

The Returns to Continuous Training in Germany: New Evidence from Propensity Score Matching Estimators

Grit Muehler^{*}, Michael Beckmann^{}, and Bernd Schauenberg^{***}**

Abstract: The present paper examines the wage effects of continuous training programs using individual-level data from the German Socio Economic Panel (GSOEP). In order to account for selectivity in training participation we estimate average treatment effects (ATE and ATT) of general and firm-specific continuous training programs using several state-of-the-art propensity score matching (PSM) estimators. Additionally, we also apply a combined matching difference-in-differences (MDiD) estimator to account for unobserved individual characteristics (e.g. motivation, ability). While the estimated ATE and ATT for general training are significant ranging between about 4 and 7.5 %, the corresponding wage effects of firm-specific training are mostly insignificant. Using the more appropriate MDiD estimator, however, we find a more precise and highly significant wage effect of about 5 to 6 %, though only for general training and not for firm-specific training. These results are consistent with standard human capital theory insofar as general training is associated with larger wage increases than firm-specific training. Furthermore, we conclude that firms may intend to use specific training to adjust to new job requirements, while career-relevant changes may be conditioned to general training.

Keywords: Continuous training; wage effect; average treatment effect; selectivity bias; propensity score matching estimators; combined matching difference-in-differences estimator.

JEL classification: C21, J24, J31, M53

The authors gratefully acknowledge financial support from the German Research Foundation (DFG). In addition, we have benefited a lot from the discussions with Andreas Ammermueller, Todd Bradley, Bernd Fitzenberger, Dietmar Harhoff, Olaf Huebler, Eva Muthmann, and Stephan Lothar Thomsen whose suggestions have been of great value. We are also indebted to Todd Bradley for improving our English. Any remaining errors are, of course, our own.

^{*} Dipl.-Kffr. Grit Muehler, Centre for European Economic Research (ZEW), Department of Labour Markets, Human Resources and Social Policy, L 7,1, D-68161 Mannheim, Phone: +49-621-1235-152, Fax: +49-621-1235-225, E-Mail: muehler@zew.de

^{**} Prof. Dr. Michael Beckmann, University of Basel, Center for Economic Sciences (WWZ), Department for Human Resources and Organization, Rosshofgasse 2, CH-4051 Basel, Switzerland, Phone: +41-61-267-3224, Fax: +41-61-267-2758, E-mail: michael.beckmann@unibas.ch.

^{***} Prof. Dr. Bernd Schauenberg, University of Freiburg, Department of Personnel and Organizational Economics, Platz der Alten Synagoge, D-79085 Freiburg im Breisgau, Germany, Phone: +49-761-203-2390, Fax: +49-761-203-2394, E-mail: bernd.schauenberg@vwl.uni-freiburg.de.

1 Introduction

Continuous training is a very important tool of human resource management toward adjusting to new firm requirements as caused, for example, by technological and organizational change. It may additionally serve as a screening device of the firm to allocate workers to jobs more efficiently or as an incentive mechanism to encourage workers to provide some extra effort. Although continuous training can typically be assumed to increase a worker's productivity, this productivity effect of training does not necessarily need to be reflected in a corresponding wage growth.¹

According to the prediction of standard human capital theory, continuous training is likely to increase post-training wages, where the wage effect of general training is expected to exceed the corresponding wage effect of firm-specific training. Becker (1962) has argued that generally-trained workers receive the total revenue of their human capital investment, while the costs and benefits from an investment in firm-specific human capital are shared between employer and employee. These (sharing) rules can be explained with different fluctuation propensities. While generally-trained workers have accumulated skills that are transferable to other firms, the skills of specifically-trained workers can only be used productively within the training firm. Hence, generally-trained workers receive wages equal to their marginal productivity (because paying them below marginal productivity would cause them to quit), while specifically-trained workers can be paid below their marginal productivity (without them leaving the firm). A first hypothesis for our empirical analysis therefore is that both general and firm-specific training are associated with positive wage increases, but the effect for general training should exceed the corresponding effect for firm-specific training.

However, the new training literature has already developed some theoretical arguments that provide an alternative interpretation of the relationship between continuous training and wages. According to this literature, productivity effects of training do not necessarily need to have a positive impact on wages (e.g. Acemoglu and Pischke 1998, 1999a, 1999b; Leuven 2005). For example, if the employer has excess bargaining power, he is at least able to pay generally trained workers below their marginal productivity, while specifically trained workers might even be paid like untrained workers. Labour market imperfections may induce a similar effect. Since labour market imperfections (e.g. asymmetric information, mobility

¹ The terms 'training' and 'continuous training' are used synonymously in this paper. By continuous training we mean the individual's participation in job-related courses and seminars. These courses and seminars are either conducted internally (i.e. within the firm) or outside by external institutions and providers. Some examples of possible training measures are language courses, courses for improving technical skills, and computer courses.

costs, or wage floors) imply that generally trained workers do not receive the full marginal return on their productivity, their post-training wage effect depends on the amount of the employer's investment in general training. If the employer bears all the training costs, a generally trained worker cannot expect a significant increase in his post-training wage. The same logic holds for specifically trained workers. According to these considerations, a second hypothesis for our empirical investigation can be derived: Neither generally trained workers nor specifically trained workers must necessarily realize significant wage increases in response to their training programs.

In addition to these human capital considerations, we provide a related explanation that is based on the particular objective behind continuous training. For example, if firms intend to use specific training quite spontaneously to adjust to new job requirements, these training programs aim at avoiding productivity declines rather than increasing productivity. As a consequence, post-training wages cannot be expected to differ significantly from pre-training wages. On the other hand, if firms intend to apply general training programs systematically, these programs may also be seen as an instrument of the firms' long-term career policy. In this case, post-training wages might significantly exceed pre-training wages.

The effects of continuous training on wages have previously been investigated in several empirical studies.² However, the empirical results are quite ambiguous, which can at least partially be attributed to the application of different data sets and estimation strategies. The international evidence mainly stems from the U.S. and Canada (see e.g. Lynch 1992; Veum 1995; Loewenstein and Spletzer 1998, 1999; Parent 1999, 2003; Frazis and Loewenstein 2003). Nearly all of these studies account for heterogeneity by applying Heckman selectivity correction or fixed effects estimators. Their resulting wage effects of training vary between 0.2 and 3.9 %. Considering European research, there is a comparatively large stock of studies using U.K. data (e.g. Booth 1991, 1993; Blundell et al. 1996, 1999; Arulampalam and Booth 2001). Applying various estimation techniques (OLS, fixed effects, quasi-differences, selection correction and instrumental variables) these studies find wage effects between 1 % and 11 % for male workers.³

Meanwhile, there are several empirical studies on the wage effects of continuous training using German data sets. Almost all of these investigations explicitly address the potential selectivity problem, which arises from not accounting for observed and unobserved

² Valuable surveys on the effects of training on wages are provided by Bishop (1997) and Leuven (2004).

³ Evidence for other European countries is rather sparse. See for France Goux and Maurin (2000), for Norway Schöne (2002), for Switzerland Gerfin (2004) and Gerfin et al. (2003), for the Netherlands Leuven and Oosterbeek (2002), and for Portugal Budria and Pereira (2004).

heterogeneity between training participants and non-participants (e.g. Büchel and Pannenberg 2004; Pischke 2001; Garloff and Kuckulenz 2006; Kuckulenz and Maier 2006; Kuckulenz and Zwick 2003, 2005; Jürges and Schneider 2005). Some of these studies are based on instrumental variables (IV) estimations. The drawback of these studies is that IV approaches require relevant and valid instruments for continuous training. Otherwise, the coefficients are likely to be biased and inconsistent. As a consequence, the estimated wage effects of continuous training resulting from IV approaches are often inflated with respect to magnitude and standard errors, which makes an interpretation quite difficult.⁴ On the contrary, other studies that do not rely on IV estimates report quite small and often insignificant wage effects of continuous training. These studies consider that selection into training might be driven by observed and unobserved characteristics (e.g. motivation and ability). For example, applying a fixed growth model Pischke (2001) finds that selection into training is due to wage growth rather than wage levels and that workers with higher wage growth participate more often in training measures. Finally, Jürges and Schneider (2006) focus on the comparison of different estimation techniques (Hausman-Taylor model, fixed-growth estimator and matching combined with difference-in-differences) using the calendar information of the German Socio Economic Panel (GSOEP) and do not find a significant causal effect of training on wages. Therefore, the conclusion is that the supposed returns to continuous training are actually returns to unobserved heterogeneity.

The objective of our study is to add to the existing literature in the following way. In contrast to the vast majority of empirical studies, we focus on the wage effects of continuous training, separated by general and firm-specific training programs. This distinction is of major importance as human capital theory predicts different wage effects for both kinds of training. To our knowledge there are only the studies of Lynch (1992), Loewenstein and Spletzer (1998, 1999), Kuckulenz and Zwick (2003), and Gerfin (2004), where only Loewenstein and Spletzer (1999) focus directly on that distinction.⁵ Applying OLS regressions they find an effect of continuous training on wages 4.4 % to 6.2 % (depending on whether or not the employee switches the employer). This effect decreases to 1.4 % if fixed-effects estimation is

⁴ Sometimes the estimated wage effects of continuous training appear to be quite unrealistic with respect to magnitude ranging from -62.5 % to 52.9 % (see e.g. Budria and Pereira 2004).

⁵ Lynch (1992) distinguishes between on-the-job training and off-the-job training, Kuckulenz and Zwick (2003) separate external from internal training and Gerfin (2004) considers firm-sponsored and work-related training. Loewenstein and Spletzer (1998) have to use a proxy for general training. They find that training pays-off more for employees who switch their employer. One study that analyses firm-sponsored training is the paper of Schøne (2004). He finds that Norwegian firms mainly finance training of general skills which has a positive impact on wages of the employees.

applied. Additionally, the authors do not find that the wage increase depend on whether the skills acquired are general or firm-specific.

In our study, we use the data of the German Socio Economic Panel (GSOEP) and apply some state-of-the-art nonparametric matching estimators that explicitly account for observed and unobserved heterogeneity between training participants and non-participants. These matching approaches are based on the estimation of propensity scores (PSM estimators). More precisely, we first apply nearest neighbour matching, kernel matching, and local linear regression matching to evaluate the wage effects of continuous training programs in Germany. Since these matching approaches only account for selectivity that can be attributed to observed individual characteristics (e.g. age or education) but not for unobserved differences between training participants and non-participants (e.g. differences in ability and motivation), the estimated effect of continuous training on wages might not be causal. Therefore, we additionally account for unobservable characteristics combining matching with a difference-in-differences estimation strategy (MDiD). The idea of the MDiD approach is to match training participants and non-participants according to observable characteristics and then to compare their wage developments. This approach allows us to control for time invariant heterogeneity that was not captured by the matching and thus contributes to our objective to estimate a causal effect of continuous training on wages.

In our contribution we go further than the above mentioned fixed-growth models applied, for example, in Pischke (2001), which calculate individual but fixed growth rates of wages. Furthermore, we extend the MDiD results, as obtained, for example, in Jürges and Schneider (2006) by estimating separate effects for general and firm-specific training. In our opinion, the most related study to our paper is Gerfin (2004), who also applies the MDiD approach using data from the Swiss Labour Force Survey and distinguishing between firm-sponsored and work-related training. Consistent with some other studies that explicitly account for unobserved heterogeneity the author finds a quite moderate wage effect of work-related training of about 2 % and an insignificant effect of firm-sponsored training.

The paper is organized as follows: Section 2 provides the data, variables and descriptive statistics. In section 3, we introduce our econometric modelling. Section 4 contains the empirical results of the reference OLS estimator, the matching estimators including our preferred MDiD approach and some robustness checks. Finally, section 5 concludes.

2 Data, variables and descriptive statistics

Our empirical analysis is based on the year 2000 and 2004 waves of the German Socio Economic Panel (GSOEP). The GSOEP is a longitudinal study of private households in Germany. Launched in 1984, the study has been surveying households annually. The panel offers information on German citizens and immigrants living in the eastern or western parts of Germany. The GSOEP data cover a wide range of areas including the individual's traits, biography, employment status, career, professional mobility, earnings, education, training, family, health, satisfaction, household composition and living situation making it one of the most commonly used individual-level data sets in Germany (see Haisken-DeNew and Frick 2005 for further details).

Our main variable of interest is participation in continuous training. The questionnaire allows us to examine the wage effects of three main treatments: continuous training (*CT*) which can further be separated into general training (*GT*) and firm-specific training (*FST*). Individuals were firstly asked about the sources they use for continuous training. Possible responses are regularly reading scientific or professional publications, attending professional conferences and participating in professionally orientated courses and seminars. We use the information on this last item to generate our variable for continuous training.

This item simultaneously works as a filter for the subsequent questions on job-related training participation within the last three years. Regarding our research question we are interested in the information whether or not continuously trained individuals would be able to transfer the skills and abilities acquired during the training to another employer. The possible responses are: (a) not at all, (b) limitedly, (c) dominantly, (d) completely. We create dummy variables for firm-specific and general training, defining responses (a) and (b) as an indication of firm-specific training and responses (c) and (d) as an indication of general training.⁶

In the panel wave of 2004 respondents could collectively name three training courses they participated in during the last three years. We choose to use the information of the middle of these courses, as the latest course might still be lasting and the least recent course has probably been attended quite a long time ago.⁷ In both panel waves, 2000 and 2004, continuous training has been one of the major topics. Since respondents are asked for information covering the last three years, our treatment period of interest ranges from 2001 to

⁶ See Loewenstein and Spletzer (1999) for an analogous approach using data from the National Longitudinal Survey of the Youth (NLSY).

⁷ However, we use the data of the least recent training course in one of our robustness checks in section 4.

2003. Thus, we can use the pre-treatment characteristics from the year 2000 questionnaire and the post-treatment characteristics from the wave of 2004.

Our outcome variable of interest is the individual's current gross monthly income.⁸ We restrict our sample to male workers to avoid another selectivity problem resulting from the a priori decision problem by women whether or not to participate in the labour market. To achieve unambiguous results we do furthermore not consider part-time workers, self-employed, apprentices, and public servants. We control for individual characteristics (years of education, marital status, nationality, age), job-related factors (tenure, fixed-term employment, overtime work, working in the certified job, occupational status), firm size, sector affiliation and region (see Table A1 in the appendix for a detailed description of the variables).

Thus, our sample consists of 1,751 individuals, where 426 receive continuous training and 1,325 do not. Table 1 displays the mean values of the dependent variable and some of the explanatory variables, separated for both groups. This comparative analysis provides first insights on specific differences between continuously-trained and untrained workers and allows us to obtain information on the determination of continuous training.

The results show the expected correlation between continuous training and increasing wages. According to the calculated means, trained workers earn on average about 800 Euros more per month than untrained workers. Of course, this preliminary finding is not very meaningful unless one controls for additional heterogeneity using multiple regression analysis. Our preliminary findings furthermore suggest substantial differences with respect to other characteristics. For example, continuously-trained workers are slightly younger than untrained workers. They are also better educated, which is indicated by the higher amount of years of schooling. This result is consistent with the commonly accepted finding that skilled or better educated employees are more likely to participate in training programs than less-skilled employees (e.g. Düll and Bellmann 1999).⁹ Moreover, continuously-trained workers are more likely to provide extra working time, and work more often in their certified job. Finally, fixed-term workers and workers of non-German origin are less likely to participate in continuous training programs. However, while these bivariate correlations may provide useful insights, they are likely to suffer from an omitted variables bias and may suggest misleading

⁸ We use the natural logarithm of this income and exclude responses below 600 Euros to avoid implausible information in our sample of full-time employed males.

⁹ The results remain unchanged when using educational degree instead of years of schooling.

conclusions. This drawback can be avoided by applying multiple regression analysis, which is the objective of our empirical investigation.

Table 1: Continuously-trained vs. untrained workers: mean values of selected variables

Variable	Means for continuously-trained workers	Means for untrained workers
Gross wage per month (<i>W</i>)	3,753.29	2,943.89
Years of education (<i>S</i>)	13.21	11.86
Worker's age (<i>AGE</i>)	41.87	43.81
Tenure (<i>TEN</i>)	11.84	12.67
Unmarried (<i>UNMARRIED</i>)	0.20	0.15
Nationality (<i>NATION</i>)	0.94	0.87
Fixed-term worker (<i>FTE</i>)	0.03	0.04
Adequate job (<i>AD_JOB</i>)	0.62	0.55
Overtime work (<i>O_TIME</i>)	4.03	2.64
West German (<i>WEST</i>)	0.81	0.76
<i>N</i>	426	1,325

Note: The calculations are restricted to individuals who do not provide item non-responses for the subsequent regression analysis. The means of the wage variable *W* are measured in Euro. *UNMARRIED*, *NATION*, *FTE*, *AD_JOB*, and *WEST* are dummy variables. Thus, the means display the fraction of individuals belonging to that certain feature. *N* indicates sample size. Source: German Socio Economic Panel (wave 2004), own calculations.

3 Econometric modelling

Our estimation strategy proceeds in three steps. First, we specify Mincer-type wage functions to evaluate the wage effect of continuous training on wages using the usual OLS procedure.¹⁰ In order to account for a possible selectivity bias arising from the fact that OLS ignores the endogeneity of continuous training participation, our second approach is to estimate an average treatment effect applying several matching methods based on propensity scores. However, as this strategy might not eliminate a bias resulting from unobserved heterogeneity, we thirdly estimate the effect of continuous training on wages using a difference-in-differences estimator on the matched sample (MDiD). Finally, we perform several robustness checks to assess the sensitivity of our results.

¹⁰ In the standard Mincer wage function, log wages is regressed on education and job experience (original and squared observations). The Mincer wage function is based upon the schooling function that assumes an exponential relation between wages and the years of schooling. For further details, see Mincer (1974).

OLS estimation

Our econometric model has the general form

$$\ln W_i = \alpha_0 + \alpha_1 S_i + \alpha_2 TEN_i + \alpha_3 (TEN_i)^2 + \gamma CT_i + \delta' X_i + v_i, \quad (1)$$

where $\ln W_i$ is the natural logarithm of gross monthly wages, S_i is the years of education and TEN_i is tenure of individual i .¹¹ CT is a dummy variable that separates employees who recently received continuous training from those who have not. X_i is a vector containing individual socio-demographic characteristics, job-related factors, information on working conditions and occupational status, and other control variables. Finally, α_i , γ and δ represent the coefficients to be estimated, and v_i is a stochastic error term with the usual assumptions, i.e., normal distribution, zero mean, and finite variance.

Since the available data allow us to distinguish between the wage effects of general training and firm-specific training, equation (1) can be modified to

$$\ln W_i = \beta_{g0} + \beta_{g1} S_i + \beta_{g2} TEN_i + \beta_{g3} (TEN_i)^2 + \gamma_g GT_i + \delta'_g X_i + v_{gi}, \quad (2)$$

or

$$\ln W_i = \beta_{f0} + \beta_{f1} S_i + \beta_{f2} TEN_i + \beta_{f3} (TEN_i)^2 + \gamma_f FST_i + \delta'_f X_i + v_{fi}, \quad (3)$$

where GT_i (FST_i) is a dummy variable that separates employees who have received general training (firm-specific training) within the last three years from those who have not received any continuous training within that time span. In the absence of selection problems with respect to training participation, the training coefficients γ , γ_g , and γ_f in specifications (1) to (3) can consistently be estimated using the conventional OLS procedure.

Importantly, the decision process whether or not to participate in a continuous training program is usually endogenous. Not accounting for the selection decision of employee and employer is typically associated with biased and inconsistent parameter estimates. More precisely, the wage effect, which has falsely been attributed to continuous training, can be driven by individual and firm characteristics influencing the probability of participating in a training program and not by continuous training itself.

¹¹ We use tenure instead of total working experience. The reason for this proceeding is that experience cannot directly be observed in our data. We would have to approximate experience by individual's age minus years of schooling minus 6 (the age at which children usually start schooling in Germany), which is a relatively imprecise measure of effective working experience.

Propensity Score Matching

In order to account for a potential selectivity bias, our second approach is the estimation of an average treatment effect by comparing trained and untrained individuals (see e.g. Heckman et al. 1997). Thus, the main requirement is to construct an adequate control group in a way that the only remaining difference between both groups is participation in training (e.g. Blundell and Costa Dias 2002; Caliendo and Kopeinig 2006). The average treatment effect for the population (ATE) describes the difference of the expected outcomes after training participation and non-participation and is defined as

$$\Delta_{ATE} = E(\Delta) = E(\ln W^1) - E(\ln W^0). \quad (4)$$

Here, $E(\ln W^1)$ is the expected log-wage for participants and $E(\ln W^0)$ the corresponding expected value for non-participants. The drawback of the ATE is that it includes the effect on individuals for whom the treatment, i.e. the training program, was never intended. Therefore, the more prominent evaluation parameter is the average treatment effect on the treated (ATT), which explicitly concentrates on the wage effects on those individuals for whom the training program is actually intended. The ATT is defined as the difference between the expected outcomes with and without participation in continuous training for the effective treatment participants:

$$\Delta_{ATT} = E(\Delta | CT = 1) = E(\ln W^1 | CT = 1) - E(\ln W^0 | CT = 1). \quad (5)$$

The second term on the right hand side of equation (5) describes a hypothetical outcome without treatment for individuals who actually received treatment and is therefore unobservable. Under the condition that $E(\ln W^0 | CT = 1) = E(\ln W^0 | CT = 0)$, the non-participants serve as an adequate control group.

This condition only holds if two main assumptions are fulfilled. First, the individuals are selected into treatment solely according to observable pre-treatment characteristics and the outcomes of non-participants would be the same if they were trained instead (Conditional Independence Assumption). This is a fairly strong assumption as our data do not contain variables for individual motivation or ability, which are likely to affect the selection into treatment and wages. However, due to the fact that we have the information on training participation for both years, 2000 and 2004, we are able to use former (year 2000 training participation) as a proxy for the general attitude of individuals towards training. Thus, we can match individuals who participate in training in the year 2004 to non-participants in 2004 with

respect to the information whether or not these individuals have already participated in continuous training in 2000.

The second necessary assumption is a positive probability to be in the treatment or in the control group to avoid comparing non-comparable individuals (Common Support Assumption). This assumption might be critical if programs are directed to certain groups of employees. However, as our sample is restricted to full-time employed male workers, the Common Support Assumptions is not unlikely to hold.

Since conditioning on a high-dimensional vector of observable characteristics is difficult, we follow Rosenbaum and Rubin (1983) and match on the treatment probability $P(Z)$, where Z contains the pre-treatment characteristics:

$$E(\ln W^0 \mid CT = 1, P(Z)) = E(\ln W^0 \mid CT = 0, P(Z)). \quad (6)$$

Since we distinguish three different treatments, we can match conditional on the probability to participate in continuous training, general or firm-specific training. Thus, we firstly estimate the propensity score and subsequently implement our matching estimator of choice. The propensity score matching (PSM) estimator for the ATT can then be written as

$$\Delta_{ATT}^{PSM} = \sum_{i \in T} \left\{ W_i - \sum_{j \in C} \Lambda_{i,j} W_j \right\} \lambda_i. \quad (7)$$

Here, W_i is the wage of the training participants, W_j is the wage of the non-participants, and T and C indicate treatment and control group. $\Lambda_{i,j}$ is the weight placed on comparison observation j for individual i .¹² Finally, λ_i reconstructs the outcome of the treated sample by re-weighting, for example, with equal weights to all observations or kernel weights (Blundell and Costa Dias 2002).¹³ Hence, the PSM estimator is the mean difference in wage outcomes over the common support, appropriately weighted by the propensity score distribution of training participants (Caliendo and Kopeinig 2006).

The PSM approach estimates the ATT or the ATE, respectively, applying a two-step procedure. In the first step, the propensity score is estimated. For this purpose, a binary choice model for CT (GT , FST) with the covariates is estimated using the probit maximum likelihood method. In the second step, the wage effect of CT (GT , FST) is estimated using matching

¹² In our analysis, we use kernel weights to account for the closeness in outcomes of participants and non-participants.

¹³ Re-weighting is necessary if the group of the treated individuals is larger than the group of controls due to oversampling of treated individuals. As this is not the case in our GSOEP data, we do not re-weight the sample.

techniques based on propensity score estimation. Specifically, we implement nearest neighbour matching, kernel matching and local linear regression matching.¹⁴

The most straightforward propensity score matching method is nearest neighbour matching (Heckman et al. 1999; Lechner 1998). The idea of nearest neighbour matching is to find for each treated individual a corresponding matching partner, i.e., a particular non-treated individual with the closest propensity score. We decided to use four matches for the nearest neighbour because we have a relative large group of controls compared to the number of treated individuals. Moreover, using four matches was found to perform well in terms of mean-squared error (Abadie and Imbens 2002; Abadie et al. 2004). In addition, we allow for replacement because this proceeding preserves us from relying on too little information and incorporating observations that are not sufficiently similar (Smith 1997). Matching with replacement (i.e. one control individual can be matched to more than one of the treated) is less beneficial than matching without replacement in terms of variance. However, it is more beneficial in terms of bias, and so matching with replacement is our preferred matching strategy.

Kernel matching (see Heckman et al. 1998 for details) and local linear regression matching represent nonparametric matching estimators that use weighted averages of (almost) all individuals in the control group to construct the counterfactual outcome.¹⁵ Hence, an advantage of these methods is the relative small variance resulting from the fact that more information is used. However, one drawback is that observations might be used that are bad matches. In principle, kernel matching is a weighted regression of the counterfactual outcome on an intercept with weights given by the kernel weights. These weights depend on the distance between each control group member and the participant observation for which the counterfactual is estimated. The difference between kernel matching and local linear regression matching is that the latter does not only include the intercept but also a linear term in the propensity score of a treated individual.

Kernel matching and local linear regression matching require choosing the kernel function and the bandwidth parameter. While the choice of the kernel function is not of major importance in practice, the choice of the bandwidth parameter involves a trade off between a small variance and an unbiased estimate of the true density function. Large bandwidth values

¹⁴ These matching estimators are commonly used in evaluative studies and are extensively described, for example, in Blundell and Costa Dias (2002), Blundell, Dearden and Sianesi (2005), Caliendo and Hujer (2006), Caliendo and Kopeinig (2006), and Cameron and Trivedi (2005).

¹⁵ The question of whether all or almost all individuals in the control group are used depends on the choice of the kernel function (Caliendo and Kopeinig 2006).

tend to decrease the variance between the estimated and the true density function but lead to a biased estimate (Caliendo and Kopeinig 2006). We chose the default Epanechnikov kernel function and a default bandwidth parameter of 0.06 for kernel matching, and the default tricube kernel function and a default bandwidth parameter of 0.8 for local linear regression matching.

The major weakness of the matching approach is that selection into training (and the wage effects as well) can only be ascribed to observed individual and firm characteristics. However, the training decision is likely to be driven by unobservable characteristics, too. In this case, the Conditional Independence Assumption is unlikely to hold. As a consequence, matching reduces but does not completely eliminate selection bias.¹⁶

Combined Matching Difference-in-Differences estimator

Our idea of additionally controlling for previous participation in continuous training might probably not suffice to satisfy the Conditional Independence Assumption. Specifically, unobserved characteristics, for example, differences in ability or motivation between training participants and non-participants, are unlikely to be completely captured by an explanatory variable for previous training participation. Since these characteristics are time-invariant, the matching difference-in-differences estimator (MDiD) compares the conditional before treatment-after treatment wages of the trained individuals to those of the untrained to control for remaining unobservable differences (see e.g. Heckman et al. 1998, Blundell and Costa Dias 2002, Caliendo and Kopeinig 2006).

The combination of matching with a DiD approach extends the method of matching by ruling out the unobserved heterogeneity which is still present. The MDiD estimator (or Conditional DiD estimator) is obtained by comparing the mean outcomes of participants with the corresponding outcomes of their matched counterparts. Extending equation (7), the MDiD estimator can be written as

$$\Delta_{ATT}^{MDiD} = \sum_{i \in T} \left\{ (W_{it_1} - W_{it_0}) - \sum_{j \in C} \Lambda_{i,j} (W_{jt_1} - W_{jt_0}) \right\} \lambda_i, \quad (8)$$

where t_1 and t_0 are time indices. The estimator is superior to DiD as it does not impose linear functional form restrictions and superior to matching as it relaxes the Conditional

¹⁶ This bias does not completely diminish because due to unobserved characteristics training participants and non-participants are selected groups earning different wages, even in the absence of continuous training programs.

Independence Assumption and allows for unobservable, time invariant differences in the outcomes between training participants and non-participants (Caliendo and Kopeinig 2006). The key assumption regarding the use of this estimator is that common time trends apply for both treated individuals and matched controls. By comparing pre- and post-training wages we additionally need to assume that the wages of the matched non-participants evolve the same way than the wages of the training participants if they had not been trained.¹⁷

4 Empirical results

OLS estimation

The first step in our empirical strategy is the estimation of a Mincer-type wage function using the cross-sectional data of the year 2004. The results of the usual OLS estimation of equations (1) to (3) are summarized in Table 2. We run separate regressions for our three types of training, namely continuous training, general and firm specific training. Thereby, Panel A displays the results of a typical Mincer wage function augmented by our variable for continuous training (*CT*, *GT*, or *FST*) and a set of common control variables for marital status, nationality, fixed-term employment, firm size, sector affiliation and region. Panel B additionally accounts for the occupational status of the individuals, overtime work and working in the certified job. The estimates of all control variables are summarized in Table A3 in the appendix.

Panel A shows similar results for all three types of training. The participation in continuous training is associated with a significant increase in wages of 9.4 %. For general training the effect is slightly higher (9.7 %) and for firm specific training somewhat lower (7.8 %). The impact of the variables for education and tenure is similar in all three equations.

Additionally controlling for occupational status and further job characteristics, we find that continuous training is no longer associated with significant wage increases. Thus, the wage effects of our training measures are quite sensitive to model specification. The wage effects of schooling reduce as well when moving from the Panel A to the Panel B specification. More precisely, the schooling effect declines from about 7 % in Panel A to about 2 % in Panel B, but is still highly significant. All estimates of our control variables are significant with the anticipated sign, except the variables indicating working in the certified job and nationality,

¹⁷ We performed robustness checks on that. See section 4 for details.

which are both insignificant (see Table A3 in the appendix).¹⁸ The remaining results can be summarized as follows: Unmarried workers earn less than married workers, fixed-term workers earn less than permanent workers, and West German workers earn more than their East German counterparts. Moreover, there are highly significant skill and hierarchical wage effects indicated by our controls for occupational status. Finally, we observe the well-known firm size wage differential.

Table 2: OLS wage effects of different forms of continuous training

Training form	Continuous training		General training		Firm-specific training	
Panel A:						
<i>S</i>	0.067***	(19.12)	0.067***	(17.80)	0.069***	(17.31)
<i>TEN</i>	0.010***	(3.84)	0.010***	(3.94)	0.009***	(3.14)
<i>(TEN)²*100</i>	-0.016**	(-2.46)	-0.017***	(-2.57)	-0.013*	(-1.91)
<i>CT</i>	0.094***	(4.80)				
<i>GT</i>			0.097***	(4.18)		
<i>FST</i>					0.078***	(2.78)
Panel B: Panel A + occupational status + overtime work + adequate job						
<i>S</i>	0.020***	(5.47)	0.018***	(4.58)	0.021***	(4.92)
<i>TEN</i>	0.008***	(3.87)	0.009***	(4.19)	0.007***	(2.73)
<i>(TEN)²*100</i>	-0.015***	(-2.65)	-0.018***	(-3.06)	-0.011*	(-1.77)
<i>CT</i>	0.015	(0.90)				
<i>GT</i>			0.008	(0.42)		
<i>FST</i>					0.021	(0.88)
<i>F</i> test	68.48***		60.43***		52.08***	
<i>R</i> ²	0.632		0.629		0.621	
<i>N</i>	1,751		1,604		1,472	

Note: */**/** indicates significance at the 10/5/1 percent level. The dependent variable is $\log W$. The coefficients in Panels A and B are estimated by OLS. The values in parentheses represent the robust *t*-statistics using heteroscedasticity-consistent standard errors of the coefficients. The basic model in Panel A additionally contains control variables for fixed-term workers, nationality, marital status, West-German workers, five firm size dummies, ten sector dummies and 15 regional dummies. The *F* test and *R*² indicate the estimation quality of the equations. *N* indicates sample size. All the latter statistics refer to Panel B. Source: German Socio Economic Panel (wave 2004), own calculations.

As previously been mentioned, the OLS estimates do not account for the endogenous decision to participate in continuous training programs.¹⁹ For example, better educated and younger workers as well as workers in skilled occupations may be more likely to attend continuous

¹⁸ The effects of the sector and regional dummies are not displayed in the Appendix. They are available from the authors upon request. Both sets of dummies show mostly insignificant coefficients.

¹⁹ The fact that we observe relatively volatile wage effects of our measures for continuous training (and schooling), while the coefficients for our control variables remain quite stable in different model specifications, may not solely indicate an omitted variables bias. It can also be viewed as an indication for a selectivity bias.

training. Therefore, our estimation results might be biased. This weakness can be rectified by using propensity score matching estimators instead of the usual OLS method.

Propensity Score Matching

One solution to overcome the selection bias is matching. We thus compare the wages of individuals who participate in continuous training to a group of people who do not participate in the treatment. To establish an adequate control group we have to match individuals according to pre-treatment characteristics. To obtain this information we go four years backwards and use the information of the wave 2000 questionnaire of the GSOEP. Since the respondents are asked in 2004 to review the last three years when answering the question, whether or not they participated in some sort of continuous training, we can use characteristics prior to this treatment.

We obtain the matching variables starting again with the classical Mincer explanatory variables and controls for regional differences. We use age instead of tenure and renounce the squared term as this term does not predict participation in training on a statistically significant level. The further matching variables are selected on the criterion of statistical significance when adding them iteratively to our baseline specification, which consists of a constant and age. Since we also aim at matching on the information of occupational status but do only have relatively few observations for managers and unskilled white collar workers (see Table A2 in the appendix for the variable means), we combine the information on occupational status into three variables indicating unskilled work (*US*), skilled work (*SK*) and high-skilled occupations (*HS*).²⁰ Table A4 in the appendix displays the results from a probit model of participation in continuous training as a function of pre-training characteristics.

Moving to a panel setting we can make use of additional information, which we could not use in the OLS regressions. Since continuous training is also a major topic in the year 2000 questionnaire, we are able to use the information whether or not the respondents have already participated in some job-related training in the three years prior to the year 2000 specifying

²⁰ *HS* is a dummy variable, which equals one when the individual executes high-skilled white collar work or managerial activities. The variable *SK* equals one if the individual executes skilled blue or white collar work. The reference category is *US* which is equal to one if the individual executes unskilled blue or white collar work.

our variable PT .²¹ This allows us to distinguish between individuals, which show a high affiliation to training arrangements from those who do not.

We apply three matching procedures, nearest neighbour matching, kernel matching and local linear regression and ensure common support by dropping treatment observations, whose propensity score is higher than the maximum or lower than the minimum propensity score of the controls. The means of the treatment group, control group and matched controls are displayed in Table 3.²² According to these results, the means for training participants are quite close to the corresponding means of their matched counterparts and quite different from the means of the non-participants.

Table 3: Means of variables in subgroups

Variable	All	Training participants	Non-participants	Matched non-participants
Gross wage per month (W)	2,731.94	3,166.28	2,592.29	2,855.20
Years of education (S)	12.16	13.17	11.84	12.45
Worker's age (AGE)	39.34	37.87	39.81	38.20
Tenure (TEN)	9.95	9.48	10.11	9.53
Unmarried ($UNMARRIED$)	0.21	0.26	0.20	0.24
Nationality ($NATION$)	0.89	0.94	0.87	0.91
Fixed-term worker (FTE)	0.04	0.03	0.05	0.04
Adequate job (AD_JOB)	0.59	0.65	0.57	0.65
Overtime work (O_TIME)	3.43	4.24	3.17	3.46
West German ($WEST$)	0.77	0.81	0.76	0.79
Skilled work (SK)	0.56	0.50	0.58	0.60
High skilled work (HS)	0.23	0.41	0.17	0.27
N	1,751	426	1,325	608

Note: These are the pre-treatment characteristics, i.e. the variables refer to the year 2000 wave of the GSOEP and thus the means differ from the corresponding values obtained for the year 2004. The matched non-participants are drawn from nearest neighbour matching for CT , allowing for replacement and using the four closest matches. N indicates sample size. Source: German Socio Economic Panel (wave 2000), own calculations.

Table 4 displays the ATE and the ATT resulting from the three different matching methods. The estimates for the ATE show that the global wage effect of continuous training ranges in the interval between 4.7 and 5.9 %. Hence, the magnitude of this wage effect lies in the

²¹ The variable we use to get this information is the same as in the 2004 questionnaire. See section 2 for details. We again regard individuals as participants in continuous training when they stated to have participated in professionally oriented courses.

²² Plotting the density of the propensity score distribution reveals that the treated and control individuals differ. Especially in the higher propensity scores we have a quite substantial amount of treated observations but relatively few controls. Nevertheless, we can assure common support.

middle of the wage effects resulting from our first and second (insignificant) OLS estimates. However, it is much more interesting to interpret the separate wage effects of general and firm-specific training. For general training both the ATE and the ATT provide significant wage effects ranging between 4.1 and 7.5 %. The corresponding wage effects of firm-specific training (ATE and ATT) are mostly insignificant. Finally, the ATT estimates are somewhat smaller the ATE estimates, which indicates that continuous training is typically not directed to the workforce as a whole but to some selected kinds of employees.

Table 4: Average treatment effects for different forms of continuous training on wages

Training form	Continuous training		General training		Firm-specific training	
Treatment variable	<i>CT</i>		<i>GT</i>		<i>FST</i>	
Matching method:	ATE	ATT	ATE	ATT	ATE	ATT
Nearest Neighbour	0.053** (0.024)	0.036 (0.030)	0.066* (0.035)	0.052* (0.031)	0.045 (0.047)	-0.007 (0.048)
Kernel	0.059*** (0.019)	0.044** (0.020)	0.075*** (0.028)	0.051** (0.024)	0.064* (0.039)	0.029 (0.028)
Local Linear Regression	0.047*** (0.017)	0.035 (0.023)	0.063** (0.030)	0.041* (0.025)	0.053 (0.043)	0.015 (0.029)
Treated	426		279		147	
<i>N</i>	1,751		1,604		1,443	

Note: */**/** indicates significance at the 10/5/1 percent level. The values in parentheses represent the bootstrapped standard errors of the coefficients. The number of bootstrap replications to be performed is 50 for the kernel matching procedure and 200 for the nearest neighbour and local linear regression matching estimators. 50-200 replications are typically adequate for standard error estimates (Mooney and Duval 1993). For nearest neighbour matching the number of matches is 4. The used control variables to estimate the propensity scores are displayed in Table A4 in the appendix. We impose common support by dropping treatment observations whose propensity score is higher than the maximum or less than the minimum propensity score of the controls. *N* indicates sample size. The number of observations for firm-specific training declines by 29 observations compared to the number given in Table 2 as one of the regional dummies causes perfectly predicted failure. Source: German Socio Economic Panel (waves 2000, 2004), own calculations.

The major weakness of the matching approach is that only observable characteristics are allowed to influence the selection into training and to affect wages. Even the application of such a rich data set like the GSOEP does not guarantee that all the heterogeneity between participants and non-participants can be captured sufficiently. Rather, matching procedures are likely to capture observed individual characteristics, while they are not able to account for unobserved characteristics. Therefore, matching reduces but does not eliminate the selection bias.

Combined Matching Difference-in-Differences estimator

The bias caused by time invariant unobservable characteristics is removed by combining matching with a difference-in-differences estimator (MDiD). The MDiD estimator can be obtained by regressing the difference in log-wages between 2004 and 2000 on a dummy variable indicating participation in continuous training and a constant, using the matched sample.²³ The results from this regression are shown in Table 5.

Table 5: MDiD estimators

Training form	Continuous training	General training	Firm-specific training
Treatment variable	<i>CT</i>	<i>GT</i>	<i>FST</i>
	0.049*** (0.015)	0.061*** (0.017)	0.018 (0.022)
<i>N</i>	1,034	777	483

Note: */**/***/ indicates significance at the 10/5/1 percent level. The values in parentheses represent the standard errors of the coefficients. The matched sample is obtained by performing nearest neighbour matching with replacement and using the four nearest neighbours. *N* indicates the sample size. Source: German Socio Economic Panel (waves 2000, 2004), own calculations.

The MDiD estimates are 4.9 % for continuous training as a whole and 6.1 % for general training. Both wage effects are highly significant. As expected, they are slightly smaller than the ATE estimates because MDiD eliminates the upward bias, which has previously been caused by unobserved heterogeneity. Firm-specific training is associated with a wage growth of 1.8 %, but this effect is again insignificant.²⁴

Robustness Checks

To check the sensitivity of our results with respect to variable specifications and estimation techniques we perform four robustness checks. First, we redefine the variables for firm-specific and general training. Second, we check if our results also hold true for the third

²³ The matched non-participants are drawn by performing the nearest neighbour matching as before (4 nearest neighbours, replacement) for our three variables of interest.

²⁴ Alternatively, one may argue that differences in prosperity between firms affect their motivation to provide general or firm-specific training. Specifically, general training may be more likely to be offered by prosperous firms (e.g. as a kind of bonus), while firm-specific training may be offered by firms regardless whether they prosper or not. As a consequence, prosper firms offering general training may be more likely to pay higher wages than less prosper firms, which tend to offer firm-specific training. Although we are not able to explicitly control for establishment prosperity, we address this issue indirectly by controlling for firm size and regional affiliation.

training course respondents could mention.²⁵ A third robustness check is exclusively dedicated to the MDiD estimation, where we check the sensitivity of the results when generating the matched sample by another than the nearest neighbour matching procedure. Finally, we check whether our estimates may be affected by temporary dips in wages that can be observed directly before training participation.

We perform a first robustness check by redefining our variables *GT* and *FST* restricting general training to acquired knowledge that is completely transferable and firm specific training to knowledge that is not transferable at all. The matching estimations for both general and firm specific training are mostly insignificant, while the MDiD estimator for general training is 6.2 % and significant at the 5 % level. Thus, the wage effect has a similar magnitude like our estimate presented above. The results are displayed in Table A5 in the appendix.

Second, we check whether the results change when performing the analysis with the furthestmost training course respondents could name. Thereby, our number of observations drops to 199 treated individuals for general training and 107 treated observations for firm-specific training. Table A6 in the appendix displays the results. The results from the matching procedures are mixed and only partially significant. The effects from the MDiD are comparable to what we have obtained before. The wage effect for general training is 4.9 % and significant at the 5 % level, and even the wage effect of firm-specific training (4.4 %) is weakly significant in this specification.

We carry out a third robustness check to test whether the results for our MDiD estimates are sensitive to the matching procedure used to establish the matched sample. Since the DiD estimator requires a quasi-experimental setting, we need to assure the adequacy of our control group and the comparability of treated and matched control individuals. Table A7 in the appendix displays the results of the MDiD estimators for the two other types of matching. The results strongly support our previously estimated wage effects. More precisely, the effect sizes of nearest neighbour matching, kernel matching and local linear regression are quite similar ranging between significant 4.8 and 6.4 % for general training. Hence, our results do not depend on the underlying matching procedure.

²⁵ We decided not to use the first training course respondents could name, neither for the analysis nor for the sensitivity check. Here, the individuals are explicitly asked for the most recent or current course and we do not expect to see wage effects from that training, mostly because of the fact that this course may just have been finished or might even last.

As a final important robustness check we have to exclude the possibility that anticipation effects or temporary individual-specific effects, which contribute to increase post-training wages, significantly affect our estimated wage effects. The literature often calls this phenomenon “Ashenfelter’s Dip” (Ashenfelter 1978; Ashenfelter and Card 1985). If temporary dips in wages occurred shortly before training participation, a faster growth in earnings would then be expected even without the treatment. However, the wage developments of training participants, non-participants and matched-controls show that temporary dips do obviously not exist (see Figure 1 and Table A8 in the appendix). However, we must admit that our conclusion would be even more convincing, if the GSOEP provided monthly wage data rather than yearly data.

All in all, the robustness checks show that our estimated wage effects are not sensitive to variable definitions or the applied matching approach.

5 Conclusion

The present paper examines the wage effects of continuous training programs using individual-level data from the German Socio Economic Panel (GSOEP). According to standard human capital theory, continuous training is expected to increase workers’ productivity and thus their wages, where general training should be associated with larger wage gains than firm-specific training. On the contrary, the new training literature allows for an alternative interpretation of the relationship between continuous training and wages, where productivity effects of training do not necessarily need to have a positive impact on wages. Previous investigations of the wage effects of continuous training provide very mixed results, largely depending on the applied data and econometric methods.

Therefore, our bid to add to the empirical literature is twofold. First, we differentiate between the wage effects of general and firm-specific training in Germany. Second, our estimation approach explicitly accounts for a potential selectivity bias caused by observed and unobserved heterogeneity between training participants and non-participants. Thereby, we at first estimate average treatment effects (ATE and ATT) of general and firm-specific continuous training programs using nearest neighbour matching, kernel matching, and local linear regression matching, which are all based on the estimation of propensity scores. Since these PSM estimators just eliminate a selectivity bias that is due to observable individual characteristics, we also apply a combined matching difference-in-differences estimator

(MDiD) to explicitly account for unobserved differences between training participants and non participants (e.g. motivation, ability).

Our results can be summarized as follows: While the estimated ATE and ATT for general training are significant ranging between about 4 and 7.5 %, the corresponding wage effects of firm-specific training are mostly insignificant. Using the more appropriate MDiD estimator, however, we find a more precise and highly significant wage effect of about 5 to 6 %, though only for general training and not for firm-specific training. These results are consistent with standard human capital theory insofar as general training is associated with larger wage increases than firm-specific training. Furthermore, we conclude that firms may intend to use specific training to adjust to new job requirements in order to prevent productivity declines, while career-relevant changes may preliminarily be conditioned to general training.

Our estimations provide some new results and thus add to the existing literature of the wage effects of continuous training. First, contrary to Loewenstein and Spletzer (1999), who find very moderate and quite similar wage effects of both general and firm-specific training, our wage effect for general training is somewhat higher and exceeds the corresponding effect for firm-specific training. Second, contrary to studies like Leuven and Oosterbeek (2002), Jürges and Schneider (2006), and Kuckulenz and Maier (2006), we can only partially conclude that wage effects of continuous training programs are returns to unobserved characteristics rather than returns to training. Third, we consider our results in line with Gerfin (2004), who can also discriminate between the wage effects of two types of training (work-related and firm-sponsored training) using Swiss individual data.

However, it should be noted that the heterogeneity of the international evidence on the wage effects of continuous training can at least partially be attributed to specific institutional differences between countries. For example, the Dual System of vocational education, which can be described as a front-loaded system, is specific to only a few countries like Germany or Switzerland. As a result, different wage effects of continuous training between countries may at least partially be caused by different arrangements with respect to the underlying system of apprenticeship training.

Our above stated interpretation that firm-specific training does not contribute to increasing wages, because firms use this tool to adjust to new job requirements, deserves some additional research. For example, if selection into training substantially depended on technological and organizational innovations, the decision whether or not to participate in specific training could be considered as a push rather than a pull decision. As a consequence, firm characteristics

should be paid more attention in future work, when estimating the wage effects of continuous training.

Given the evidence provided so far, the influence of continuous training on wages is rather modest. However, there seem to be some indications that training has a more pronounced positive influence on promotions (e.g. Pannenberg 1997; Krueger and Rouse 1998) and on wages in future firms (e.g. Loewenstein and Spletzer 1998; Booth and Bryan 2005). These results are consistent with our interpretation that general training is more likely to support the career of employees than firm-specific training.

Finally, it should be mentioned that we have not explicitly addressed the problem of different time durations of the training courses in our study. Since we examined the wage effects of continuous training that took place within a time span of three years, it would be interesting to obtain more precise information about the returns to training courses with a different length of time spell. In the present study, we were not able to address this problem adequately because our focus lied on disentangling the wage effects of general from firm-specific training. Additionally accounting for the length of time spells of general and firm-specific training would have substantially reduced our treated individuals. For example, we would have had to estimate the wage effects for specifically trained workers who participated in a course that lasted only one month and compare this effect with the corresponding effect for those workers who participated in a course that lasted six months, and so on. Since our groups of the treated are much smaller than the control groups (e.g. about 1/10 for specific training), examining the wage effects of continuous training for different time spells would surely have contributed to additionally decrease the number of the treated relative to the controls. As a consequence, we will leave an interesting question open for future research.

References

- Abadie, A., Imbens, G.W. (2002): Simple and bias-corrected matching estimators. NBER Technical Working Paper No. 0283, Cambridge, MA.
- Abadie, A., Drukker, D., Leber Herr, J., Imbens, G.W. (2004): Implementing matching estimators for average treatment effects in Stata. *Stata Journal* 1, 1-18.
- Acemoglu D., Pischke J.-S. (1998): Why do firms train? theory and evidence. *Quarterly Journal of Economics* 113, 79-119.
- Acemoglu D., Pischke J.-S. (1999a): Beyond Becker: training in imperfect labour markets. *Economic Journal* 109, F112-F142.
- Acemoglu D., Pischke J.-S. (1999b): The structure of wages and investment in general training. *Journal of Political Economy* 107, 539-572.
- Arulampalam, W.; Booth, A. (2001): Learning and earning: Do multiple training events pay? A decade of evidence from a cohort of young British men. *Economica* 68, 379-400.
- Ashenfelter, O. (1978): Estimating the effect of training program on earnings. *Review of Economics and Statistics* 60, 47-57.
- Ashenfelter, O., Card, D. (1985): Using the longitudinal structure of earnings to estimate the effect of training programs. *Review of Economics and Statistics* 67, 648-660.
- Becker, G.S. (1962): Investment in human capital: a theoretical analysis. *Journal of Political Economy* 70 (Supplement), 9-49.
- Bishop, J. (1997): What we know about employer-provided training: a review of the literature. *Research in Labor Economics* 16, 19-87.
- Blundell, R., Costa Dias, M. (2002): Alternative approaches to evaluation in empirical microeconomics. *Portuguese Economic Journal* 1, 91-115.
- Blundell, R., Dearden, L., Sianesi, B. (2005): Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey. *Journal of the Royal Statistical Society Series A* 168, 473-512.
- Blundell, R., Dearden, L., Meghir, C. (1996): The determinants and effects of work-related training in Britain. London: Institute of Fiscal Studies.
- Blundell, R., Dearden, L., Meghir, C. (1999): Work-related training and earnings. Mimeo, Institute of Fiscal Studies.

- Booth, A. (1991): Job-related formal training: who receives it and what is it worth? *Oxford Bulletin of Economics and Statistics* 53, 281–294.
- Booth, A. (1993): Private sector training and graduate earnings. *Review of Economics and Statistics* 75, 164–170.
- Booth, A.L., Bryan, M.L. (2005): Testing some predictions of human capital theory: new training evidence from Britain. *Review of Economics and Statistics* 87, 391–394.
- Büchel, F., Pannenberg, M. (2004): Berufliche Weiterbildung in West- und Ostdeutschland – Teilnehmer, Struktur und individueller Ertrag. *Zeitschrift für Arbeitsmarktforschung* 37, 73–126.
- Budria, S., Pereira, P.T. (2004): On the returns to training in Portugal. IZA Discussion Paper No. 1429, Bonn.
- Caliendo, M., Hujer, R. (2006): The microeconomic estimation of treatment effects – an overview. *Journal of the German Statistical Society / Allgemeines Statistisches Archiv* 90, 197–212.
- Caliendo, M., Kopeinig, S. (2006): Some practical guidance for the implementation of propensity score matching. Forthcoming in: *Journal of Economic Surveys*.
- Cameron, A.C., Trivedi, P.K. (2005): Microeconometrics: methods and applications. Cambridge et al.: Cambridge University Press.
- Düll, H., Bellmann, L. (1999): Der unterschiedliche Zugang zur betrieblichen Weiterbildung nach Qualifikation und Berufsstatus: Eine Analyse auf der Basis des IAB-Betriebspanels 1997 für West- und Ostdeutschland. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* 32, 70–84.
- Frazis, H.; Loewenstein, M. (2003): Re-examining the returns to training: functional form, magnitude, and interpretation. Working Paper 367, Bureau of Labor Statistics.
- Garloff, A., Kuckulenz, A. (2006): Training, mobility, and wages: specific versus general human capital. *Jahrbücher für Nationalökonomie und Statistik* 226/1, 55–81.
- Gerfin, M. (2004): Work-related training and wages: an empirical analysis for male workers in Switzerland. IZA Discussion Paper No. 1078, Bonn.
- Gerfin, M., Leu, R.E., Nyffeler, R. (2003): Berufliche Weiterbildung in der Schweiz. Diskussionspapier 03-18, Universität Bern, Volkswirtschaftliches Institut.

- Goux, D., Maurin, E. (2000): Returns to firm-provided training: Evidence from French worker-firm matched data. *Labour Economics* 7, 1–19.
- Haisken-DeNew, J., Frick, J.(2005): DTC – Desktop Companion to the German Socio Economic Panel (GSOEP), Deutsches Institut für Wirtschaftsforschung, Berlin.
- Heckman, J., LaLonde R. J., Smith, J.A. (1999): The economics and econometrics of active labor market programs. In: Ashenfelter, O., Card. D. (eds.): *Handbook of Labor Economics*, Vol. 3A. Amsterdam, Elsevier, 1865-2097.
- Heckman, J., Ichimura, H., Smith, J.A., Todd, P. (1998): Characterizing selection bias using experimental data. *Econometrica* 65, 1017-1098.
- Heckman, J., Ichimura, H., Todd, P. (1997): Matching as an econometric evaluation estimator: evidence from evaluating a job training program. *Review of Economic Studies* 64, 605-654.
- Jürges, H., Schneider, K. (2006): Dynamische Lohneffekte beruflicher Weiterbildung. In: Weiss, M. (Hrsg.): *Tagungsband des Bildungsökonomischen Ausschusses des Verein für Socialpolitik 2004/2005*. Berlin: Duncker & Humblodt.
- Kuckulenz, A., Maier, M. (2006): Heterogeneous returns to training. *Jahrbücher für Nationalökonomie und Statistik* 226/1, 24-40.
- Kuckulenz, A., Zwick, T. (2003): The impact of training on earnings – differences between participant groups and training forms. ZEW Discussion Paper No. 03-57, Mannheim.
- Kuckulenz, A., Zwick, T. (2005): Heterogene Einkommenseffekte betrieblicher Weiterbildung. *Die Betriebswirtschaft* 65, 258-275.
- Krueger, A., Rouse, C. (1998): The effect of workplace education on earnings, turnover, and job performance. *Journal of Labor Economics* 16, 61-94.
- Lechner, M. (1998): *Training the East German labor force: microeconomic evaluations of continuous vocational training after unification*. Heidelberg, Physica.
- Leuven, E. (2004): A review of the wage returns to private sector training. Paper presented at the EC-OECD Seminar on Human Capital and Labour Market Performance: Evidence and Policy Challenges. Brussels, December 8th.
- Leuven, E. (2005): The economics of private sector training: a survey of the literature. *Journal of Economic Surveys* 19, 91-111.

- Leuven, E., Oosterbeek, H. (2002): A new approach to estimate the wage returns to work-related training. IZA Discussion Paper No. 526, Bonn.
- Leuven, E., Sianesi, B. (2003): PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. <http://ideas.repec.org/c/boc/bocode/s432001.html>.
- Loewenstein, M.A., Spletzer, J.R. (1998): Dividing the costs and returns to general training. *Journal of Labor Economics* 16, 142-171.
- Loewenstein, M.A., Spletzer, J.R. (1999): General and specific training: evidence and implications. *Journal of Human Resources* 34, 710-733.
- Lynch, L. (1992): Private sector training and the earnings of young workers. *American Economic Review* 82, 299-312.
- Mincer, J. (1974): Schooling, experience, and earnings. New York: National Bureau of Economic Research.
- Mooney, C.Z., Duval, R.D. (1993): Bootstrapping: a nonparametric approach to statistical inference. Newbury Park, CA: Sage.
- Pischke, J.-S. (2001): Continuous training in Germany. *Journal of Population Economics* 14, 523-548.
- Pannenberg, M. (1997): Financing on-the-job-training: shared investment or promotion based systems? Evidence from Germany. *Zeitschrift für Wirtschafts und Sozialwissenschaften* 117, 525-543.
- Parent, D. (1999): Wages and mobility: the impact of employer-provided training. *Journal of Labor Economics* 17, 298-317.
- Parent, D. (2003): Employer-supported training in Canada and its impact on mobility and wages. *Empirical Economics* 28, 431-459.
- Rosenbaum, P., Rubin, D. (1983): The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 41-50.
- Schøne, P. (2002): Why is the return to training so high? Mimeo, Institute for Social Research, Oslo, Norway.
- Schøne, P. (2004): Firm-financed training: Firm specific or general skills? *Empirical Economics* 29, 885-900.

- Smith, H (1997): Matching with multiple controls to estimate treatment effects in observational studies. *Sociological Methodology* 27, 325-353.
- Veum, J. R. (1995): Sources of training and their impact on wages. *Industrial and Labor Relations Review* 48, 812–826.

Appendix:

Table A1: Description of variables

Variable	Description
Dependent variable	
Log wage ($\ln W$)	Natural logarithm of individual's recent gross earnings
Individual characteristics	
Years of education (S)	Years of individual's education
Unmarried ($UNMARRIED$)	Dummy variable indicating whether or not individual is unmarried
Age (AGE)	Age of the individual
West-German ($WEST$)	Dummy variable indicating whether or not individual currently lives in West-Germany
Nationality ($NATION$)	Dummy variable indicating whether or not individual is a German citizen
Job-related factors	
Continuous training (CT)	Dummy variable indicating whether or not individual has participated in continuous training within the past three years
Continuous general training (GT)	Dummy variable indicating whether or not individual has participated in continuous general training within the past three years
Continuous firm-specific training (FST)	Dummy variable indicating whether or not individual has participated in continuous firm-specific training within the past three years
Previous continuous training (PT)	Dummy variable indicating whether or not individual has participated in continuous training within the three years prior to 2000.
Fixed-term Employment (FTE)	Dummy variable indicating whether individual is temporarily or permanently employed
Tenure (TEN)	Years of individual's intra-firm job tenure
Overtime work (O_TIME)	Number of hours of overtime work per month
Adequate job (AD_JOB)	Dummy variable indicating whether or not individual is working in his original job
Occupational status	Six dummy variables for managers (MAN), highly skilled white collar workers ($HSWC$), skilled white collar workers (SWC), skilled blue collar workers (SBC), unskilled white collar workers (UWC), and unskilled blue collar workers (UBC)
Regional dummies	Sixteen dummy variables indicating the federal state in which the household resides
Sector dummies	Eleven dummy variables indicating the sector affiliation of the firm
Firm size dummies	Six dummy variables indicating different firm size classes (< 5 employees, 5-19 employees, 20-99 employees, 100-199 employees, 200-1,999 employees, 2,000 and more employees)

Source: German Socio Economic Panel (waves 2000, 2004), own calculations.

Table A2: Descriptive statistics of the variables

Variable	Min-Max	Mean	Std. dev.
Gross monthly wage (W)	720-13,600	3,140.81	1,523.31
Log wage ($\ln W$)	6.58-9.52	7.96	0.43
Continuous training (CT)	0-1	0.24	0.43
Continuous general training (GT)	0-1	0.16	0.37
Continuous firm-specific training (FST)	0-1	0.08	0.28
Age (AGE)	23-65	43.34	8.83
Years of education (S)	7-18	12.19	2.49
Unmarried ($UNMARRIED$)	0-1	0.16	0.37
Nationality ($NATION$)	0-1	0.89	0.31
Tenure (TEN)	0-51.4	12.47	9.67
Adequate job (AD_JOB)	0-1	0.57	0.50
Fixed-term Employment (FTE)	0-1	0.03	0.18
Overtime work (O_TIME)	0-23.1	2.98	4.06
Managers (MAN)	0-1	0.03	0.17
High skilled white collar workers ($HSWC$)	0-1	0.22	0.41
Skilled white collar workers (SWC)	0-1	0.17	0.38
Skilled blue collar workers (SBC)	0-1	0.37	0.48
Unskilled blue collar workers (UBC)	0-1	0.17	0.37
Unskilled white collar workers (UWC)	0-1	0.04	0.19
West-German ($WEST$)	0-1	0.78	0.42

Note: The calculations are restricted to individuals who do not provide item non-responses for the subsequent regression analysis. The sample size is $N = 1,751$.

Source: German Socio Economic Panel (wave 2004), own calculations.

Table A3: OLS estimates for the explanatory variables not displayed in Table 2

Dependent Variable	ln <i>W</i>	
<i>CONST</i>	6.961***	(106.06)
<i>UNMARRIED</i>	-0.080***	(-4.63)
<i>NATION</i>	0.012	(0.62)
<i>WEST</i>	0.364***	(3.47)
<i>FTE</i>	-0.096***	(-2.78)
<i>AD_JOB</i>	0.005	(0.35)
<i>O_TIME</i>	0.015***	(8.02)
Sector dummies	yes	
Occupational status		
Manager (<i>MAN</i>)	0.862***	(15.10)
High skilled white collar worker (<i>HSWC</i>)	0.501***	(16.76)
Skilled white collar worker (<i>SWC</i>)	0.257***	(10.02)
Skilled blue collar worker (<i>SBC</i>)	0.118***	(5.81)
Unskilled white collar worker (<i>UWC</i>)	0.119***	(2.57)
Firm size		
5 to 19 employees	0.074**	(2.04)
20 to 99 employees	0.119***	(3.02)
100 to 199 employees	0.189***	(4.73)
200 to 1,999 employees	0.212***	(5.79)
2,000 and more employees	0.260***	(6.92)
Regional dummies	yes	

Note: */**/** indicates significance at the 10/5/1 percent level. The coefficients are estimated by OLS. The values in parentheses represent the robust *t*-statistics using heteroscedasticity-consistent standard errors of the coefficients. The estimates refer to equation (1), Panel B. The reference group for occupational status is the group of unskilled blue collar workers (*UBC*). The reference group for the firm size dummies are small establishments employing less than 5 workers.

Source: German Socio Economic Panel (wave 2004), own calculations.

Table A4: Probit estimates of the propensity scores

Training form	Continuous training		General training		Firm-specific training	
Dependent variable	<i>CT</i>		<i>GT</i>		<i>FST</i>	
<i>PT</i>	0.992***	(12.91)	1.037***	(11.90)	0.771***	(7.26)
<i>S</i>	0.037**	(2.04)	0.048**	(2.36)	0.009	(0.35)
<i>AGE</i>	-0.021***	(-4.64)	-0.022***	(-4.37)	-0.014**	(-2.24)
<i>SK</i>	0.314***	(2.78)	0.331**	(2.49)	0.240	(1.55)
<i>HS</i>	0.652***	(4.52)	0.610***	(3.64)	0.643***	(3.23)
<i>NATION</i>	0.215	(1.57)	0.112	(0.72)	0.307	(1.53)
<i>CONST</i>	-1.653***	(-4.49)	-2.179***	(-4.76)	-1.682***	(-3.52)
Firm size dummies	yes	yes	yes	yes	yes	yes
Regional dummies	yes	yes	yes	yes	yes	yes
LR test	411.17***		322.77***		184.49***	
Pseudo- R^2	0.212		0.218		0.194	
<i>N</i>	1,751		1,604		1,443	

Note: */**/** indicates significance at the 10/5/1 percent level. The coefficients are estimated by the probit maximum likelihood method. The values in parentheses represent the z-statistics. The variable *SK* is a dummy variable indicating skilled work, which is formed by the two variables *SWC* and *SBC*. The variable *HS* is a dummy variable indicating high skilled work, which is formed by the two variables *HSWC* and *MAN*. The reference category is *US* formed by *USBC* and *USWC*. The LR (Likelihood Ratio) test and Pseudo- R^2 indicate the estimation quality of the equations. *N* indicates sample size. The number of observations for firm-specific training declines by 29 observations compared to the number given in Table 2 as one of the regional dummies causes perfectly predicted failure. The explanatory variables are described in tables A1 and A2.

Source: German Socio Economic Panel (waves 2000, 2004), own calculations.

Table A5: Matching estimators for an alternative specification of the dependent variables general and firm-specific training

Training form		General training	Firm-specific training
Treatment variable		<i>GT</i>	<i>FST</i>
Matching estimator			
Nearest Neighbour	ATE	0.055 (0.077)	0.092 (0.147)
	ATT	0.081* (0.048)	0.128 (0.095)
MDiD		0.062** (0.028)	0.059 (0.040)

Note: */**/** indicates significance at the 10/5/1 percent level. The values in parentheses represent the (bootstrapped) standard errors of the coefficients. In these specifications, *GT* is defined as continuous training that is completely transferable to other employers, while *FST* is defined as continuous training that is not transferable at all. Treated observations are 106 for general training and 32 for firm-specific training. The number of observations *N* in the nearest neighbour matching is 1,402 (general training) and 1,301 (firm-specific training), respectively. For the MDiD estimations *N* is 358 (general training) and 122 (firm-specific training), respectively.

Source: German Socio Economic Panel (waves 2000, 2004), own calculations.

Table A6: Matching estimators for different forms of continuous training on wages using information from the least recent training

Training form		Continuous training	General training	Firm-specific training
Treatment variable		<i>CT</i>	<i>GT</i>	<i>FST</i>
Matching estimator				
Nearest Neighbour	ATE	0.040 (0.026)	0.070* (0.037)	0.051 (0.050)
	ATT	0.018 (0.036)	0.048 (0.040)	0.027 (0.051)
Kernel	ATE	0.053** (0.024)	0.090*** (0.034)	0.064* (0.038)
	ATT	0.042* (0.023)	0.049** (0.025)	0.051 (0.033)
Local Linear Regression	ATE	0.042** (0.020)	0.058* (0.031)	0.061 (0.048)
	ATT	0.033 (0.023)	0.035 (0.024)	0.025 (0.033)
MDiD		0.044*** (0.016)	0.049** (0.020)	0.044* (0.026)
Treated		306	199	107
<i>N</i>		1,631	1,524	1,422

Note: */**/** indicates significance at the 10/5/1 percent level. The values in parentheses represent the (bootstrapped) standard errors of the coefficients. Treated observations on the common support are indicated in the table. *N* indicates the sample size. For further explanations see the comments on tables 4 and 5. Referring to the MDiD estimates *N* is 807 (continuous training), 588 (general training) and 387 (firm-specific training), respectively.

Source: German Socio Economic Panel (waves 2000, 2004), own calculations.

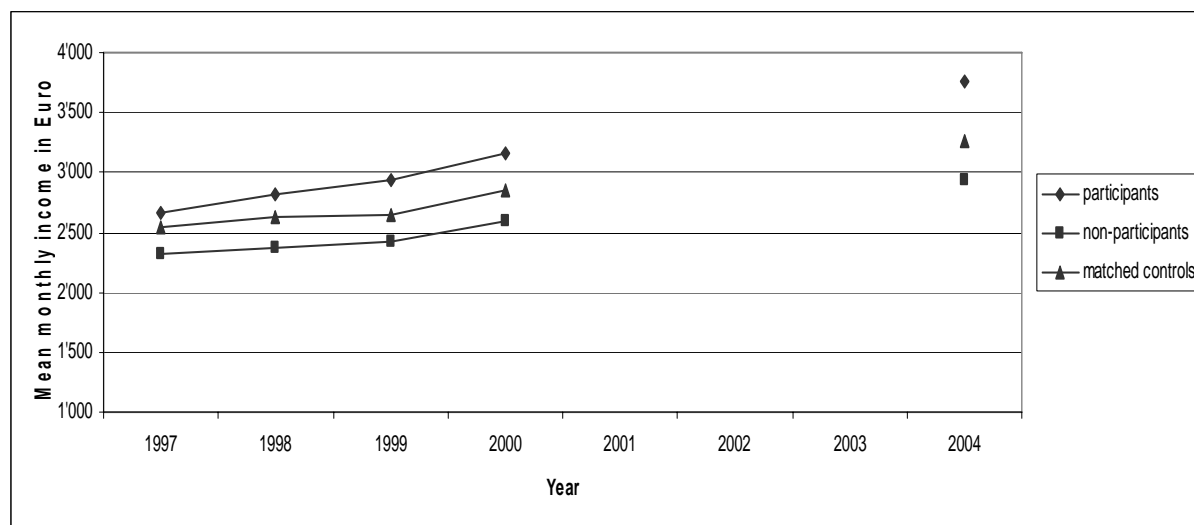
Table A7: MDiD estimators resulting from different matching procedures

Training form	Continuous training	General training	Firm-specific training
Treatment variable	<i>CT</i>	<i>GT</i>	<i>FST</i>
Kernel	0.058*** (0.013)	0.064*** (0.015)	0.043** (0.021)
Local Linear Regression	0.033* (0.018)	0.048** (0.020)	0.035 (0.029)
<i>N</i> (Kernel / LLR)	1,746 / 673	1,587 / 470	1,357 / 265

Note: */**/** indicates significance at the 10/5/1 percent level. The values in parentheses represent the standard errors of the coefficients. *N* indicates sample size.

Source: German Socio Economic Panel (waves 2000, 2004), own calculations.

Figure 1: Yearly wage development of training participants, non-participants and matched controls



Note: The matched non-participants are drawn from the nearest neighbour matching for CT, allowing for replacement and using the four closest matches.

Source: German Socio Economic Panel (waves 1997-2004), own calculations.

Table A8: Yearly wage development of training participants, non-participants and matched controls

Year	Participants		Non-participants		Matched controls	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
1997	2,667	1,225	2,314	946	2,540	1,084
1998	2,813	1,241	2,374	1,015	2,631	1,183
1999	2,945	1,318	2,415	1,013	2,651	1,143
2000	3,166	1,356	2,592	1,155	2,855	1,298
2004	3,753	1,584	2,944	1,450	3,261	1,615

Note: The matched non-participants are drawn from the nearest neighbour matching for CT, allowing for replacement and using the four closest matches.

Source: German Socio Economic Panel (waves 1997-2004), own calculations.