ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Caliendo, Marco; Mahlstedt, Robert; Mitnik, Oscar A.

Working Paper

Unobservable, but Unimportant? The Influence of Personality Traits (and Other Usually Unobserved Variables) for the Evaluation of Labor Market Policies

IZA Discussion Papers, No. 8337

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Caliendo, Marco; Mahlstedt, Robert; Mitnik, Oscar A. (2014) : Unobservable, but Unimportant? The Influence of Personality Traits (and Other Usually Unobserved Variables) for the Evaluation of Labor Market Policies, IZA Discussion Papers, No. 8337, Institute for the Study of Labor (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/101909

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

DISCUSSION PAPER SERIES

IZA DP No. 8337

Unobservable, but Unimportant? The Influence of Personality Traits (and Other Usually Unobserved Variables) for the Evaluation of Labor Market Policies

Marco Caliendo
Robert Mahlstedt
Oscar A. Mitnik

July 2014

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Unobservable, but Unimportant? The Influence of Personality Traits (and Other Usually Unobserved Variables) for the Evaluation of Labor Market Policies

Marco Caliendo
University of Potsdam, IZA, DIW Berlin and IAB

Robert Mahlstedt
IZA and University of Potsdam

Oscar A. Mitnik
Federal Deposit Insurance Corporation and IZA

Discussion Paper No. 8337
July 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 8337
July 2014

ABSTRACT

Unobservable, but Unimportant? The Influence of Personality Traits (and Other Usually Unobserved Variables) for the Evaluation of Labor Market Policies*

Many commonly used treatment effects estimators rely on the unconfoundedness assumption ("selection on observables") which is fundamentally non-testable. When evaluating the effects of labor market policies, researchers need to observe variables that affect both treatment participation and labor market outcomes. Even though in many countries it is possible to access (very) informative administrative data, concerns about the validity of the unconfoundedness assumption remain. The main concern is that the observed characteristics of the individuals may not be enough to properly address potential selection bias. This is especially relevant in light of the research on the influence of personality traits and attitudes on economic outcomes. We exploit a unique dataset that contains a rich set of administrative information on individuals entering unemployment in Germany, as well as several usually unobserved characteristics like personality traits, attitudes, expectations, and job search behavior. This allows us to empirically assess how estimators based on the unconfoundedness assumption perform when alternatively including or not these usually unobserved variables. Our findings indicate that these variables play a significant role for selection into treatment and labor market outcomes, but do not make for the most part a significant difference in the estimation of treatment effects, compared to specifications that include detailed labor market histories. This suggests that rich administrative data may be good enough to draw policy conclusions on the effectiveness of active labor market policies.

JEL Classification: C21, D04, J68

Keywords: matching, unconfoundedness, unobervables, selection bias, heterogeneity, personality traits, active labor market policy

Corresponding author:

Marco Caliendo
University of Potsdam
Chair of Empirical Economics
August-Bebel-Str. 89
14482 Potsdam
Germany
E-mail: caliendo@empwifo.uni-potsdam.de

* This study uses the *IZA Evaluation Dataset IAB Linked* (for more information, see Arni, Caliendo, Künn, and Zimmermann, 2014). The IAB (Nuremberg) kindly gave us permission to use the administrative data. The linked administrative data were prepared and provided for this research project only. The authors thank Michela Bia, Hilmar Schneider and Jeffrey A. Smith for helpful comments and suggestions. We also thank the participants in seminars at IAB Nuremberg and CEPS Luxembourg for valuable comments.

1 Introduction

Evaluating the *causal* effect(s) on outcomes of an intervention or treatment(s) has become the key empirical objective in many areas of Economics, Statistics, and other fields like Sociology, Political Science, Epidemiology, and Medicine. Among the most exhaustively studied interventions are Active Labor Market Policies (ALMP), both using experimental and nonexperimental methods. After the influential study by LaLonde (1986) raised concerns on the ability of nonexperimental methods to replicate the results of ALMP experiments, a very large literature developed analyzing methodological aspects related to ALMP evaluation, and nonexperimental methods in general.[1] A key ever-present question that nonexperimental ALMP evaluations face is whether the data can account fully for all the factors that explain both the participation in, and the outcomes of, a program. The objective of this paper is to address this question, relying on unique data on several characteristics usually not observed in the context of ALMP evaluations, for individuals entering unemployment in Germany.

In an ideal experiment where units (e.g. individuals) are randomly assigned to the alternative treatment levels, the average causal effect of any given treatment level compared to another can be obtained by just comparing the average outcomes across the treatment levels. However, if the assignment to treatment is nonexperimental, assumptions are needed to identify the treatment effects of interest. One of the most popular approaches is based on the assumption referred alternatively as unconfoundedness, exogeneity, ignorability, or selection on observables (Heckman, LaLonde, and Smith, 1999; Imbens and Wooldridge, 2009). In a binary setting where units are either treated or used as comparisons (controls), the assumption implies that after controlling for differences in observed covariates between the two groups, any remaining differences are as if they had been generated by random assignment to the groups. Many commonly used treatment effects estimators (e.g. propensity score matching and weighting) rely on this unconfoundedness assumption, which is fundamentally non-testable. Furthermore, a leading critique is that it is not realistic. In the context of ALMP evaluations the argument is that researchers need to observe all the variables that affect both treatment participation and labor market outcomes. The main concern is that there may be *unobserved* characteristics that simultaneously explain the

[1] See Heckman, LaLonde, and Smith (1999) for a survey of the ALMP evaluation literature and the early debate on the LaLonde (1986) study; see Kluve (2010) and Card, Kluve, and Weber (2010) for an overview of ALMP evaluation in Europe; see Dehejia and Wahba (1999, 2002), Smith and Todd (2005) and Dehejia (2005) on the debate on using propensity score matching to evaluate the training program in the LaLonde study; see Imbens and Wooldridge (2009) for a recent survey of econometric methods used in program evaluation.

particular treatment individuals received, and the treatment outcomes.[2] In this case, estimators based on the unconfoundedness assumption become biased, either underestimating or overestimating the causal effects of the treatment. Looking back at the last decade, the development is twofold. On the one hand, many countries now offer access to (very) informative and complete administrative data – including detailed information on the labor market history of individuals – increasing the likelihood that the unconfoundedness assumption is satisfied. On the other hand, the recent literature showing the influence of variables such as personality traits or preferences on economic outcomes (e.g. Heckman, Stixrud, and Urzua, 2006; Osborne Groves, 2005; Bowles, Gintis, and Osborne, 2001), should be a cause of concern about the validity of the unconfoundedness assumption; these variables might be important on many dimensions in the context of ALMP (e.g. job search behavior, selection into programs, overall labor market performance) but have not been used previously as conditioning variables in this context.

In this paper we address this concern explicitly. Our focus is on a family of estimators that rely on comparing treated and control individuals based on the propensity score, the probability of receiving treatment, as a way of reducing a multi-dimensional problem to a one-dimensional problem (Rubin, 1974). Most of the technical details regarding the properties and features of estimators based on the propensity score are at this point well understood. Intuitively, propensity score-based estimators are based on comparing individuals with similar scores. The key is that even if individuals in the two treatment arms have similar values of their estimated propensity scores (and thus in the observed characteristics used to estimate the propensity score), they could still differ in other characteristics not observed by the researcher. The data we exploit is from Germany and unique because, similar to many other ALMP evaluations (in particular in Europe), it not only contains a rich set of administrative-based information for a fresh inflow sample into unemployment, but it also contains information on characteristics usually not observed in the context of ALMP evaluations, like personality traits, attitudes, expectations, and job search behavior.[3] This allow us to empirically assess how estimators based on the unconfoundedness assumption perform when alternatively including or not these usually unobserved variables.

As the variables measuring personality traits, attitudes, expectations, and job search behav-

[2]Even though the literature uses "selection on observables" as a way of referring to the unconfoundedness assumption, and the term "unobservables" is also commonly used, we prefer to use the term "unobserved" to highlight the fact that the observability of a particular variable will vary for different contexts and data.

[3]For example Gerfin, Lechner, and Steiger (2005) for Switzerland, Sianesi (2004) for Sweden, or Lechner, Miquel, and Wunsch (2011) and Biewen, Fitzenberger, Osikominu, and Paul (2014) for Germany, make use of comprehensive administrative data in order to evaluate ALMP programs in (Western) European countries. However, those studies generally suffer from the absence of information about personality traits, attitudes and expectations.

ior are usually not observed in the context of ALMP evaluations, there is no prior research on the actual consequences, when using propensity score-based estimators, of not observing these types of variables. However, this paper relates to the prior literature dealing with the sensitivity of unconfoundedness-based estimators. Imbens (2003), Altonji, Elder, and Taber (2005) and Ichino, Mealli, and Nannicini (2008) have proposed methods to assess the sensitivity of unconfoundedness-based estimators to the presence of unobserved variables. With methodological differences in their approaches, these studies try to assess how large should the effect of hypothetically not observed variables be to invalidate the results obtained from applying propensity score-based estimators in different situations. Oster (2013) relates Altonji, Elder, and Taber (2005) to coefficient movements when adding controls in linear models. Lechner and Wunsch (2013) explore, using a German dataset, how sensitive matching estimators are to the inclusion of a variety of usually observed (but rich) characteristics, and find that those rich characteristics can remove selection bias.

Building upon this previous literature, we estimate different treatment selection models using alternative sets of variables, for three typical ALMP programs – short-term training, long-term training and wage subsidies – which are very common in many countries. We examine the resulting propensity score distributions and ranks. Based on these selection models we estimate average treatment effects on the treated using different treatment effects estimators, and compare the effects associated to the alternative variable sets. We also examine the resulting matching quality and perform extensive sensitivity analyses. Our findings indicate that personality traits and other usually unobserved variables play a significant role for selection into treatment and labor market outcomes. However, comprehensive control variables (in particular labor market histories) are able to capture partly the same information provided by the usually unobserved variables. Thus, the difference in treatment effects between including and excluding the usually unobserved variables is in general small. Overall we interpret our results as suggesting that rich administrative data that includes detailed labor market histories may be good enough to draw policy conclusions on the effectiveness of active labor market policies.

The paper is structured as follows. The next section gives a short summary on the identification of treatment effects and the role of potentially unobserved variables. Section 3 describes the institutional background and the dataset, and presents some descriptives statistics. Section 4 presents the results, while Section 5 concludes.

2 Identifying Treatment Effects - A Short Summary

2.1 Basic Setup

We base our discussion on the well known potential outcomes framework (Roy, 1951; Rubin, 1974), where we have two potential outcomes (Y^1 with treatment, Y^0 without treatment), a treatment indicator ($W = 1$ if individual received treatment) and a set of observed characteristics X. We focus on the usual parameter of interest in most evaluations, the average treatment effect on the treated (ATT):

$$\tau_{ATT} = E(Y^1 \mid D = 1) - E(Y^0 \mid D = 1). \tag{1}$$

The fundamental evaluation problem arises because the last term on the right hand side of equation (1) is not observed and selection bias arises if participants and non-participants are selected groups in terms of (un)observed characteristics who would have different potential outcomes even in the absence of treatment. To correct for this selection bias, propensity score matching estimators rely on the unconfoundedness or conditional independence assumption (CIA), which states that conditional on observed characteristics (X) the counterfactual outcome is independent of treatment. In addition to the CIA, we also assume overlap which implies that there are no perfect predictors which determine participation.[4] Additionally, as direct matching on X can be difficult when X is of high dimension ("curse of dimensionality"), Rosenbaum and Rubin (1983) suggest using balancing scores such as the propensity score. These two assumptions

$$Y^0 \perp D | P(X) \quad \text{and} \quad Pr(D = 1 \mid P(X)) < 1, \text{for all X} \tag{2}$$

are sufficient for identification of the ATT based on matching. The CIA is a strong assumption and its justification depends crucially on the availability of data which allow the researcher to control for all relevant variables that simultaneously influence the participation decision and the outcome variable. If there are unobserved variables which affect assignment into treatment and the outcome variable simultaneously, a *hidden bias* might arise to which matching estimators are not robust (see, e.g. Rosenbaum, 2002, for an extensive discussion). To see what happens, let us assume that the participation probability is not only determined by X, but also by a set of (unobserved) variables U. Then

$$P_i = P(X_i, U_i) = P(D_i = 1 \mid X_i, U_i) = F(\beta X_i + \gamma U_i), \tag{3}$$

[4] See Rubin (2008) for key design issues that need to accompany these assumptions for a nonexperimental study to be considered an appropriate substitute for a randomized experiment.

where γ is the effect of U_i on the participation decision. If the study is free of hidden bias, γ will be zero and the participation probability will solely be determined by X_i. However, if there is hidden bias, two individuals with the same observed covariates X have different chances of receiving treatment. Let us assume we have a matched pair of individuals i and j and further assume that F is the logistic distribution. The odds that individuals receive treatment are then given by $\frac{P_i}{(1-P_i)}$ and $\frac{P_j}{(1-P_j)}$, and the odds ratio is given by $\frac{\frac{P_i}{1-P_i}}{\frac{P_j}{1-P_j}} = \frac{P_i(1-P_j)}{P_j(1-P_i)} = \frac{\exp(\beta x_i+\gamma u_i)}{\exp(\beta x_j+\gamma u_j)}$. If both units have identical observed covariates – as implied by the matching procedure – the X-vector cancels out implying that the odds ratio is $exp[\gamma(u_i - u_j)]$. Both individuals differ in their odds of receiving treatment by a factor that involves the parameter γ and the difference in their unobserved covariates U.

Here is key to consider the relationship between X and U. If there is high correlation between X and U, it is very likely that the $u_i - u_j$ difference will be small. For example, motivation may be part of U. If motivation is uncorrelated with the observed characteristics of the individuals, differences in motivation may be large for individuals with the same values of X. However, in ALMP evaluations it is common to include in the vector X variables related to the labor market histories of individuals. In particular, assume X includes lagged values of the outcome variable, $Y_{t-1}, ..., Y_{t-l}$. If motivation is correlated with contemporaneous values of the outcome, it is very likely to be correlated with lagged values of the outcome. In that case, the difference $u_i - u_j$ is very likely to be small for individuals with similar (or in the extreme identical) outcome histories.

2.2 The Role of Unobserved Variables in Labor Market Policy Evaluation

The extent to which γ and U play an empirical role will depend on the empirical context. The importance of the unobserved characteristics U clearly depends on the extent of the observed characteristics. A more informative set of control variables X reduces the likelihood that γ has an effect on the participation decision. Previous studies suggest that socio-demographic and regional information as well as labor market histories of participants play an important role when evaluating treatment effects (e.g. Mueser, Troske, and Gorislavsky, 2007; Heckman, Ichimura, Smith, and Todd, 1998). Especially, the improved availability and quality of administrative data in recent years allows investigating the effects of certain characteristics on potential treatment effects in a systematic way (e.g. Lechner and Wunsch, 2013; Huber, Lechner, and Wunsch, 2013; Biewen, Fitzenberger, Osikominu, and Paul, 2014). However, at the same time a variety of studies shows the importance of variables previously not extensively considered in economics in general and for ALMP evaluations, like personality traits (Nyhus and Pons, 2005), cognitive and non-

cognitive skills (Heckman, Stixrud, and Urzua, 2006) or preferences and attitudes (Pannenberg, 2010; Belzil and Leonardi, 2007).

In this context several variables which are usually not observed when evaluating labor market policies, might be of special interest. For example, Mueller and Plug (2006) find for the U.S. that the 'Big Five' personality traits – extraversion, agreeableness, conscientiousness, neuroticism and openness – have an impact on earnings similar to that found for cognitive abilities. Similar results are found by Nyhus and Pons (2005) for the Netherlands. Moreover, several empirical studies investigate how an individual's locus of control might be related to labor market performance. Locus of control refers to a general expectation about internal versus external control of reinforcement (Rotter, 1966). People with a more external locus of control believe that much of what happens in life is beyond their control, while people with an internal locus of control see life's outcomes as dependent on their own decisions and behavior. Several studies find a significant effect of the locus of control on individual earnings (e.g. Andrisani, 1977; Heineck and Anger, 2010; Semykina and Linz, 2007) and job search strategies (e.g. Caliendo, Cobb-Clark, and Uhlendorff, 2014; McGee, 2014).

Another strand of the literature analyzes the effect of job search choices, like the use of different search channels (e.g. Holzer, 1988; Van den Berg and Van der Klaauw, 2006) or reservation wages (e.g. Shimer and Werning, 2007; Brown, Roberts, and Taylor, 2010), on individual labor market performance. Based on theoretical job search models (Mortensen, 1986), these variables are significant in determining the unemployment duration and the speed at which job-seekers are reintegrated into the labor market. Moreover, the intergenerational transmission of human capital (Black, Devereux, and Salvanes, 2005) and attitudes (Dohmen, Falk, Huffman, and Sunde, 2012), as well as social networks (Montgomery, 1991; Bayer, Ross, and Topa, 2008) seem to play an important role in determining an individual's labor market performance. Combining these different strands of literature the natural question that arises is whether these variables also play a role when evaluating the effects of active labor market programs.

[Insert Table 1 about here]

Table 1 gives an overview of the covariates of interest in our study. First of all, we divide the variables in five categories: 1) socio-demographic variables, 2) labor market history, 3) personality traits, 4) job search and employment outlook and 5) socio-cultural characteristics. The first category includes, beside typical *socio-demographic variables*, like age, gender, marital status, educational level, health status, etc., also control variables for the month of entry into unemployment, the time between entry and the interview, and regional information. The second block,

labor market history, can be divided into short-term and long-term history. The short-term history summarizes information up to two years prior to the beginning of the current unemployment spell, while the long-term history includes information for the prior ten years. Additionally, we define three groups of variables containing information on usually unobserved variables. *Personality traits* contain the 'Big Five' factors of personality openness, conscientiousness, extraversion and neuroticism (agreeableness is not included due to missing items) (see Digman, 1990, for an overview) and locus of control. The *job search and employment outlook variables* contain information on the individual reservation wage, the search intensity, search channels, as well as expectations about future employment prospects and program participation. Finally, the block *socio-cultural characteristics* includes variables like the number of good friends outside the family, problems with child care, father's education and individual life satisfaction.

3 Institutional Background, Data and Descriptives

3.1 Institutional Background

Germany has a long tradition of ALMP and the German Social Security Code provides a large set of programs geared towards helping unemployed individuals, like training programs, job search assistance, job creation schemes, start-up subsidies or benefits to increase the job-seeker's labor market mobility. Table 2 shows the entries into different programs in Germany between 2005 and 2011. As they are most relevant (in terms of number of participants) for supporting unemployed job-seekers and very typical for many OECD countries, we investigate the effect of three programs in detail: 1) Short-term training, 2) long-term training and 3) wage subsidies. While the short-term training represents a more recent group of programs shifting the focus towards more 'activating' elements, long-term training and wage subsidies represent more traditional programs, which aim to remove disadvantages in education, work experience or productivity. Since these programs represent very different re-integration strategies and are targeted at different types of unemployed individuals, this potentially allows us to examine the role of usually unobserved variables for these different selection processes. Let us briefly summarize some institutional details for these three programs.

[Insert Table 2 about here]

Short-term training measures, introduced in 1998, have a maximum duration of eight weeks. The courses can either serve as test of the participant's occupation-specific aptitude, or aim to improve the general employability. For example, the courses teach the unemployed how to apply

effectively for a new job or how to behave in job interviews, but can also consist of computer or language classes. Some of the courses impart knowledge on starting a business to founders of start-ups, while others are concerned with the special needs of certain 'hard-to-place' job-seekers. Caseworkers can also use them to attain additional information on the participant's abilities and willingness to work. Courses are conducted full- or part-time and last from two days up to eight weeks; an individual's time spent in short-term training programs is limited to twelve weeks in total. While in a short-term training program an unemployed person cannot earn additional wages; however she continues to receive unemployment benefits and coverage of the costs associated to participation (e.g. transportation, child care) (see Wolff and Jozwiak, 2007).

Long-term training programs have been a well established part of the German labor market policy for many decades. These programs can last from three months to up to three years. Historically, a caseworker would assign an unemployed individual to a specific course aimed at improving her occupational skills, and facilitating reintegration into the labor market. Previous studies find positive effects only in the very long-run (e.g. Fitzenberger, Osikominu, and Völter, 2008; Lechner, Miquel, and Wunsch, 2011) or even partly negative effects on employment (e.g. Lechner and Wunsch, 2008). With the 'Hartz reforms' at the beginning of the century, the German government reduced the usage of (long) vocational training programs. From 2003 onwards caseworkers no longer choose a specific course for the unemployed but hand out a training voucher to the job-seeker who is then allowed to find an appropriate training program for herself (see Bernhard and Kruppe, 2012).

Wage subsidies are one of the oldest instruments used to reintegrate unemployed individuals into the labor market. The aim of the subsidy is to reduce the labor costs for the firm, potentially bridging any deficiencies in a worker's productivity. Wage subsidies (or temporary employment with a wage subsidy) can also be used as a screening device, lowering uncertainty and, hopefully, creating stable employer-employee relationships. Whether or not an unemployed person is supported with a targeted wage subsidy is a decision that is made by her caseworker. In addition, the caseworker determines the properties of the subsidy (restricted by the legal framework and guidelines): up to 50 percent of the monthly wage can be covered by the subsidy for at most 12 months. Extensions are possible if the wage subsidy aims at the integration of older or handicapped workers. Employers of subsidized workers agree to employ workers who are younger than 50 years for a follow-up period after the subsidy ends. This follow-up period is usually as long as the subsidized period itself. In case the worker is dismissed for reasons that are not attributable

to her performance, the employer has to return a portion of the subsidy. Previous research shows relatively large favorable effects on employment prospects of hard-to-place workers (e.g. Bernhard and Wolff, 2008; Jaenichen and Stephan, 2011), but potential substitution/displacement effects and deadweight losses may be present.

3.2 The IZA Evaluation Dataset and our Estimation Sample

This study uses the *IZA Evaluation Dataset IAB Linked* which combines survey information and administrative data on individuals who entered unemployment between June 2007 and May 2008 in Germany (see Caliendo et al., 2011). The dataset contains a 9% random sample, from the monthly unemployment inflows of approximately 206,000 individuals identified in the administrative records, who are selected for interview. From this gross sample of individuals aged between 16 and 54 years, representative samples of about 1,450 individuals are interviewed each month so that after one year twelve monthly cohorts were gathered (see Arni et al., 2014, for details on represantativeness etc.). The first wave of interviews takes place shortly after the entry into unemployment. Besides the extensive set of individual-level characteristics and labor market outcomes, the individuals are asked a variety of non-standard questions about search behavior, social networks, psychological factors, cognitive and non-cognitive skills, subjective assessments on future outcomes, and attitudes. For the 88% of individuals who agreed, these survey data were then merged to administrative information from the *Integrated Employment Biographies* (IEB) provided by the Institute for Employment Research (IAB).[5] The IEB integrates different sources, e.g., employment history, benefit recipient history, training participation history and job search history and therefore contains detailed information on employment subject to social security contributions, unemployment and participation in active labor market policy including wages and transfer payments. The data additionally include a broad range of socio-economic characteristics including education, family status and health restrictions. The data do not contain information about working hours or periods in self-employment, working as a civil servant, or time spent in inactivity. Altogether, this amounts to a total of 15,274 realized interviews with a time lag from seven to fourteen weeks between the unemployment registration and the interview.

For the purpose of the study, we restrict our estimation sample to all individuals who are still unemployed and do not participate in any ALMP program when the interview takes place. Furthermore, we include only respondents who state that they are actively searching for new employment. We define job seekers as participants if they attend short-/long-term training, or

[5]This study is based on a weakly anonymized sample of the Integrated Employment Biographies by the IAB (V.901).

receive a wage subsidy within the first twelve months after the entry into unemployment, and as non-participants if they do not participate in any ALMP program within this period. This leaves us with 3,769 non-participants, while 1,607 individuals participate in short-term training, 694 in long-term training and 501 receive a wage subsidy.[6]

3.3 Some Descriptives for Outcomes and (Usually Un)Observed Variables

We observe every job-seeker in our sample for a period of 30 months after entering unemployment. To evaluate the influence of usually unobserved variables on the treatment effects we focus on labor market outcomes which are typically used in the evaluation of ALMP programs. In particular we concentrate our analysis on the employment probabilities after 12 months and at the end of our observation period after 30 months. Additionally, we observe the time spent in employment as well as the cumulative earnings within the observation period of 30 months. The upper part of Table 3 shows the differences between participants and non-participants with respect to these labor market outcomes. We observe significantly lower employment probabilities for participants after 12 months which tend to vanish over time. For recipients of a wage subsidy, the raw employment probabilities are higher after 30 months. However, the overall time spent in employment within 30 months after the entry is significantly lower for participants, irrespective of the type of treatment. The cumulative earnings are also lower for participants in short- and long-term training, but higher for recipients of wage subsidies. The lower employment probabilities in the short-run are induced by 'lock-in' effects, i.e. a reduction of participants' search effort before and during program participation (e.g. Van Ours, 2004; Jespersen, Munch, and Skipper, 2008). Lock-in effects are largest for participants in long-term training, as programs are still ongoing 12 months after entering unemployment.

[Insert Table 3 about here]

Additionally, Table 3 shows differences with respect to the socio-demographic variables and the labor market history, i.e. the baseline control variables. ALMP participants are in general

[6]The choice of the period for the split is arbitrary and could be debated (see Sianesi, 2004); nevertheless it is a standard procedure in the evaluation of ALMP. In our case, choosing 12 months as the treatment period covers about 89% of all individuals who participate in an ALMP program within our complete observation period of 30 months and ensures that we observe individuals for a sufficiently long time window after the treatment. Moreover, increasing the treatment period, has the disadvantage that the non-participation in later periods is to some extent simply the consequence of a successful job search in earlier periods. Therefore, it becomes less clear whether the estimated effects are causal to the program participation. Alternatively, duration models would allow us to control for the exact timing of the treatment, however additional distributional assumptions would be necessary (see for example Card, Kluve, and Weber, 2010, for an overview of potential estimation strategies when evaluating 199 worldwide ALMP programs).

more likely to be female, older and have more children than non-participants. Moreover, participants are generally better educated, we observe a higher share with an upper secondary school and an university degree, have spent more time in employment in the past (short- and long-term) and are more likely to live in West-Germany. Comparing the three programs of interest in detail the overall pattern is very similar. However, we observe some interesting deviations. For example, participants in long-term training had a significantly higher income before the beginning of the unemployment spell, while among recipients of wage subsidies, there is a significantly lower share of individuals with health restrictions or disabilities and a higher share of people from East-Germany.

[Insert Table 4 about here]

We have discussed in Section 2.2 the role that some usually unobserved variables might play for the selection into the different types of programs. Table 4 shows significant differences between participants and non-participants for a variety of these variables. We follow our definition from Table 1 and distinguish three categories: 1) Regarding the personality traits, participants in short-term training show a higher level of neuroticism, a lower internal locus of control and a lower level of extraversion. The latter is also true and the observed differences are even larger for participants in long-term training, while recipients of wage subsidies show a significantly higher level of conscientiousness. 2) However, the observed differences between participants and non-participants with respect to personality traits seem to be relatively low compared to the second group of covariates, the job search and employment outlook variables. All groups of participants show a higher job search intensity (measured as the number of applications) and use significantly more search channels than non-participants. Additionally, participants in short-term training and recipients of wage subsidies have also a significantly lower reservation wage. Considering the importance of these variables within the job search process it is obvious that these differences are simultaneously related to the selection into the treatment and the labor market performance. Moreover, participants of all programs, ex-ante expect a higher probability to participate in a program, while participants in short- and long-term training programs have a lower expectation of their future job finding probability. Since these personal expectations are likely to be correlated with other unobserved cognitive or non-cognitive abilities, controlling for these differences in expectations potentially reduces hidden bias when evaluating the effect of active labor market policies (van den Berg, Bergemann, and Caliendo, 2009). 3) Finally, there are some differences in the socio-cultural characteristics. In general, all participants report a lower life satisfaction and a lower likelihood that their father has an A-level qualification, while

participants in long-term training are also more likely to have problems with child care. We will examine in the next Section now in detail what role these variables play in the selection process and for the estimation of treatment effects.

4 Empirical Results

4.1 Estimation Strategy

The objective of our study is to examine how estimators based on the unconfoundedness assumption perform when alternatively including – or not – usually unobserved variables. The implementation of propensity score matching and weighting estimators is a two-step-procedure where in a first step the participation model is estimated. The resulting participation probabilities are then used in a second step to match participants with similar non-participants. In the next subsections we evaluate the effect of the usually unobserved variables on each of these steps.

To conduct our analysis in a systematic way, we start by estimating 'standard' specifications for the propensity score as baseline results and subsequently include the three groups of usually unobserved variables. Our specifications mirror the definitions in Table 1 which also contains the exact definition of the three specifications: The first model (Base 1) only uses socio-demographic characteristics, family characteristics and variables related to unemployment entry and local economic conditions at unemployment entry. The second model (Base 2) additionally includes labor market history-related variables, which are consistently found to be key drivers of selection into training in the literature (e.g. Dolton and Smith, 2011). These two baseline specifications allow us to identify the estimated ATT assuming that the CIA from condition (2) holds for two different sets of covariates X. The third model (Full) additionally includes the personality traits, job search and employment outlook variables variables and socio-cultural characteristics.[7]

Based on these specifications, in Section 4.2 we asses the relevance of the usually unobserved variables on the participation probability by calculating different goodness of fit measures, and by examining the consequences for the propensity score distribution and the resulting rank order. In Section 4.3 we examine the resulting matching quality (in terms of the mean standardized bias), and in Section 4.4 we show the consequences for estimating the treatment effects of the different programs on employment and earnings. The key intuition behind our approach is that we identify the ATT assuming that the CIA holds alternatively for the usually observed

[7]For reasons of clarity and brevity we only present the estimates of the specifications Base 1, Base 2 and Full. Estimation results for sequential additions of each group of variables can be found in Appendix A.1.

covariates X, and the usually unobserved covariates U. Comparing the estimated treatment effects allows us to determine the sign and magnitude of potential hidden bias due the exclusion of the usually unobserved variables.

4.2 Relevance for Propensity Score Estimation

We start the analysis by estimating the propensity score for each program using a logit model, as is standard in the literature. Table 5 presents the average marginal effects from propensity score models for ALMP participation in short-term training, long-term training and wage subsidies, with all tables depicting the specifications from Table 1.

Significance of Variables The estimation results in Table 5 show that (except for short-term training) older job-seekers are more likely to participate while a positive effect of being female disappears once we control for the usually unobserved variables. For all programs it holds that the short-term labor market history has a strong influence on the selection into the treatment, while the long-term history is less important. Furthermore, several of the usually unobserved variables have a significant impact on the participation probabilities. For example, extraversion has a strong negative impact on the probability of participating in long-term training, while an external locus of control increases the probability of short-term training. However, none of the personality traits has a significant impact on the receipt of wage subsidies.

[Insert Table 5 about here]

The job search and outlook variables have a very strong impact irrespective of the program under consideration. Job-seekers who put more effort on their job search, in terms of number of applications and search channels, are also more likely to participate in ALMP programs; reservation wages have a negative significant effect for all programs, except for long-term training. Moreover, the effect of the subjective probability of receiving the treatment is quite large, about 15 percentage points for participation in long-term training, consistent with the individuals possessing private information which is not captured by other variables.[8] In the same way, a high expected probability of finding a job in the next 6 months decreases the probability of long-term training participation by 6.5 percentage points. As a reference, the overall participation probability is 15.6 percentage points. Within the socio-cultural characteristics, only a higher life satisfaction has a negative significant impact on the participation probabilities. Testing

[8] However, this does not hold for recipients of wage subsidies. This might be caused by the fact that the receipt of a wage subsidy is related to new employment and respondents do not consider this as a 'typical treatment'.

separately the joint significance of the different groups of usually unobserved variables it can be seen that the job search and employment outlook variables have the strongest impact on the participation decision for all programs (significant at the 1%-level). Personality traits have a significant impact on the participation in training programs (at the 10%-level for short-term training and 1%-level for long-term training), but not for wage subsidies, and the socio-cultural characteristics have no significant impact on any participation decision.

Goodness-of-Fit Overall, the differences, with respect to the marginal effects between the specific models, tend to be relatively small, while the percentage of observations correctly predicted between the Base 1 specification and the Full specification increases between 4.7 percentage points for wage subsidies and 9.5 percentage points for long-term training. However, about half of the increased hitrate can be explained by the labor market history, while the other half can be attributed to usually unobserved variables. To summarize the results in the prior tables, Table 6 shows three different goodness of fit measures for the previous models.[9] The first measure, the McFadden (1974) R^2 is defined as the 'likelihood ratio index' of the estimated model and a model with all coefficients but the intercept constrained to zero. The second measure, the R^2 of McKelvey and Zavoina (1975) corresponds to the regression variation divided by the total variation in the latent index function model, where the calculations are based on predicting the continuous latent variable underlying the observed binary variable. Finally, the R^2 of Efron (1978) is the sum of squared model residuals divided by the total variability in the dependent variable. It is equal to the squared correlation between the predicted values and the actual values.

[Insert Table 6 about here]

As it is typical for the evaluation of ALMP programs using individual-level data, the overall level of the R^2 is relatively low (e.g Dolton and Smith, 2011; Huber, Lechner, and Wunsch, 2013). However, we observe substantial differences when including or not the different groups of control variables. For example, considering McKelvey and Zavoina's R^2, the base specification 1, including only socio-demographics, achieves only between 28% (long-term training) and 40% (wage subsidies) of the goodness of fit of the full model, including all variables; additionally including the labor market history increases it up to 49% (long-term training) and up to 80% (wage subsidies). Hence, the labor market histories seem to explain the participation in wage subsidies especially well. Including the usually unobserved variables has relatively high additional

[9] We show three measures of goodness-of-fit as there is no agreed-upon measure for non-linear models.

explanatory power when estimating the participation in long-term training, while it only slightly increases it for wage subsidies. The effect varies between 20% (for wage subsidies) and 51% (for long-term training). The overall pattern looks very similar for McFadden's and Efrons's R^2 while the absolute values tend to be lower. The results suggest that the baseline variables do very well in predicting the selection into wage subsidies, while the job search and outlook variables especially increase the goodness of fit for long-term training substantially.

Score Distribution Figure 1 presents the propensity score distribution for each specification, separately for the three types of treatment and by treatment arm, for the full estimation sample. Unsurprisingly, including additional control variables reduces the concentration of the propensity scores around the mean and shifts more probability mass to the tails of the distribution. Consistent with the observed improvements in goodness of fit, those shifts imply more mass at lower propensity score values for non-participants and at higher propensity score values for participants. With respect to the different types of treatment, the shifts are lower for wage subsidies than for short- and long-term training. Comparing the baseline specification 1 and the full specification for short-term training the propensity score increases for 61.2% of the participants, with a mean change of 1.5 percentage points, and 39.3% of the non-participants, with a mean change of -3.6 percentage points. Adding the labor market history to the base specification has only little impact: The propensity score increases for 59.4% of the participants, with a mean change of 1.9 percentage points, and 42.9% of the non-participants, with a mean change of -0.8 percentage points. For wage subsidies, the overall pattern looks very similar. Switching from the base specification 1 to the full specification raises the propensity score for 61.2% of the participants (mean change: 4.5 percentage points) and 38.6% of the non-participants (mean change: -0.6 percentage points). Using specification 2 as the baseline, the propensity score increases for 62.3% (mean change: 1.5 percentage points) and 44.4% of the non-participants (mean change: -0.2 percentage points).

[Insert Figures 1 and Table 7 about here]

Score Correlations Table 7 provides several correlation measures between specifications, for the full sample and by treatment arm. The propensity score and rank correlation between the base and full specifications is the highest for wage subsidies and the lowest for long-term training, which suggests that the usually unobserved variables have the highest impact on the selection into long-term training and the lowest on that into wage subsidies. For the latter, the usually unobserved variables provide little extra information that is not captured by the

socio-demographic and labor market history variables. Finally, we compare the propensity score distributions of both base specifications with that of the full specification. For example, the paired t-test shows no distributional differences for the full sample, but significant variation when comparing treated and non-treated individuals separately, irrespective of the treatment under consideration. With respect to the rank distribution, the Friedman test shows highly significant differences for all programs.

4.3 Consequences for Matching Quality

What has to be kept in mind in the first implementation step is, that the overall aim of the procedure is not to predict the propensity score as well as possible, but to balance the distribution of covariates between participants and non-participants, and we have to check this explicitly. One suitable indicator is the mean standardized bias (Rosenbaum and Rubin, 1983), which assesses the distance in marginal distribution of the covariates before and after matching. For each covariate X it is defined as the difference of sample means in the treated and matched control subsamples as a percentage of the square root of the average of sample variances in both groups:

$$SB(x) = \frac{100(\bar{x}_c - \bar{x}_t)}{\sqrt{(s_{xc}^2 + s_{xt}^2)/2}}, \tag{4}$$

with $\bar{x}_c$ being the mean of the control group, $\bar{x}_t$ the mean of the treatment group, s_{xc}^2 the variance of the control group and s_{xt}^2 the variance of the treatment group. In Table 8 we present the mean of $SB(x)$, MSB, for groups of variables and overall, for the different specifications, treatments and samples. In our setting, it is especially interesting to assess the matching quality for the usually unobserved variables for those propensity score specifications (Base 1 and Base 2) that do *not* rely on the usually unobserved variables; it is a useful way to summarize the degree to which the socio-demographic characteristics and labor market history can proxy for the usually unobserved variables in the selection into treatment process.

The first column of Table 8 presents the *raw* MSB, i.e. prior to matching, while the next three columns present the MSB when matching with the alternative specifications of the propensity score: only using socio-demographics (Base 1), using socio-demographics and the labor market history (Base 2), and adding the usually unobserved variables (Full).

[Insert Table 8 about here]

Using the full propensity score specification reduces the overall MSB down to 1.9 for short-term training, to 3.0 for long-term training, and to 3.7 for wage subsidies, values all below

the target of 3% to 5% usually considered appropriate in empirical studies. However, on closer examination we find substantial differences with respect to the different programs and groups of control variables. For all programs, we find the largest raw bias with respect to the job search and employment outlook variables (MSB of 15.0 for wage subsidies, 16.5 for short-term training and 24.3 for long-term training). These MSB values are substantially larger than for most other groups of variables, except for the labor market history in the wage subsidies treatment (14.1). However, the latter is reduced to 2.6 by including both sets of baseline variables as controls. The same is not true with regard to the job search and employment outlook variables: the MSB is at best only slightly reduced, for all three treatments, when using the propensity score specification Base 1 or Base 2 (the MSB even increases for long-term training and wage subsidies under Base 2). Once the Full specification for the propensity score is used in matching, i.e. including the usually unobserved variables, the MSB for these variables is reduced substantially. In most cases the reductions are enough to bring the MSB down to below the 5% threshold. With respect to the personality variables we have the interesting finding that conditioning on Base 2 even increases the MSB for long-term training and does not do much for short-term training. Balancing for these variables is better when using the Full specification, even though the MSB remains at 6.2 for wage subsidies. These results are in line with our previous results which indicate that the job search and employment outlook variables have the strongest impact on the selection into all types of programs. More importantly, estimating the propensity score using only the baseline variables does not appear to be successful in eliminating the differences in the usually unobserved variables, which appear as very important for the selection into treatment process.

4.4 Consequences for Treatment Effects

In this section we present the consequences of using the alternative propensity score specifications for the estimation of the treatment effects of each program. There are several possible estimators for the Average Treatment on the Treated (ATT) parameters we are interested in obtaining (e.g. Imbens and Wooldridge, 2009). For the sake of clarity, we focus our analysis on a particular estimator, kernel matching, which is heavily used in evaluation studies. When relying on kernel matching estimators researchers need to specify a kernel function and a bandwidth parameter.[10] We specify an epanechnikov kernel, and a bandwidth of 0.06. In Appendix A.2 we conduct a sensitivity analysis where we specify alternative estimators, and bandwidth parameters for the

[10] In contrast to the choice of the bandwidth parameter, where a trade-off between a small variance and an unbiased estimate of the true conditional mean function arises, the choice of the kernel type appears to be relatively less important in practice (see the discussion in Caliendo and Kopeinig, 2008; Galdo, Smith, and Black, 2008).

kernel estimators. Our results are qualitatively similar using any estimator.

Table 9 presents the differences in mean outcomes (raw gap) as well as the ATT from using the kernel estimator. We use the same three main specifications for the propensity score as defined before (see Table 1).[11] As the outcomes of interest we analyze the employment probability 12 and 30 months, respectively, after the entry into unemployment, the overall time spent in employment and the cumulated earnings within 30 months. The left panel of the Table shows the ATTs, while the right panel calculates the difference in ATTs, both absolute and in percentage terms.

Shortly summarizing the estimated effects of the different programs, we find a negative effect of all treatments on the employment probability after 12 months, while the negative effect gets smaller over time and we even find a positive and significant effect for wage subsidies after 30 months. Moreover, there is a negative effect of all treatments on the overall time spent in employment within 30 months, with the most unfavorable effect for long-term training. Regarding cumulated earnings within 30 months, there is a positive effect of wage subsidies, while all other programs reduce participants earnings.

Our main interest is in the comparison of the estimated treatment effects using alternative propensity score specifications. It can be seen that the overall absolute differences are relatively small. However, some of these differences are quite large in percentage terms. For example, for short-term training, the differences in the ATT for cumulated earnings within 30 months between the Base 1 and Full specifications is as large as 27.8% (€758). However, when comparing the Base 2 and Full specifications, the difference in ATT gets reduced to 4.1% (€85). For the same treatment, however, the percentage difference in ATTs for employed after 30 months is -21.1% between the Full and Base 1 specifications, and it gets reduced only slightly to -17.8% between the Full and Base 2 specifications (in both cases, in absolute terms the difference is still small, -.002, out of an ATT of around 0.01). In general, the most distinctive differences can be generally found when estimating the effect on cumulated earnings, while the differences are relatively small when considering the employment probability at a certain point in time.

Comparing the different programs the effect of including the usually unobserved variables is stronger on the treatment effect of wage subsidies compared to the two types of training programs. Also, for the majority of the estimated effects the differences are substantially larger when comparing the Base 1 specification with the Full model, which supports the idea that a large part of the usually unobserved characteristics, especially those that are constant over time,

[11] Estimation results for a larger set of specifications including each group of usually unobserved variables separately can also be found in Appendix A.1.

can be captured by controlling for the previous labor market performance.

Note that only a few differences in ATTs are statistically significant, and only when comparing the Full and Base 1 specifications. No statistically significant differences remain when controlling also for the market history variables in Base 2.

[Insert Table 9 and Figure 2 about here]

The results for the ATTs seem surprising given the importance for the selection process that the usually unobserved variables seem to have, as indicated by the goodness-of-fit measures, the propensity score distributions, and the measures of matching quality. To better understand this issue, in Figure 2 we present matched propensity score differences, for the different treatments. Each panel presents a box plot of the difference in the Full specification propensity scores between the treated and *matched* control individuals, based on the deciles of the propensity score under the Base 2 specification. Each box plot shows the median as a line in the box, the inter-quantile range (IQR) with the height of the box, and the distance 1.5×IQR with whiskers outside the box; outliers are represented by dots outside the whiskers. Consistent with the findings in Table 6 the median differences in propensity scores are larger for long-term training, the treatment for which a larger percentage of the overall variation is explained by the usually unobserved variables. For short-term training and wage subsidies the medians across deciles are closer to zero, indicating that there are no large differences in propensity scores between the matched individuals, either when using the Base 2 or the Full specification. However, as a fair amount of variation around the median remains, it is consistent with the results in Table 8 that show that the differences in usually unobserved variables, as measured by the MSB, remain high when using the Base 2 propensity score specification.

It is useful to relate the characteristics of the two programs for which selection based on the usually unobserved variables seem to matter the most and the least, long-term training and wage subsidies respectively, to their selection process. For both programs the results related to the usually unobserved variables are intuitive: long-term training requires a high degree of commitment by the trainees, to endure a program that can be as long as three years. It is not surprising then that personality traits and job search and employment outlook variables matter for selection into this type of training. Wage subsidies are oriented to help individuals find employment in markets where the demand for labor may be weak for the particular skills of these individuals. Therefore, it is not unexpected that the labor market histories of the individuals have the main explanatory power in the selection process (in terms of goodness-of-fit the usually unobserved variables explain marginally no more than 20% of the observed variation,

with respect to the Base 2 specification, and in terms of matching quality, the MSB improves only slightly overall with the addition of these variables). In particular, prior subsidized employment influences the participation in this treatment positively compared to all other treatments (see Table 5).

Returning to the treatment effects in Table 9 it is clear that the distinct selection processes suggested by the results are not reflected by the estimated treatment effects. Given the differences in treatment effects between the Full and Base 1 specifications, compared to the differences between the Full and Base 2 specifications, it is clear that the labor market history is able to capture most of the information contained in the usually unobserved variables. This point is even stronger if some of the usually unobserved variables are stable over time as, for example, personality traits are expected to be. In general, the higher is the correlation between the usually unobserved variables and the labor market history, the smaller the additional value of the usually unobserved variables will be. Finally, note that our results are qualitatively similar when further splitting the sample by characteristics of the participants, like gender or skill level.[12]

5 Conclusion

The aim of this study was to investigate the effect of personality traits and other usually unobserved variables on the selection into active labor market policy programs and in consequence on the estimated average treatment effects. The results present a clear picture. The usually unobserved variables matter in terms of the selection process into treatment, in a different manner for the individuals treated under each of the programs. This is consistent with the three programs representing distinct re-integration strategies targeted to different types of unemployed individuals. Even though we find that the usually unobserved variables matter for selection, when estimating the effects of ALMP programs on labor market outcomes in a second step, the overall influence of including or excluding them is rather small. Our results suggests that unobserved variables have a slightly stronger impact when evaluating the effect of wage subsidies, compared to training measures, and can be stronger for different subgroups like high-skilled workers. But even in those cases the differences in treatment effects are in general not statistically significant.

The relatively small overall impact on treatment effects of the usually unobserved variables seems to be explained by the comprehensive baseline control variables, predominately the labor market history. Assuming that the usually unobserved variables are constant over time, they not only affect selection into programs and future labor market outcomes, but they are probably

[12]These results are available upon request.

correlated with past labor market performances. Thus, conditioning on labor market histories implicitly captures a large part of the information contained in the usually unobserved variables. Our results show that given our set of usually unobserved characteristics, the influence of these variables on the effect of ALMP programs in Germany is generally limited when informative administrative data are available. This suggests that lacking these usually unobserved characteristics does not affect in a fundamental way the assessment of public policies: as long as a large enough set of covariates, which include labor market history, is available; any expected biases associated to not observing some of the personality traits and job search and employment outlook variables, are likely to be sufficiently small as to not fundamentally affect policy conclusions.

It is necessary to be prudent in generalizing our results outside the setting of this study: the effects can clearly differ among different types of programs, different countries and the population of interest. Nevertheless, our study does show that valid concerns about the role of unobserved variables, when using a "selection on observables" assumption for the estimation of treatment effects, may be less relevant when observable information is available that is sufficiently correlated to the unobservable variables. This clearly seems to be the case in settings, like in many European countries these days, where policy evaluation is based on detailed administrative data.

References

Altonji, J. G., T. E. Elder, and C. R. Taber (2005): "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," Journal of Political Economy, 113(1), 151–184.

Andrisani, P. J. (1977): "Internal-external attitudes, personal initiative, and the labor market experience of black and white men," Journal of Human Resources, pp. 308–328.

Arni, P., M. Caliendo, S. Künn, and K. Zimmermann (2014): "The IZA Evaluation Dataset Survey: A Scientific Use File," IZA Journal of European Labor Studies, 3:6.

Bayer, P., S. L. Ross, and G. Topa (2008): "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," Journal of Political Economy, 116(6), 1150–1196.

Belzil, C., and M. Leonardi (2007): "Can risk aversion explain schooling attainments? Evidence from Italy," Labour Economics, 14(6), 957–970.

Bernhard, S., and T. Kruppe (2012): "Effectiveness of further vocational training in Germany: empirical findings for persons receiving means-tested unemployment benefit," Discussion Paper 10/2012, IAB, Nuremberg.

Bernhard, S., and J. Wolff (2008): "Contracting out placement services in Germany: Is assignment to private providers effective for needy job-seekers?," Discussion Paper 05/2008, IAB, Nuremberg.

Biewen, M., B. Fitzenberger, R. Osikominu, and M. Paul (2014): "The Effectiveness of Public Sponsored Training Revisited: The Importance of Data and Methodological Choices," forthcoming in: Journal of Labor Economics.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005): "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital," American Economic Review, 95(1), 437–449.

Bowles, S., H. Gintis, and M. Osborne (2001): "Incentive-enhancing preferences: Personality, behavior, and earnings," American Economic Review, pp. 155–158.

Brown, S., J. Roberts, and K. Taylor (2010): "Reservation wages, labour participation and health," Journal of the Royal Statistical Society A, 173, 501–529.

Caliendo, M., D. Cobb-Clark, and A. Uhlendorff (2014): "Locus of control and job search strategies," IZA Discussion Paper No. 4750, forthcoming in: Review of Economics and Statistics.

Caliendo, M., A. Falk, L. Kaiser, H. Schneider, A. Uhlendorff, G. van den Berg, and K. Zimmermann (2011): "The IZA Evaluation Dataset: towards evidence-based labor policy making," International Journal of Manpower, 32, 731–752.

Caliendo, M., and S. Kopeinig (2008): "Some practical guidance for the implementation of propensity score matching," Journal of Economic Surveys, 22(1), 31–72.

Card, D., J. Kluve, and A. Weber (2010): "Active Labour Market Policy Evaluations: A Meta-Analysis*," Economic Journal, 120(548), F452–F477.

Dehejia, R. (2005): "Practical propensity score matching: a reply to Smith and Todd," Journal of Econometrics, 125(1), 355–364.

Dehejia, R. H., and S. Wahba (1999): "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," Journal of the American Statistical Association, 94(448), 1053–1062.

Dehejia, R. H., and S. Wahba (2002): "Propensity Score-Matching Methods for Nonexperimental Causal Studies," Review of Economics and Statistics, 84(1), 151–161.

Digman, J. M. (1990): "Personality structure: Emergence of the five-factor model," Annual Review of Psychology, 41(1), 417–440.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2012): "The intergenerational transmission of risk and trust attitudes," Review of Economic Studies, 79(2), 645–677.

Dolton, P. J., and J. A. Smith (2011): "The Impact of the UK New Deal for Lone Parents on Benefit Receipt," Discussion paper, Institute for the Study of Labor (IZA).

Efron, B. (1978): "Regression and ANOVA with zero-one data: Measures of residual variation," Journal of the American Statistical Association, 73(361), 113–121.

Fitzenberger, B., A. Osikominu, and R. Völter (2008): "Get training or wait? Long-run employment effects of training programs for the unemployed in West Germany," Annales d'Economie et de Statistique, pp. 321–355.

Galdo, J., J. A. Smith, and D. Black (2008): "Bandwidth Selection and the Estimation of Treatment Effects with Unbalanced Data," Annals of Economics and Statistics / Annales d'Économie et de Statistique, 91/92, 189–216.

Gerfin, M., M. Lechner, and H. Steiger (2005): "Does subsidised temporary employment get the unemployed back to work? Aneconometric analysis of two different schemes," Labour Economics, 12(6), 807–835.

Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998): "Characterizing selection bias using experimental data," Econometrica, 66(5), 1017–1098.

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999): "The economics and econometrics of active labor market programs," Handbook of Labor Economics, 3, 1865–2097.

Heckman, J. J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," Journal of Labor Economics, 24(3), 411–482.

Heineck, G., and S. Anger (2010): "The returns to cognitive abilities and personality traits in Germany," Labour Economics, 17(3), 535–546.

Holzer, H. J. (1988): "Search Method Use by Unemployed Youth," Journal of Labor Economics, 6(1), 1–20.

Huber, M., M. Lechner, and C. Wunsch (2013): "The performance of estimators based on the propensity score," Journal of Econometrics.

Ichino, A., F. Mealli, and T. Nannicini (2008): "From temporary help jobs to permanent employment: What can we learn from matching estimators and their sensitivity?," Journal of Applied Econometrics, 23(3), 305–327.

Imbens, G. W. (2003): "Sensitivity to exogeneity assumptions in program evaluation," American Economic Review, 93(2), 126–132.

Imbens, G. W., and J. M. Wooldridge (2009): "Recent Developments in the Econometrics of Program Evaluation," Journal of Economic Literature, 47(1), 5–86.

Jaenichen, U., and G. Stephan (2011): "The effectiveness of targeted wage subsidies for hard-to-place workers," Applied Economics, 43(10), 1209–1225.

Jespersen, S. T., J. R. Munch, and L. Skipper (2008): "Costs and benefits of Danish active labour market programmes," Labour Economics, 15(5), 859–884.

Kluve, J. (2010): "The effectiveness of European active labor market programs," Labour Economics, 17(6), 904–918.

LaLonde, R. J. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," American Economic Review, 76(4), 604–620.

Lechner, M., R. Miquel, and C. Wunsch (2011): "Long-Run effects of public sector Sponsored Training in West Germany," Journal of the European Economic Association, 9(4), 742–784.

Lechner, M., and C. Wunsch (2008): "What did all the money do? On the general ineffectiveness of recent West German labour market programmes," Kyklos, 61(1), 134–174.

——— (2013): "Sensitivity of matching-based program evaluations to the availability of control variables," Labour Economics.

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," Frontiers in Econometrics, pp. 105–142.

McGee, A. (2014): "How the Perception of Control Influences Unemployed Job Search," Industrial and Labor Relations Review, forthcoming.

McKelvey, R. D., and W. Zavoina (1975): "A statistical model for the analysis of ordinal level dependent variables," Journal of Mathematical Sociology, 4(1), 103–120.

Montgomery, J. D. (1991): "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis," American Economic Review, 81(5), 1407–18.

Mortensen, D. T. (1986): "Job search and labor market analysis," in Handbook of Labor Economics, ed. by O. Ashenfelter, and R. Layard, vol. 2, pp. 849 – 919. Elsevier.

Mueller, G., and E. Plug (2006): "Estimating the effect of personality on male and female earnings," Industrial and Labor Relations Review, pp. 3–22.

Mueser, P. R., K. R. Troske, and A. Gorislavsky (2007): "Using state administrative data to measure program performance," Review of Economics and Statistics, 89(4), 761–783.

Nyhus, E. K., and E. Pons (2005): "The effects of personality on earnings," Journal of Economic Psychology, 26(3), 363–384.

Osborne Groves, M. (2005): "How important is your personality? Labor market returns to personality for women in the US and UK," Journal of Economic Psychology, 26(6), 827–841.

Oster, E. (2013): "Unobservable Selection and Coefficient Stability: Theory and Validation," Working Paper 19054, National Bureau of Economic Research (NBER).

Pannenberg, M. (2010): "Risk attitudes and reservation wages of unemployed workers: evidence from panel data," Economics Letters, 106(3), 223–226.

Rosenbaum, P., and D. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," Biometrika, 70(1), 41–50.

Rosenbaum, P. R. (2002): Observational Studies. Springer, New York.

Rotter, J. B. (1966): "Generalized expectancies for internal versus external control of reinforcement.," Psychological monographs: General and applied, 80(1), 1.

Roy, A. (1951): "Some Thoughts on the Distribution of Earnings," Oxford Economic Papers, 3(2), 135–145.

Rubin, D. (1974): "Estimating Causal Effects to Treatments in Randomised and Nonrandomised Studies," Journal of Educational Psychology, 66, 688–701.

Rubin, D. B. (2008): "For Objective Causal Inference, Design Trumps Analysis," The Annals of Applied Statistics, 2(3), 808–840.

Semykina, A., and S. J. Linz (2007): "Gender differences in personality and earnings: Evidence from Russia," Journal of Economic Psychology, 28(3), 387–410.

Shimer, R., and I. Werning (2007): "Reservation wages and unemployment insurance," Quarterly Journal of Economics, 122, 1145–1185.

Sianesi, B. (2004): "An evaluation of the Swedish system of active labor market programs in the 1990s," Review of Economics and Statistics, 86(1), 133–155.

Smith, J. A., and P. E. Todd (2005): "Does matching overcome LaLonde's critique of non-experimental estimators?," Journal of Econometrics, 125(1), 305–353.

van den Berg, G., A. Bergemann, and M. Caliendo (2009): "The Effect of Active Labor Market Programs on Not-Yet Treated Unemployed Individuals," Journal of the European Economic Association, 7(2-3), 606–616.

Van den Berg, G. J., and B. Van der Klaauw (2006): "Counseling and monitoring of unemployed workers: Theory and evidence from a controlled social experiment," International Economic Review, 47(3), 895–936.

Van Ours, J. C. (2004): "The locking-in effect of subsidized jobs," Journal of Comparative Economics, 32(1), 37–55.

Wolff, J., and E. Jozwiak (2007): "Does short-term training activate means-tested unemployment benefit recipients in Germany?," Discussion Paper 29/2007, IAB, Nuremberg.

Tables and Figures

Table 1: Overview - Control Variables

Category	Control variables	Specification		
		Base 1	Base 2	Full
1) Socio-demographic/ baseline variables	*Socio-demographics*			
	Gender			
	Educational level			
	Marital status			
	German citizenship			
	Migration background			
	Number of children	✓	✓	✓
	Health problems			
	Searching for full- or part-time employment			
	Employment status of partner			
	Month of entry into unemployment			
	Time between entry into unemployment and interview			
	Regional information			
	Living in West-Germany			
	Local unemployment rate			
2) Labor market history	*Short-term labor market history*			
	Employment status before entry into unemployment			
	Last daily income			
	Last job was full-time employment			
	Reason for termination of last employment			
	Months in .... in last 6 months / 24 months			
	- employment			
	- unemployment			
	- out of labor force			
	Number of ... in last 24 months			
	- employers			
	- program participations			
	- unemployment spells		✓	✓
	- out of labor force spells			
	Long-term labor market history			
	Months in ... in last 10 years			
	- employment			
	- unemployment			
	- program			
	- out of labor force			
	Number of ... in last 10 years			
	- employers			
	- program participations			
	- unemployment spells			
	- out of labor force spells			
	Time with last employer			
	Duration of last unemployment spell			
3) Personality traits	Openness			
	Conscientiousness			
	Extraversion			✓
	Neuroticism			
	Locus of control			
4) Job search and employment outlook	Reservation wage			
	Search intensity(number of applications)			
	Number of search channels used			✓
	Subjective (overall) probability of treatment			
	Expected probability to find a job in the next 6 months			
5) Socio-cultural characteristics	Number of good friends outside the family			
	Problems with child care			✓
	Father has A-level qualification			
	Life satisfaction			

Table 2: Entries in ALMP Programs in Germany (in 1,000)

	2005	2007	2009	2011
Entries into unemployment	8,427	8,155	9,253	8,218
Entries into ALMP programs				
Short-term training	901	1,092	1,194	1,201
Long-term training	132	365	618	305
Wage subsidies	134	266	266	187
Job creation schemes	78	70	11	1
Start-up subsidies	91	126	137	137

Note: Since 2009, the former mobility assistances are included into the new intermediation budget.
Source: Statistic of the German Federal Employment Agency

Table 3: Descriptive Statistics: Labor Market Outcomes and Baseline Variables

	NP	STT	LTT	WS	Any
No. of observations	3,769	1,607	694	501	3,092
Labor market outcomes					
Regular employed					
after 12 months	0.54	***0.43***	***0.34***	***0.40***	***0.40***
after 30 months	0.59	0.57	0.57	***0.69***	***0.55***
Months employed within 30 months	14.54	***12.48***	***10.96***	**13.25**	***11.54***
Cumulated earnings within 30 months	26,818	***22,382***	***21,633***	***29,955***	***21,586***
Socio-demographic characteristics					
Female	0.48	**0.52**	**0.53**	0.47	***0.53***
Age in years	35.18	35.07	***37.29***	***38.00***	***36.14***
Married (or cohabiting)	0.38	0.39	*0.42*	0.41	**0.41**
German citizenship	0.95	0.94	0.95	0.95	0.95
West Germany	0.66	*0.69*	**0.71**	**0.60**	**0.70**
Migration background (1=yes)	0.14	0.14	0.15	0.13	*0.15*
Children					
No children	0.68	0.68	**0.64**	0.69	**0.65**
One child	0.19	0.19	0.20	0.18	0.19
Two (or more) children	0.13	0.13	*0.16*	0.13	**0.15**
School leaving degree					
None, special needs, other	0.02	0.03	0.02	0.03	0.02
Lower secondary school	0.28	***0.34***	0.27	0.30	***0.32***
Middle secondary school	0.41	***0.46***	0.43	**0.46**	**0.44**
Specialized upper secondary school	0.28	***0.18***	0.29	***0.21***	***0.22***
Vocational training					
None	0.09	0.09	0.08	*0.07*	0.10
Internal or external professional training, others	0.69	***0.78***	0.70	***0.76***	***0.73***
Technical college or university degree	0.22	***0.13***	0.22	**0.17**	***0.17***
Local UE rate at interview	0.09	**0.09**	0.09	***0.10***	***0.09***
Health restriction or disability	0.08	0.08	0.07	**0.05**	0.08
Labour market history					
Employment status before unemployment					
Employed	0.66	0.66	**0.70**	**0.71**	0.66
Subsidized employment	0.07	0.07	0.06	***0.10***	0.07
School, apprentice, military, etc.	0.16	0.15	***0.10***	*0.12*	**0.14**
Maternity leave	0.04	*0.05*	***0.07***	0.03	***0.06***
Other	0.08	0.07	0.07	***0.04***	0.07
Months employed in last 6 months	4.77	**4.95**	4.91	**5.09**	4.88
Last daily income	47.93	***44.61***	**51.22**	46.21	*46.17*
Last job was full-time employment	0.95	*0.94*	0.95	0.95	0.95
Months employed in last 2 years	15.14	***16.27***	*15.77*	***16.40***	***15.82***
No. of employers in last 2 years	1.67	1.65	1.62	1.67	1.63
Months employed in last 10 years	49.56	***52.75***	***54.00***	***55.26***	***52.46***
No. of employers in last 10 years	3.60	3.55	3.70	***4.02***	3.61

Notes: All numbers denote shares unless otherwise indicated, measured at the entry into unemployment. Italic (bold) (italic and bold) numbers indicate statistically significant differences between each group of participants and non-participants at the 10% (5%) (1%)-level based on a two-tailed t-test on equal means. NP - non-participants; STT - participants in short-term training; LTT - participants in long-term training; WS - recipients of wage subsidies; Any - participants in any ALMP.

Table 4: Descriptive Statistics: Usually Unobserved Variables

	NP	STT	LTT	WS	Any
No. of observations	3,769	1,607	694	501	3,092
Personality traits					
Openness	5.06	5.01	4.98	5.11	**4.99**
Conscientiousness	6.27	6.26	6.31	*6.35*	6.27
Extraversion	5.23	*5.17*	***5.07***	5.21	***5.14***
Neuroticism	3.74	**3.82**	3.75	3.68	3.78
Locus of control	5.06	***4.96***	5.02	5.01	***4.99***
Job search and employment outlook variables					
Reservation wage (in Euro)	7.24	***6.61***	7.19	**6.78**	***6.88***
Number of own applications (mean)	15.07	**16.49**	***19.46***	***19.12***	***17.02***
Number of search channels	4.95	***5.31***	***5.38***	***5.36***	***5.30***
Subjective (overall) probability of treatment					
low	0.44	***0.29***	***0.25***	***0.36***	***0.31***
middle	0.25	**0.27**	***0.19***	0.25	0.26
high	0.31	***0.43***	***0.55***	***0.39***	***0.44***
Expected probability to find a job in the next 6 months					
improbable	0.11	0.10	**0.14**	***0.07***	0.12
probable	0.30	***0.38***	***0.41***	*0.34*	***0.38***
very probable	0.49	***0.43***	***0.36***	0.49	***0.40***
Socio-cultural characteristics					
Number of good friends outside the family	4.83	4.84	4.78	5.12	4.81
Problems with childcare (0: n/a, 1: None, 6: Very)	0.64	0.65	**0.75**	0.60	**0.71**
Father has A-level qualifications (1=yes)	0.16	***0.11***	0.15	**0.12**	***0.13***
Life satisfaction					
low	0.10	0.11	0.10	0.12	0.11
middle	0.35	**0.38**	*0.38*	**0.40**	**0.38**
high	0.55	***0.50***	*0.52*	**0.49**	***0.51***

Notes: All numbers denote shares unless otherwise indicated, measured at the entry into unemployment. Personality traits are measured with different items on a 7-Point Likert-Scale. Italic (bold) (italic and bold) numbers indicate statistically significant differences between each group of participants and non-participants at the 10% (5%) (1%)-level based on a two-tailed t-test on equal means. NP - non-participants; STT - participants in short-term training; LTT - participants in long-term training; WS - recipients of wage subsidies; Any - participants in any ALMP.

Table 5: Logit Estimates - Participation Active Labor Market Policy Programs

	Short-term training			Long-term training			Wage subsidies		
	Base 1	Base 2	Full	Base 1	Base 2	Full	Base 1	Base 2	Full
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Personality traits									
Openness (standardized)			-.003			-.006			0.008
Conscientiousness (standardized)			-.002			0.008			0.004
Extraversion (standardized)			-.009			-.022***			-.006
Neuroticism (standardized)			-.006			-.003			-.008
Locus of control (standardized)			-.016**			0.003			-.001
Job search and employment outlook variables									
Ln(Reservation wage)			-.062***			0.004			-.029*
Number of own applications (mean)			0.0002			0.0006***			0.0004**
Search channel(s) (Number of)			0.02***			0.012***			0.009***
Subjective probability of treatment									
middle									
high			0.058***			0.116***			0.022
low			-.063***			-.033***			-.013
Expected probability to find a job									
probable									
very probable			-.030**			-.064***			0.009
improbable			-.029			0.032			-.053***
Socio-cultural characteristics									
Number of good friends			0.001			0.001			0.0009
Problems with child care			0.0002			-.002			-.003
Father: A-level qualifications (1=yes)			-.023			-.007			-.020
Life satisfaction									
middle									
high			-.026*			-.014			-.020*
low			0.0007			-.016			0.007
Socio-demographic characteristics									
Female	0.04***	0.025*	0.017	0.027**	0.019	0.017	0.005	-.003	0.0002
Age 45-55 years	-.005	-.016	-.013	0.087***	0.093***	0.077***	0.094***	0.099***	0.103***
Married (or cohabiting)	0.027	0.022	0.02	-.005	-.014	-.017	-.012	-.015	-.013
Migration background (1=yes)	-.014	-.010	-.018	0.027	0.03	0.016	0.014	0.019	0.025
Two (or more) children	-.006	-.008	-.008	0.032*	0.021	0.021	0.027	0.024	0.033
(Specialized) upper secondary school	-.109***	-.102***	-.079**	0.033	0.033	0.048	-.071***	-.056*	-.050
Technical college or university degree	-.043*	-.041	-.023	-.012	-.021	-.023	0.002	0.005	0.01
Health restriction or disability	-.018	-.021	-.028	-.026	-.017	-.021	-.055***	-.050***	-.052***
Searching for full-time employment	-.004	-.003	0.004	0.017	0.019	0.026	0.027*	0.03*	0.028*
Partner is full-time employed	-.026	-.027	-.021	-.009	-.008	-.005	0.012	0.016	0.015
Short-term labour market history									
Employment status before unemployment									
Subsidized employment		0.018	0.026		-.029	-.032*		0.037*	0.035*
Maternity leave		0.077**	0.098**		0.144***	0.127***		0.015	0.051
Last daily income		-.0005**	-.0002		-.00007	0.00008		-.0006***	-.0006***
Last job was full-time employment		-.056**	-.057**		0.015	0.01		-.002	0.004
Months employed in last 6 months		-.015***	-.014**		-.006	-.004		-.008*	-.009**
Months employed in last 2 years		0.007***	0.006***		0.004**	0.003*		0.002	0.002
No. of employers in last 2 years		-.006	-.007		-.013*	-.012*		-.019***	-.021***
No. of unemployment spells in last 2 years		-.034***	-.031**		0.019*	0.02*		0.008	0.01
No. of programs in last 2 years		0.029**	0.025*		0.043***	0.039***		0.026***	0.024**
Long-term labour market history									
Months employed in last 10 years		0.0004	0.0006		-.0005	-.0004		0.0007	0.0008
No. of employers in last 10 years		-.0006	-.001		0.002	0.002		0.01***	0.01***
No. of unemployment spells in last 10 years		-.008	-.007		-.011***	-.009*		-.010**	-.011**
No. of programs in last 10 years		0.019**	0.015*		-.001	-.006		0.009	0.009
Obs.	5,376	5,376	5,376	4,463	4,463	4,463	4,270	4,270	4,270
log-Likelihood	-3208.80	-3141.05	-3065.67	-1876.79	-1830.22	-1714.90	-1487.20	-1429.19	-1399.56
P-value joint significance									
Personality traits			0.061			0.004			0.421
Search variables			0.000			0.000			0.000
Socio-cultural variables			0.282			0.767			0.162
Personality and search variables			0.000			0.000			0.000
Personality and socio-cultural variables			0.048			0.038			0.290
All unobservables			0.000			0.000			0.000
Mean Value	0.299	0.299	0.299	0.156	0.156	0.156	0.117	0.117	0.117
Hitrate	0.543	0.586	0.622	0.57	0.611	0.665	0.611	0.644	0.658

Notes: Depicted are average marginal effects for selected control variables. In all estimations we control for additional information on socio-demographics, short- and long-term labour market history (except for column 1, 4 and 7), the regional unemployment rate, the month of entry into unemployment and the time between the entry into unemployment and the interview. Full estimation results can be found in the appendix. ***/**/* indicate statistically significance at the 1%/5%/10%-level.

Table 6: Consequences for Estimation Quality: Goodness of Fit

Specification	Base 1	Base 2	Full
	Short-term training		
McFadden's R^2	0.021	0.042	0.065
% of full model R^2	32.9	64.7	100
Difference in %-points	32.9	31.8	35.3
McKelvey and Zavoina's R^2	0.041	0.078	0.12
% of full model R^2	33.8	64.7	100
Difference in %-points	33.8	30.9	35.3
Effron's R^2	0.025	0.050	0.076
% of full model R^2	33.2	65.8	100
Difference in %-points	33.2	32.6	34.2
	Long-term training		
McFadden's R^2	0.027	0.051	0.11
% of full model R^2	24.2	46.0	100
Difference in %-points	24.2	21.8	54.0
McKelvey and Zavoina's R^2	0.056	0.099	0.20
% of full model R^2	27.7	48.8	100
Difference in %-points	27.7	21.1	51.2
Effron's R^2	0.025	0.049	0.11
% of full model R^2	23.0	44.0	100
Difference in %-points	23.0	21.0	56.0
	Wage subsidies		
McFadden's R^2	0.037	0.074	0.093
% of full model R^2	39.3	79.5	100
Difference in %-points	39.3	40.2	20.5
McKelvey and Zavoina's R^2	0.074	0.15	0.19
% of full model R^2	39.6	80.0	100
Difference in %-points	39.6	40.4	20.0
Effron's R^2	0.027	0.062	0.080
% of full model R^2	33.6	77.6	100
Difference in %-points	33.6	44.0	22.4

Note: Depicted are three different goodness of fit measures comparing the three specifications of interest. Base 1 - Baseline specification including only socio-demographics, Base 2 - Baseline specification including socio-demographics and labor market history, Full - Full specification including socio-demographics, labor market history and all usually unobserved variables.

Table 7: Consequences for Propensity Scores and Ranks: Correlation Coefficients

	Base 1 vs. Full			Base 2 vs. Full		
	Full sample	Participants	Non-participants	Full sample	Participants	Non-participants
			Short-term training			
Propensity score correlation						
Pearson's r	0.57	0.50	0.58	0.81	0.79	0.80
	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}
Rank correlation						
Spearman's rho	0.57	0.49	0.58	0.80	0.77	0.78
	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}
Concordance						
Kendall's tau	0.40	0.34	0.41	0.60	0.58	0.59
	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}
Distribution comparison						
Paired t-test	0.00	-.036	0.015	0.00	-.019	0.008
	{1.00}	{0.00}	{0.00}	{1.00}	{0.00}	{0.00}
Wilcoxon signed-ranks test	3.29	-11.79	12.12	1.06	-9.38	7.80
	{0.00}	{0.00}	{0.00}	{0.29}	{0.00}	{0.00}
Rank distribution comparison						
Friedman test	33.00	80.20	171.94	10.12	57.13	75.94
	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}
			Long-term training			
Propensity score correlation						
Pearson's r	0.47	0.47	0.44	0.66	0.68	0.63
	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}
Rank correlation						
Spearman's rho	0.50	0.49	0.47	0.68	0.68	0.66
	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}
Concordance						
Kendall's tau	0.34	0.34	0.33	0.49	0.49	0.47
	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}
Distribution comparison						
Paired t-test	0.00	-.070	0.013	0.00	-.050	0.009
	{1.00}	{0.00}	{0.00}	{1.00}	{0.00}	{0.00}
Wilcoxon signed-ranks test	7.50	-11.61	14.35	6.41	-10.46	12.59
	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}
Rank distribution comparison						
Friedman test	181.90	47.73	311.19	160.75	54.23	287.52
	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}
			Wage subsidies			
Propensity score correlation						
Pearson's r	0.60	0.64	0.58	0.88	0.90	0.87
	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}
Rank correlation						
Spearman's rho	0.60	0.67	0.58	0.88	0.89	0.87
	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}
Concordance						
Kendall's tau	0.43	0.48	0.41	0.70	0.72	0.69
	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}
Distribution comparison						
Paired t-test	0.00	-.045	0.006	0.00	-.015	0.002
	{1.00}	{0.00}	{0.00}	{1.00}	{0.00}	{0.00}
Wilcoxon signed-ranks test	7.26	-8.68	11.41	3.08	-6.09	5.88
	{0.00}	{0.00}	{0.00}	{0.29}	{0.00}	{0.00}
Rank distribution comparison						
Friedman test	128.24	29.22	196.69	20.52	30.20	46.58
	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}

Note: Depicted are correlation coefficients and distribution comparison test between specification Base 1 (Base 2) and Full. Base 1 - Baseline specification including only socio-demographics, Base2 - Baseline specification including socio-demographics and labor market history, Full - Full specification including socio-demographics, labor market history and all usually unobserved variables. P-values are shown in braces based two-tailed t-tests on zero correlation, respectively equal distributions.

Table 8: Consequences for the Matching Quality: Mean Standardized Bias (MSB)

	Raw	Base 1	Base 2	Full
		Short-term training		
Socio-demographics	4.55	2.10	2.56	2.00
Labor market history	6.77	5.50	1.99	1.51
Personality traits	6.00	4.68	5.72	3.25
Job search and employment outlook	16.5	14.6	13.8	1.86
Socio-cultural characteristics	5.70	2.20	4.16	1.75
Overall	6.33	4.27	3.57	1.90
		Long-term training		
Socio-demographics	4.44	2.91	2.83	3.29
Labor market history	9.06	7.38	2.96	2.02
Personality traits	6.48	5.90	9.54	1.70
Job search and employment outlook	24.3	23.6	24.6	5.16
Socio-cultural characteristics	3.92	2.27	6.47	3.73
Overall	7.51	6.05	5.21	3.00
		Wage subsidies		
Socio-demographics	6.60	3.43	3.89	3.74
Labor market history	14.1	9.22	2.58	2.73
Personality traits	5.15	4.32	4.22	6.20
Job search and employment outlook	15.0	14.2	15.5	5.09
Socio-cultural characteristics	7.69	4.95	6.52	4.58
Overall	9.49	6.17	4.60	3.74

Note: Depicted is the mean standardized bias separated for each group of control variables and for all control variables in total measured before matching (Raw), after matching on the specification Base 1, Base 2 and Full. Base 1 - Baseline specification including only socio-demographics, Base 2 - Baseline specification including socio-demographics and labor market history, Full - Full specification including socio-demographics, labor market history and all usually unobserved variables.

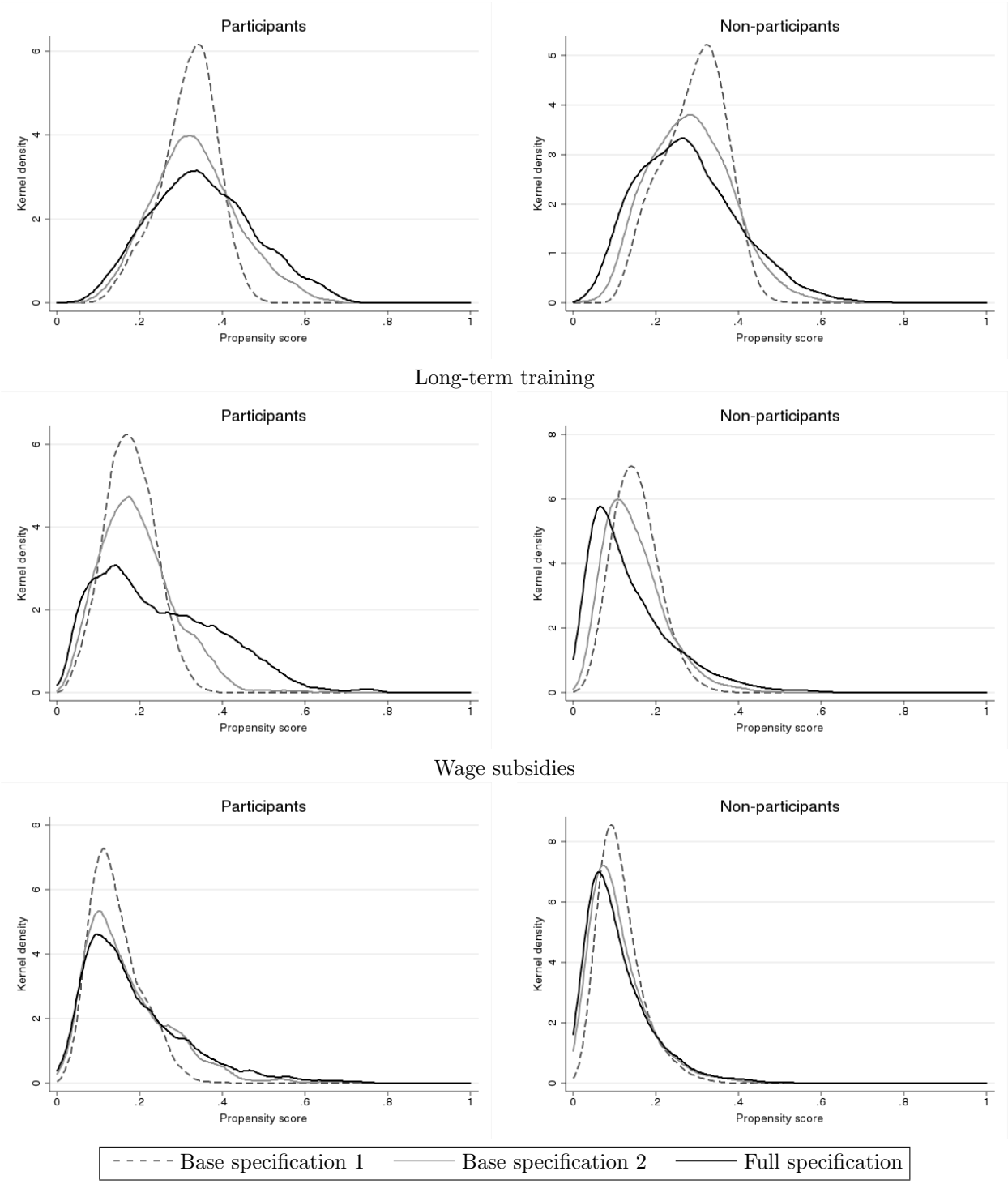
Table 9: Consequences for the Average Treatment Effects on the Treated (ATT)

Specification	Treatment effects				Differences			
	Raw gap	Base 1	Base 2	Full	Base 1 vs. Full		Base 2 vs. Full	
					absolute	in %	absolute	in %
			Short-term training					
Employed after 12 months	-.105***	-.101***	-.095***	-.096***	0.005	4.5	-.001	-1.4
	(0.015)	(0.014)	(0.014)	(0.014)	(0.006)		(0.005)	
Employed after 30 months	-.017	-.009	-.009	-.011	-.002	-21.1	-.002	-17.8
	(0.015)	(0.016)	(0.017)	(0.017)	(0.006)		(0.004)	
Months employed within 30 months	-2.060***	-1.853***	-1.651***	-1.708***	0.145	8.1	-.056	-3.6
	(0.304)	(0.287)	(0.283)	(0.279)	(0.130)		(0.110)	
Cumulated earnings within 30 months	-4,439***	-2,724***	-2,051***	-1,966***	758***	27.8	85	4.1
	(727)	(585)	(572)	(558)	(270)		(212)	
Obs.	5,376	5,376	5,376	5,376				
			Long-term training					
Employed after 12 months	-.193***	-.201***	-.210***	-.210***	-0.009	-4.4	0.0004	1.9
	(0.02)	(0.02)	(0.02)	(0.022)	(0.012)		(0.010)	
Employed after 30 months	-.022	-.032	-.043**	-.040*	-0.008	-24.1	0.003	7.4
	(0.020)	(0.021)	(0.022)	(0.024)	(0.012)		(0.011)	
Months employed within 30 months	-3.580***	-3.726***	-3.948***	-3.757***	-0.031	-0.8	0.191	4.7
	(0.427)	(0.378)	(0.386)	(0.441)	(0.270)		(0.220)	
Cumulated earnings within 30 months	-5,185***	-5,789***	-6,227***	-6,053***	-265	-4.6	173	2.8
	(1,058)	(832)	(823)	(989)	(630)		(561)	
Obs.	4,463	4,463	4,463	4,463				
			Wage subsidies					
Employed after 12 months	-.142***	-.149***	-.157***	-.167***	-0.018*	-12.1	-0.011	-7.0
	(0.024)	(0.023)	(0.024)	(0.025)	(0.010)		(0.009)	
Employed after 30 months	0.100***	0.093***	0.082***	0.07***	-0.023**	-24.7	-0.012	-14.6
	(0.023)	(0.023)	(0.023)	(0.025)	(0.009)		(0.009)	
Months employed within 30 months	-1.292***	-1.381***	-1.527***	-1.805***	-0.420**	-30.7	-0.280	-18.2
	(0.488)	(0.381)	(0.393)	(0.442)	(0.210)		(0.200)	
Cumulated earnings within 30 months	3,137**	3,368***	3,616***	3,049***	-319	-9.5	-567	-15.7
	(1,218)	(904)	(897)	(996)	(538)		(462)	
Obs.	4,270	4,270	4,270	4,270				

Note: Depicted are estimated average treatment effects on the treated as the difference in mean outcomes between participants and matched non-participants using epanechnikov kernel propensity score matching with bandwidth 0.06, as well as the difference between base specification 1 (including only socio-demographics) and the full specification, as well as the difference between base specification 2 (including socio-demographics and the labor market history) and the full specification, for estimated average treatment effects. Standard errors are in parentheses and based on bootstrapping with 251 replications. ***/**/* indicate statistically significance at the 1%/5%/10%-level.

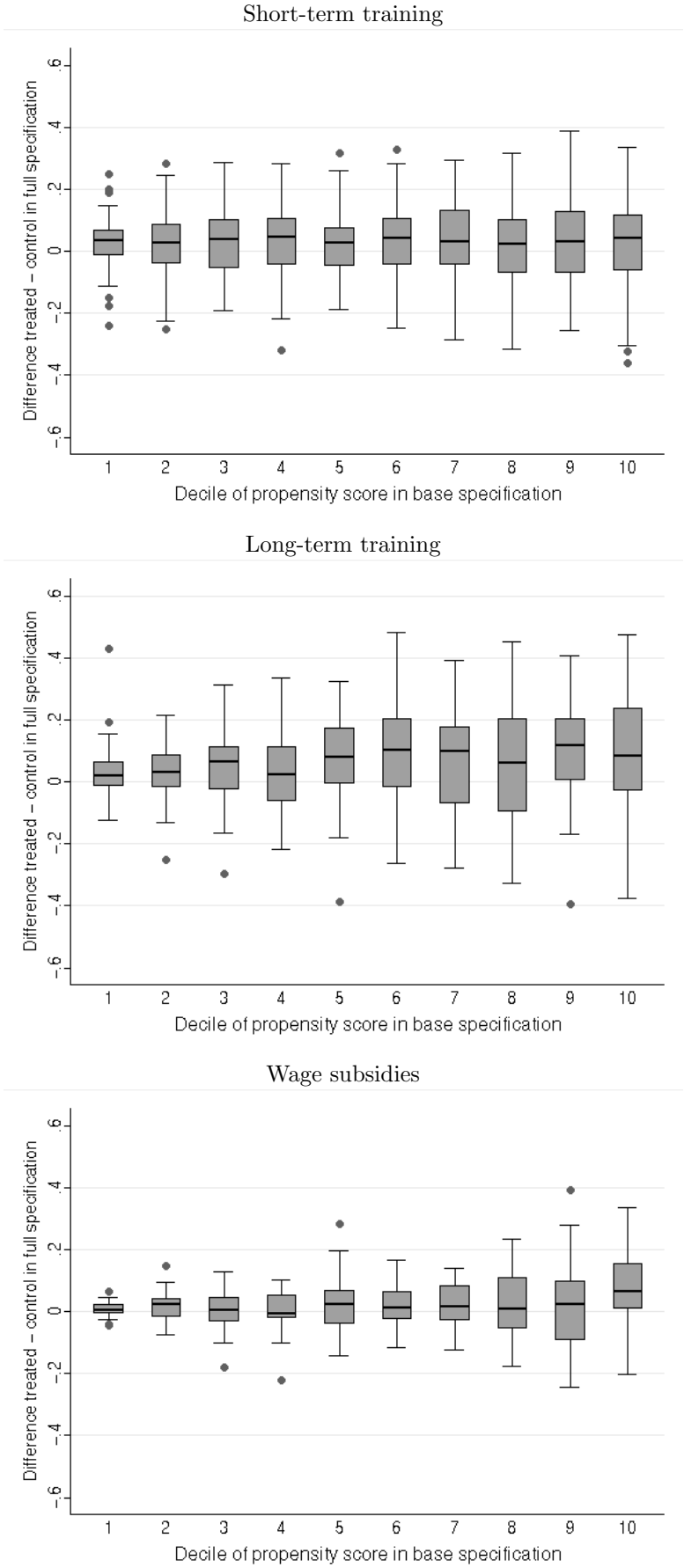
Figure 1: Propensity Score Distribution

Short-term training



Note: Depicted are epanechnikov kernel densities (bandwith=0.06) of the propensity score after matching on three different specifications. Base 1 - Baseline specification including only socio-demographics, Base 2 - Baseline specification including socio-demographics and labor market history, Full - Full specification including socio-demographics, labor market history and all usually unobserved variables.

Figure 2: Matched Propensity Score Differences Between Treated and Controls - Base 2 vs. Full Specification



Note: Depicted is the distribution of the propensity score difference between participants and matched non-participants when including all usually unobserved variables using epanechnikov kernel propensity score matching with bandwidth 0.06 for each decile of the propensity score distribution in specification Base 2 (including socio-demographics and labor market histories).

A Appendix - Supplementary Tables and Figures

A.1 Sequential Inclusion of Usually Unobserved Variables

Table A.1: Consequences for Average Treatment Effects on the Treated: Sequential Inclusion of Usually Unobserved Variables

	Raw gap	(1)	(2)	(3)	(4)	(5)	(6)
				Short-term training			
Employed after 12 months	-.105***	-.101***	-.095***	-.095***	-.096***	-.094***	-.096***
	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Employed after 30 months	-.017	-.009	-.009	-.008	-.013	-.007	-.011
	(0.015)	(0.016)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Months employed within 30 months	-2.060***	-1.853***	-1.651***	-1.648***	-1.747***	-1.608***	-1.708***
	(0.304)	(0.287)	(0.283)	(0.28)	(0.276)	(0.288)	(0.279)
Cumulated earnings within 30 months	-4,439***	-2,724***	-2,051***	-2,010***	-2,026***	-1,964***	-1,966***
	(727)	(585)	(572)	(572)	(551)	(579)	(558)
Obs.	5,376	5,376	5,376	5,376	5,376	5,376	5,376
				Long-term training			
Employed after 12 months	-.193***	-.201***	-.210***	-.210***	-.208***	-.210***	-.210***
	(0.02)	(0.02)	(0.02)	(0.021)	(0.022)	(0.021)	(0.022)
Employed after 30 months	-.022	-.032	-.043**	-.041*	-.043*	-.041*	-.040*
	(0.020)	(0.021)	(0.022)	(0.022)	(0.024)	(0.022)	(0.024)
Months employed within 30 months	-3.580***	-3.726***	-3.948***	-3.942***	-3.745***	-3.918***	-3.757***
	(0.427)	(0.378)	(0.386)	(0.389)	(0.433)	(0.393)	(0.441)
Cumulated earnings within 30 months	-5,185***	-5,789***	-6,227***	-6,166***	-5,883***	-6,139***	-6,053***
	(1,058)	(832)	(823)	(841)	(955)	(823)	(989)
Obs.	4,463	4,463	4,463	4,463	4,463	4,463	4,463
				Wage subsidies			
Employed after 12 months	-.142***	-.149***	-.157***	-.155***	-.169***	-.156***	-.167***
	(0.024)	(0.023)	(0.024)	(0.024)	(0.024)	(0.024)	(0.025)
Employed after 30 months	0.100***	0.093***	0.082***	0.082***	0.071***	0.084***	0.07***
	(0.023)	(0.023)	(0.023)	(0.024)	(0.025)	(0.024)	(0.025)
Months employed within 30 months	-1.292***	-1.381***	-1.527***	-1.490***	-1.820***	-1.512***	-1.805***
	(0.488)	(0.381)	(0.393)	(0.405)	(0.413)	(0.411)	(0.442)
Cumulated earnings within 30 months	3,137**	3,368***	3,616***	3,682***	3,094***	3,729***	3,049***
	(1,218)	(904)	(897)	(909)	(975)	(906)	(996)
Obs.	4,270	4,270	4,270	4,270	4,270	4,270	4,270
Control variables							
Socio-demographic variables		✓	✓	✓	✓	✓	✓
Labor market history			✓	✓	✓	✓	✓
Personality traits				✓			✓
Job search and employment outlook					✓		✓
Socio-cultural characteristics						✓	✓

Note: Depicted are estimated average treatment effects on the treated as the difference in mean outcomes between participants and matched non-participants using epanechnikov kernel propensity score matching with bandwith 0.06. Standard errors are in parentheses and based on bootstrapping with 251 replications. ***/**/* indicate statistically significance at the 1%/5%/10%-level.

A.2 Different Matching/Weighting Estimators

Table A.2: Average Treatment Effects on Treated for Different Matching/Weighting Estimators - Short-term Training

	Raw gap	Base 1	Base 2	Full
Employed after 12 months				
IPW	-.105***	-.102***	-.093***	-.097***
	(0.015)	(0.014)	(0.014)	(0.014)
Kernel 1 (bw=0.02)		-.102***	-.093***	-.094***
		(0.015)	(0.015)	(0.015)
Kernel 2 (bw=0.06)		-.101***	-.095***	-.096***
		(0.014)	(0.014)	(0.014)
Kernel 3 (bw=0.2)		-.100***	-.102***	-.099***
		(0.015)	(0.014)	(0.014)
Employed after 30 months				
IPW	-.017	-.009	-.007	-.011
	(0.015)	(0.016)	(0.017)	(0.017)
Kernel 1 (bw=0.02)		-.011	-.009	-.007
		(0.016)	(0.017)	(0.017)
Kernel 2 (bw=0.06)		-.009	-.009	-.011
		(0.016)	(0.017)	(0.017)
Kernel 3 (bw=0.2)		-.011	-.014	-.015
		(0.016)	(0.016)	(0.016)
Months employed within 30 months				
IPW	-2.060***	-1.889***	-1.580***	-1.692***
	(0.304)	(0.288)	(0.282)	(0.279)
Kernel 1 (bw=0.02)		-1.892***	-1.594***	-1.633***
		(0.293)	(0.291)	(0.287)
Kernel 2 (bw=0.06)		-1.853***	-1.651***	-1.708***
		(0.287)	(0.283)	(0.279)
Kernel 3 (bw=0.2)		-1.897***	-1.907***	-1.848***
		(0.285)	(0.274)	(0.268)
Cumulated earnings within 30 months				
IPW	-4,439***	-2,793***	-1,966***	-1,981***
	(727)	(586)	(574)	(567)
Kernel 1 (bw=0.02)		-2,778***	-1,890***	-1,825***
		(595)	(582)	(576)
Kernel 2 (bw=0.06)		-2,724***	-2,051***	-1,966***
		(585)	(572)	(558)
Kernel 3 (bw=0.2)		-3,250***	-2,999***	-2,627***
		(594)	(562)	(537)
Obs.	5,376	5,376	5,376	5,376

Note: Depicted are estimated average treatment effects on the treated as the difference in mean outcomes between participants and matched non-participants using inverse probability weighting (IPW) and 3 types of epanechnikov kernel propensity score matching. Standard errors are in parentheses and based on bootstrapping with 251 replications. ***/**/* indicate statistically significance at the 1%/5%/10%-level.

Table A.3: Average Treatment Effects on Treated for Different Matching/Weighting Estimators - Long-term Training

	Raw gap	Base 1	Base 2	Full
Employed after 12 months				
IPW	-.193***	-.204***	-.211***	-.210***
	(0.02)	(0.02)	(0.02)	(0.022)
Kernel 1 (bw=0.02)		-.202***	-.207***	-.204***
		(0.02)	(0.021)	(0.023)
Kernel 2 (bw=0.06)		-.201***	-.210***	-.210***
		(0.02)	(0.02)	(0.022)
Kernel 3 (bw=0.2)		-.196***	-.202***	-.204***
		(0.02)	(0.02)	(0.021)
Employed after 30 months				
IPW	-.022	-.035*	-.042**	-.045*
	(0.020)	(0.021)	(0.021)	(0.023)
Kernel 1 (bw=0.02)		-.030	-.043*	-.043*
		(0.021)	(0.022)	(0.024)
Kernel 2 (bw=0.06)		-.032	-.043**	-.040*
		(0.021)	(0.022)	(0.024)
Kernel 3 (bw=0.2)		-.025	-.032	-.036
		(0.021)	(0.021)	(0.022)
Months employed within 30 months				
IPW	-3.580***	-3.787***	-3.942***	-3.869***
	(0.427)	(0.38)	(0.382)	(0.423)
Kernel 1 (bw=0.02)		-3.687***	-3.863***	-3.772***
		(0.382)	(0.4)	(0.447)
Kernel 2 (bw=0.06)		-3.726***	-3.948***	-3.757***
		(0.378)	(0.386)	(0.441)
Kernel 3 (bw=0.2)		-3.647***	-3.772***	-3.746***
		(0.362)	(0.363)	(0.398)
Cumulated earnings within 30 months				
IPW	-5,185***	-5,965***	-6,269***	-6,035***
	(1058)	(833)	(796)	(940)
Kernel 1 (bw=0.02)		-5,810***	-6,143***	-5,928***
		(846)	(879)	(1019)
Kernel 2 (bw=0.06)		-5,789***	-6,227***	-6,053***
		(832)	(823)	(989)
Kernel 3 (bw=0.2)		-5,379***	-5,727***	-5,742***
		(832)	(794)	(859)
Obs.	4,463	4,463	4,463	4,463

Note: Depicted are estimated average treatment effects on the treated as the difference in mean outcomes between participants and matched non-participants using inverse probability weighting (IPW) and 3 types of epanechnikov kernel propensity score matching. Standard errors are in parentheses and based on bootstrapping with 251 replications. ***/**/* indicate statistically significance at the 1%/5%/10%-level.

Table A.4: Average Treatment Effects on Treated for Different Matching/Weighting Estimators - Wage Subsidies

	Raw gap	Base 1	Base 2	Full
Employed after 12 months				
IPW	-.142***	-.150***	-.159***	-.169***
	(0.024)	(0.023)	(0.024)	(0.024)
Kernel 1 (bw=0.02)		-.150***	-.159***	-.163***
		(0.023)	(0.024)	(0.025)
Kernel 2 (bw=0.06)		-.149***	-.157***	- -.167***
		(0.023)	(0.024)	(0.025)
Kernel 3 (bw=0.2)		-.144***	-.148***	-.155***
		(0.023)	(0.023)	(0.023)
Employed after 30 months				
IPW	0.100***	0.095***	0.081***	0.069***
	(0.023)	(0.023)	(0.024)	(0.024)
Kernel 1 (bw=0.02)		0.094***	0.081***	0.071***
		(0.023)	(0.025)	(0.025)
Kernel 2 (bw=0.06)		0.093***	0.082***	0.07***
		(0.023)	(0.023)	(0.025)
Kernel 3 (bw=0.2)		0.098***	0.094***	0.084***
		(0.022)	(0.023)	(0.023)
Months employed within 30 months				
IPW	-1.292***	-1.374***	-1.553***	-1.817***
	(0.488)	(0.389)	(0.4)	(0.414)
Kernel 1 (bw=0.02)		-1.370***	-1.554***	-1.710***
		(0.392)	(0.408)	(0.44)
Kernel 2 (bw=0.06)		-1.381***	-1.527***	-1.805***
		(0.381)	(0.393)	(0.442)
Kernel 3 (bw=0.2)		-1.316***	-1.387***	-1.546***
		(0.36)	(0.365)	(0.38)
Cumulated earnings within 30 months				
IPW	3,137**	3,485***	3,674***	3,131***
	(1,218)	(910)	(877)	(911)
Kernel 1 (bw=0.02)		3,475***	3,687***	3,293***
		(929)	(916)	(1004)
Kernel 2 (bw=0.06)		3,368***	3,616***	3,049***
		(904)	(897)	(996)
Kernel 3 (bw=0.2)		3,178***	3,399***	3,166***
		(934)	(878)	(908)
Obs.	4,270	4,270	4,270	4,270

Note: Depicted are estimated average treatment effects on the treated as the difference in mean outcomes between participants and matched non-participants using inverse probability weighting (IPW) and 3 types of epanechnikov kernel propensity score matching. Standard errors are in parentheses and based on bootstrapping with 251 replications. ***/**/* indicate statistically significance at the 1%/5%/10%-level.