

Discussion Paper No. 04-67

**Using Methods of Treatment Evaluation  
to Estimate the Wage Effect of IT Usage**

Alexandra Spitz

**ZEW**

Zentrum für Europäische  
Wirtschaftsforschung GmbH

Centre for European  
Economic Research

Discussion Paper No. 04-67

## **Using Methods of Treatment Evaluation to Estimate the Wage Effect of IT Usage**

Alexandra Spitz

Download this ZEW Discussion Paper from our ftp server:

**<ftp://ftp.zew.de/pub/zew-docs/dp/dp0467.pdf>**

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

---

Discussion Papers are intended to make results of ZEW research promptly available to other economists in order to encourage discussion and suggestions for revisions. The authors are solely responsible for the contents which do not necessarily represent the opinion of the ZEW.

## Non-technical Summary

The widening wage structure in most industrialized countries in the 1980s has often been attributed to the impact of skill-biased technological change (SBTC), founded upon the notion that employer's demand for more skilled workers has been brought about by the extensive spread of information technologies (IT) at the workplace. The idea is that IT increase productivity, but that only some employees possess the necessary skills to use them. Employees who use IT at the workplace then receive a wage payoff for their higher productivity owing to the use of IT.

The exercise of estimating the wage effect of IT use is one of the classical problems of selection bias. Employers choose employees to use IT or employees themselves choose to use IT, hence, the assignment of persons to treatment is non-random. It is not clear whether differences in measured wages among employees are due to IT use or due to other factors that result in employees earning different wages even if there were no causal effect of IT use.

In this study, I primarily use two methods: regression based matching and matching on the propensity score. Both approaches rest upon the assumption that, conditional on a set of covariates, the difference in wages for IT users and IT non-users is attributable on average to the use of IT. The data set backs this assumption because it is particularly rich in variables that have been identified as important determinants in the IT use-wage relationship.

I find a robust positive relationship between wages and computer usage in West Germany in 1998/99. Given the extensive and rapid changes that have occurred in the workplace in recent decades, IT users would be worse off had they not started to work with computers. In addition to the methodological issues, this study contributes more recent evidence on the computer wage differential to literature.

# Using Methods of Treatment Evaluation to Estimate the Wage Effect of IT Usage<sup>§</sup>

ALEXANDRA SPITZ\*

September 2004

## Abstract

In this study, I analyze the relationship between IT use and wages in West Germany in 1998/99. I use two estimation approaches: regression based matching and matching on the propensity score. The richness of the data set favors the use of these approaches. The variable of main interest is the average treatment effect for the treated. I find that, given the extensive changes in workplaces that have occurred in recent decades, IT users would be worse off in terms of wages had they not started to use IT.

**JEL–classification:** J30, J31

**Keywords:** Computer wage differential, matching

---

<sup>§</sup>**Acknowledgments:** The paper was written as part of the research project “ICT, Workplace Organization, and Wages Across and Within Skill Groups” commissioned by the Landesstiftung Baden-Württemberg foundation. Helpful comments by Joshua Angrist, Irene Bertschek, Bernd Fitzenberger, Julia Häring, and Thomas Hempell are gratefully acknowledged.

\* ZEW Mannheim, Centre for European Economic Research, Research Group Information and Communication Technologies, P.O. Box 103443, D-68034 Mannheim, Germany, email: [spitz@zew.de](mailto:spitz@zew.de).

# 1 Introduction

The widening wage structure in most industrialized countries in the 1980s has often been attributed to the impact of skill-biased technological change (SBTC), founded upon the notion that employer's demand for more skilled workers has been brought about by the extensive spread of information technologies (IT) at the workplace.<sup>1</sup> The idea is that IT increase productivity, but that only some employees possess the necessary skills to use them. Employees who use IT at the workplace then receive a wage payoff for their higher productivity owing to the use of IT.

The empirical evidence with respect to the relationship between the use of IT and wages on the individual level strongly depends on methodological issues. Cross-sectional analyses typically find significant positive wage markups for IT users. However, these rewards are strongly reduced in studies using panel data methods. As a result, most studies favor the argument that the observed positive correlations are spurious, that is, attributable to unobserved heterogeneity.

The exercise of estimating the wage effect of IT use is one of the classical problems of selection bias, a special case of the more general problem of causal inference from non-experimental data.<sup>2</sup> Employers choose employees to use IT or employees themselves choose to use IT, hence, the assignment of persons to treatment is non-random. It is not clear whether differences in measured wages among employees are due to IT use or due to other factors that result in employees earning different wages even if there were no causal effect of IT use. In the absence of genuine experimental data, some assumptions must be made to account for the problem of selection bias. The choice of assumptions must be justified in an economically convincing way depending on the context and the availability of data.

In this study, I primarily use two methods: regression based matching and matching on the propensity score. Both approaches rest upon the assumption that, conditional on a set of covariates, the difference in wages for IT users and IT non-users is attributable on average to the use of IT.<sup>3</sup> The particular feature of the data set used in this study backs this assumption. It has several advantages over other individual-level data sets:

1. It includes a large number of control variables that have been identified as important determinants of IT usage and wages in previous analysis. In contrast to other individual-level data sets that are usually limited to information about employees, it includes information on employers. This information is important since employers may differ in a way that systematically influences both individual wages and IT

---

<sup>1</sup>Comprehensive reviews of this literature can be found in Katz and Autor (1999), Acemoglu (2002) and Chennells and van Reenen (2002). For a critical assessment see Card and DiNardo (2002). Machin (2001) summarizes and discusses the most prominent points of criticism.

<sup>2</sup>See, for example, Heckman and Robb (1986).

<sup>3</sup>The measure of technology in this study is computer usage, which serves as a proxy for IT usage.

use. It is, for example, very likely that an IT user would be employed at another employer in the counterfactual situation. The data set allows me to take various company characteristics into account, such as company size, industry affiliation, innovation strategy and company performance.

2. It is a sample of West German employees that includes both IT non-users, the potential comparison group, and IT users, the treatment group. They answered identical questionnaires. Wages and all other characteristics are therefore measured in the same way for both groups.
3. As the data set includes information on the regional affiliation of employees and employer characteristics such as company size and company performance, IT users and IT non-users operating in a common economic environment can be compared.

The most interesting parameter is the average treatment effect for the treated ( $ATT$ ), that is, the expected benefit of IT usage on the users themselves, as opposed to the average treatment effect ( $ATE$ ), the expected benefit of IT use on a person drawn randomly from the work force. It is often argued in literature that the latter is not a very useful concept in the case of IT use because computers are productive only in certain occupations (for example, for financial market analysts but not for ballet dancers). I additionally argue that it is important to allow for the possibility that treatment effects differ between individuals. How much IT increases the productivity of employees crucially depends on the tasks they have to perform, their qualification and the characteristics of the company. The results of Entorf and Kramarz (1998), for example, suggest that the increase in productivity is related to the experience employees have with new technologies.

In spite of the above arguments, I will also estimate the  $ATE$ . It provides a benchmark for assessing how important the arguments actually are. In addition, it allows me to compare my results with those found in other studies.

As in most cross-section studies, I find a robust positive relationship between wages and computer usage in West Germany in 1998/99. In addition to the methodological issues, this study contributes more recent evidence on the computer wage differential to literature. I do not interpret the result as causal in the sense that it suffices to give a computer to randomly drawn people of the working population for their wages to rise. Instead, I argue that given the extensive and rapid changes that have occurred in the workplace in recent decades, IT users would be worse off had they not started to work with computers.<sup>4</sup> This interpretation is in the spirit of Entorf, Gollac and Kramarz (1999), who found that computer users in France are less vulnerable to job losses.

This paper is organized in 6 sections. In the next section, I introduce the estimation problem at hand in a slightly more formal way following the program evaluation literature. I also discuss previous empirical results and present the empirical strategy of this study. After a general introduction, I discuss the regression based matching approach and matching on the propensity score. Regression based matching aims at purging the

---

<sup>4</sup>See Spitz (2004) for empirical evidence on how skill requirements in the workplace in West Germany have changed in recent decades.

regression specification from potential correlation between unobservables and IT use. The purpose of matching, on the other hand, is to re-establish the condition of an experiment when no such data are available. The differences in assumptions and interpretation of both methods are also discussed. In Section 3, I present the data set and the variables of interest. Section 4 presents and discusses the estimation results. Section 5 concludes.

## 2 Empirical Strategy

### 2.1 Why Is It So Difficult to Estimate the Returns from IT Usage?

This study focuses on the problem of estimating the wage impact of IT usage. To state this more formally, I designate wages  $Y$ . Wages are assumed to be determined by a set of exogenous variables  $X$ , and by a dummy variable  $D$ , such that  $D_i = 1$  indicates that employee  $i$  uses IT on the job, and  $D_i = 0$  that employee  $i$  does not. Assuming, for the ease of illustration, that the decision to use IT was made in period  $s$ , so that, in each period  $t$ ,

$$Y_{it} = \begin{cases} X_{it}\beta + D_i\alpha_i + U_{it} & \text{if } t > s \\ X_{it}\beta + U_{it} & \text{if } t \leq s \end{cases} \quad (1)$$

where  $\alpha_i$  measures the heterogeneous impact of IT usage on wages. If  $\alpha_i$  is constant across individuals, the impact is homogeneous. The set of parameters  $\beta$  defines the relationship between the observable variables  $X$  and the outcome  $Y$ , and  $U_{it}$  is an error term of mean zero and is assumed to be uncorrelated with  $X$ .

Most of the discussions on the wage effects of IT use evolve because the group of IT users is most probably not random. This situation arises because the decision of individual  $i$  or, more likely, the decision of an employer  $e$  to choose individual  $i$  to use IT is based on personal characteristics  $P$ , workplace characteristics  $W$  and employer characteristics  $E$  that may also affect wages  $Y$ . If this is the case, and the set of observable variables  $X$  is only a subset of characteristics  $P$ ,  $W$  or  $E$ , then the error term  $U$  and the IT use variable  $D$  are very likely to be correlated. Hence, as a consequence of such a non-random assignment, IT use ( $D_i = 1$ ) would be correlated with the error term  $U_{it}$  in the wage equation, which results in the standard econometric approach that regresses  $Y$  on  $X$  and  $D$ , producing biased results.

The decision rule may be described in a very stylized form as a latent single index function  $IN_e(i)$ , which specifies the gain in profitability of employer  $e$  if employee  $i$  (working at  $e$ ) uses IT.  $IN_e(i)$  thus represents the difference between the productivity increases of employer  $e$  owing to the implementation of IT and the higher wages employer  $e$  has to pay to employee  $i$  for the IT usage.  $IN_e(i)$  is a function of observable variables  $X$  and unobservable variables  $V$ :  $IN_e(i) = X\gamma + V$ . The entire set of characteristics  $I$ , which are relevant to the decision about IT usage and wages, are then the combination of variables

$X$  and  $V$ ,  $I = (X, V)$ . Employer  $e$  decides that employee  $i$  uses IT on the job if  $IN_e(i) > 0$ , in which case ( $D_i = 1$ ) is observed. A simple behavioral model is one in which employers select employees into IT usage based on a comparison of the present value of expected profits with and without employee  $i$  using IT:  $IN_e(i) = PV_e(1) - PV_e(0)$ , where  $PV(1)$  is the present value of expected profits if employee  $i$  uses IT on the job and  $PV(0)$  without IT usage.<sup>5</sup>

Previous empirical studies have addressed the issue of non-random assignment of computers in one form or another. All of these studies consider observed differences between computer users and non-users, though there is a large difference with respect to the number of available observables. The methods used to account for unobserved heterogeneity range from including proxies for individual ability into the regression specification to applying panel data methods.

Krueger (1993), who was the first to address the question of whether workers who use computers at work are paid more as a result of their computer skills, uses four empirical strategies to take into account unobserved heterogeneity in cross-section data. First, he considers computer use at home because computer use at home may reveal unobserved characteristics on the basis of which employers may select workers to use computers at work. Including computer usage at home in the regression specification then captures at least some of the unobserved heterogeneity that is associated with computer usage at work. Second, he restricts the sample to narrowly defined occupational groups, for example, secretaries. Third, he includes test score results and parental education in the specification. These variables are commonly used as proxies for individual ability. Fourth, he analyzes the relationship between changes in the proportion of computer users and changes in hourly earnings on the occupational level. The successive inclusion of additional controls attenuates the raw logarithmic wage differential for computer use of more than 30 percent to around 10 percent in the 1980s. However, none of the specifications alters the main result of a significant positive relationship between computer use and wages.

In subsequent years, various studies have presented empirical evidence suggesting that Krueger's results are driven by unobserved heterogeneity. DiNardo and Pischke (1997) make their point by showing that there is also a wage premium for workers using a pencil, calculator or telephone at work or who work while sitting. Entorf and Kramarz (1997) and Entorf, Gollac and Kramarz (1999) use French data and estimate fixed-effects models. They do not find a wage premium for computer usage once they control for time-invariant unobserved individual and company heterogeneity, while the corresponding cross-section results lead to a computer-wage differential of around 16 percent – a common figure in cross-section analyses of the early 1990s. The results indicate that it is the increased productivity owing to the experience employees gain with new technologies that leads to the wage premium rather than the effect of merely using new technologies. Bell (1996), on the other hand, finds that 60 percent of the cross-sectional computer wage premium still remains in first-difference regressions using UK data.

---

<sup>5</sup>For a more comprehensive model of computer assignment see, for example, Borghans and ter Weel (2004).

The panel methods hinge crucially on the assumption of time-invariant unobserved heterogeneity. In the presence of changing returns on unobserved skills, differencing the data will not remove the wage effect of unobservables that might be correlated with computer use. As DiNardo and Pischke (1997) already emphasized, this factor may be a plausible explanation for the differing results of the two studies because the wage structure has widened in the UK since the early 1980s but not in France. In addition, Dolton and Makepeace (2004) criticize the simple panel methods for not allowing the coefficient to vary between individuals or, in other words, for ignoring variation in the parameter values. Exploring information on whether individuals used a computer in 1991 and 2000 (stayers), in 1991 only (leavers) or in 2000 only (enterers), they find, for example, a positive wage effect of 9 percent for male leavers in panel regressions.

As this short review of research demonstrates, previous results are mixed. To my knowledge, none of the studies considers explicitly that IT usage may be interpreted as a treatment. This notion is very attractive for this question of analysis because the main interest lies on the *ATT*. By contrast, previous studies focus on estimating the average wage effect of IT usage.

## 2.2 Estimation Methods

The empirical strategy in this study is to use methods most often applied in the active labor market program evaluation literature. In the tradition of this literature, I interpret IT usage as a treatment  $D_i$ , that is, estimating the coefficient of IT usage translates into a problem of estimating an average treatment effect (*ATE*).<sup>6</sup> The *ATE* is usually estimated using a counterfactual framework (Rubin, 1974), where each individual  $i$  has an outcome with ( $D_i = 1$ ) and without ( $D_i = 0$ ) treatment (conventionally denoted as  $Y_{1i}$  and  $Y_{0i}$  respectively). Since both states are never observed for one person simultaneously, different methods have been developed in literature to overcome this “omitted variable” problem depending on the specific circumstances of the question of interest.

The observed outcome may be expressed as:

$$Y_i = Y_{0i}(1 - D_i) + Y_{1i}D_i = Y_{0i} + D_i(Y_{1i} - Y_{0i}). \quad (2)$$

This body of literature distinguishes between the average treatment effect (*ATE*),  $E(Y_{1i} - Y_{0i})$ , that is, the expected effect of treatment on a person drawn randomly from the population, and the average treatment effect on the treated (*ATT*),  $E(Y_{1i} - Y_{0i} | D_i = 1)$ , that is, the expected effect of treatment of those who actually have been treated. Since  $E(Y_{1i} | D_i = 1)$  is directly identifiable from the sample, the problem of estimating the *ATT* is equivalent to the problem of estimating the counterfactual average,  $E(Y_{0i} | D_i = 1)$ . Thus, in order to be able to estimate the average non-treatment outcome for treated employees based on the outcomes for non-treated, identifying assumptions are needed. Estimating the *ATE*, on the other hand, involves estimating both counterfactual averages,

---

<sup>6</sup>*ATE* is an average partial effect in the case of a binary variable.

$E(Y_{0i}|D_i = 1)$  for the non-treated workers in the sample and  $E(Y_{1i}|D_i = 0)$  for the treated workers in the sample. In the case of IT usage, the effect on the subpopulation of treated employees (*ATT*) is probably of more interest than the effect on the overall population (*ATE*). As already discussed earlier in this paper, one of the reasons is that computers are of value only in certain jobs, for example, for computer scientists but not for artistic painters. The overall focus of the analysis will therefore be on the *ATT*. However, for ease of comparison with previous analyses and for the purpose of having a benchmark, I will also estimate the *ATE*.

The observed difference in outcomes between those treated and those untreated equals  $E(Y_{1i} - Y_{0i}|D_i = 1)$  plus a bias term:

$$\begin{aligned} E(Y|D_i = 1) - E(Y|D_i = 0) &= E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 0) & (3) \\ &= E(Y_{1i} - Y_{0i}|D_i = 1) & (4) \\ &+ \underbrace{[E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)]}_{\text{bias term}}. \end{aligned}$$

If IT use were randomly assigned to employees, that is, if the treatment and control group had the same distribution of both observed and unobserved characteristics, this would be an easy task: the *ATE* and *ATT* would be the difference in sample average of those using IT and those not using IT at the workplace.<sup>7</sup> However, this scenario is very unlikely, in particular as recent empirical studies document that IT users differ systematically from IT non-users, for example, with respect to their education. Table 9 shows the non-randomness of IT usage based on the data set at hand. The major differences are in the distribution of formal education between IT users and IT non-users. In contrast to IT being randomly assigned, employers select employees with specific favorable traits for treatment. The decision about who uses IT on the job is related to the question of how much an employer’s profitability increases from IT use and, hence, to the outcomes  $Y_{1i}$  and  $Y_{0i}$ .

Since data from a social experiment or instrumental variables that allow purging of the regression from the presence of a possible omitted variable bias are not available for this research project, my strategy is based on the assumption of “strict ignorability” of treatment conditional on a set of controls (Rosenbaum and Rubin, 1983):  $(Y_{1i}, Y_{0i}) \perp\!\!\!\perp D_i|X$ . That is, conditional on  $X$ ,  $D_i$  and  $(Y_{1i}, Y_{0i})$  are independent, where  $X$  denotes a vector of observable covariates,  $D_i$  denotes the treatment (in our case  $D_i = 1$  : IT use,  $D_i = 0$  : no IT use) and  $Y_{1i}$  is the logarithm of hourly wages for IT users and  $Y_{0i}$  for IT non-users. This assumption implies that  $E(Y_{0i}|X, D_i) = E(Y_{0i}|X)$  and  $E(Y_{1i}|X, D_i) = E(Y_{1i}|X)$ , referred to as the conditional mean independence assumption (CIA) in literature. That is, conditional on  $X$ , the non-treatment outcome for the treated and non-treated are

---

<sup>7</sup>An additional assumption is that there are no general equilibrium effects in the sense that IT use by one employee does not affect wages of other employees (so-called stable unit treatment value assumption: SUTVA; see, for example, Angrist, Imbens and Rubin, 1996). This is synonymous to stating that there are no macroeconomic effects. The analysis in this study is based on a random sample, which implies SUTVA.

comparable in expectation. For the estimators used in this study it suffices that this weaker assumption holds.

The intuition behind this assumption is that systematic differences in wages between IT users and IT non-users with the same values for the covariates  $X$  are attributable to the use of IT. Since the moments of the distribution of  $Y_{1i}$  for the treated can be estimated directly, it suffices for the assumption  $E(Y_{0i}) \perp D_i | X$  to hold in order to be able to estimate the *ATT*.

The present data set is particularly suited for this strategy because it contains a large number of observables that have been identified as important determinants of the IT use-wage relationship in earlier studies. In spite of this merit, one must keep in mind that the CIA is a strong assumption and that, strictly speaking, the estimation strategy accounts solely for selection on observables. Differences between IT users and IT non-users in the distribution of unobserved attributes such as ability are not considered. Heckman, Ichimura, Smith and Todd (1998b) decompose the bias term of equation 3 into three components:

$$\text{bias term} = [E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)] = B_1 + B_2 + B_3 \quad (5)$$

The first two components,  $B_1$  and  $B_2$ , arise due to differences in the distribution of observed characteristics  $X$  whereas the last component,  $B_3$ , refers to the “classical” selection bias resulting from selection on unobservable attributes.  $B_1$  corresponds to the bias due to differences in the support of regressors between IT users and IT non-users.  $B_2$  corresponds to the bias owing to differences in the shape of distributions of regressors in the two groups in the region of common support.

I use two methods to overcome the bias due to  $B_2$ , regression based matching and matching on the propensity score. The two approaches differ in how they deal with the bias due to  $B_1$ , which I will discuss in Section 2.3. None of them is able to resolve problems due to the presence of  $B_3$ . However, Heckman, Ichimura and Todd (1997) and Heckman, Ichimura, Smith and Todd (1998b) find that the selection bias part,  $B_3$ , is the smallest of the three components of bias in their application. How important the different bias components are depends, of course, on the application. However, the richness of the data set in the present study largely reduces the scope for a potential bias owing to unobservables.

Both approaches have already been widely discussed in literature (see, for example, Heckman and Robb, 1985, Heckman, Ichimura and Todd, 1997 and 1998a, or Wooldridge, 2002, Chapter 18). I will review the basic concepts in order to demonstrate their usefulness for the estimation problem at hand.

### 2.2.1 Regression Based Matching

The aim of regression based matching (RBM) is to purge the outcome equation from potential correlation between the error term  $U_{it}$  and  $D_i$  (see equation 1). The basic logic

of the RBM is analogous to the proxy solution to the omitted variable problem.<sup>8</sup>

To illustrate this approach it is useful to decompose the counterfactual outcomes for IT user  $Y_{1i}$  and IT non-user  $Y_{0i}$  into their means and a stochastic part:<sup>9</sup>

$$Y_{0i} = \mu_0 + U_{0i} \quad E(U_{0i}) = 0 \quad (6)$$

$$Y_{1i} = \mu_1 + U_{1i} \quad E(U_{1i}) = 0 \quad (7)$$

where  $\mu_j = E(Y_{ji})$ ,  $j = 0, 1$ . As in the previous part of the paper,  $i$  refers to individuals and  $j$  to the treatment status. For the case of linear regressions  $\mu_j = X\beta_j$ . Inserting these into equation (2) of observed outcomes gives:

$$Y_i = \mu_0 + (\mu_1 - \mu_0)D_i + U_{0i} + (U_{1i} - U_{0i})D_i \quad (8)$$

Under the CIA assumption and assuming that  $E(U_{1i}|X) = E(U_{0i}|X)$  it follows that standard regression methods can be used to estimate ATE. This is because equation (8) may be rewritten as:

$$E(Y_i|D_i, X) = \mu_0 + (\mu_1 - \mu_0)D_i + E(U_{0i}|X) + \underbrace{E(U_{1i} - U_{0i}|X)}_{=0} D_i \quad (9)$$

$$= \mu_0 + \alpha D_i + g_0(X) \quad (10)$$

where  $\alpha = \mu_1 - \mu_0 = E(Y_{1i} - Y_{0i}|X) = ATE = ATT$  and  $g_0(X) = E(U_{0i}|X)$ . In addition,  $E(U_{0i}|X)$  might be expressed as a function of some vector function  $h_0(X)$ , such as  $E(U_{0i}|X) = \eta_0 + h_0(X)\beta_0$ . Then, equation (9) translates into

$$E(Y_i|D_i, X) = \gamma_0 + \alpha D_i + h_0(X)\beta_0 \quad (11)$$

where  $\gamma_0 = \mu_0 + \eta_0$ .  $h_0(X)\beta_0$  is an example of a control function and its purpose is to provide an approximation for  $E(U_{0i}|X)$ . When inserted into equation (9) and therefore implicitly subtracted from  $U_{0i}$ , it purges the equation from the covariance between  $D_i$  and  $U_{0i}$ . The idea is that the purged disturbance term  $[U_{0i} - h_0(X)\beta_0]$  is orthogonal to

---

<sup>8</sup>The omitted variable bias can be mitigated, or even eliminated, if a proxy variable is available for an unobserved variable  $q$ . The formal requirements for a proxy variable can be found, for example, in Wooldridge (2002, p.63ff.). The intuition is that, conditional on the proxy variable, the unobservable variable  $q$  is uncorrelated to each of the observable variables  $X$  in the specification.

<sup>9</sup>The structure of illustration follows Wooldridge (2002), chapter 18.

all variables on the right-hand side of the equation. If the assumptions hold, the control function accounts for a possible self-selection bias.<sup>10</sup>

The previous specification rests upon the assumption that there are on average no person-specific gains conditional on  $X$ ,  $E(U_{1i} - U_{0i}|X) = 0$ . Relaxing this assumption implies that  $ATE$  and  $ATT$  are no longer identical. However, a regression specification can still be used:

$$E(Y_i|D_i, X) = \mu_0 + \alpha D_i + g_0(X) + D_i[g_1(X) - g_0(X)] \quad (12)$$

where  $\alpha = ATE$ ,  $g_0(X) \equiv E(U_{0i}|X)$  and  $g_1(X) \equiv E(U_{1i}|X)$ .  $g_0(X)$  and  $g_1(x)$  are both operationalized by replacing them with a parametric function of  $X$ . The intuition is analogous to the approach in equation (11).

Assuming that these functions are both linear in  $X$ , equation (12) can be written as

$$E(Y_i|D_i, X) = \gamma_0 + \alpha D_i + X\beta_0 + D_i[X - E(X)]\delta \quad (13)$$

where  $\beta_0$  and  $\delta$  are unknown parameters (the intermediate steps are shown in the appendix to this chapter). As can be seen, the control function then involves both the level effect of  $X$  and the interactions of  $X$  (which have been previously demeaned by  $E(X)$ , approximated with the sample averages) with the treatment indicator. Based on the  $\hat{ATE}$ , the  $ATT$  is estimated using<sup>11</sup>

$$\hat{ATT} = \hat{\alpha} + \left(\sum_{i=1}^N D_i\right)^{-1} \left(\sum_{i=1}^N D_i(X - \bar{X})\hat{\delta}\right). \quad (14)$$

I take account of the fact that the demeaned variables are generated regressors by applying a bootstrap method to construct the standard errors of the estimated coefficients. All bootstrap results are based on 100 resamples.

---

<sup>10</sup>This is a special case of a control function that does not rely on instruments. For a discussion on the control function approach involving instruments refer, for example, to Heckman and Robb (1985) or Blundell, Dearden and Sianesi (2003). In contrast to the instrument approach that rests upon the exclusion restriction, identification in the approach used in this study is achieved through a functional form restriction.

<sup>11</sup>A method to estimate the  $ATT$  directly would be to demean the  $X$  using the averages of the “treated” sample instead of the averages of the “whole” sample.

## 2.2.2 Propensity Score Matching

The CIA assumption, introduced in Section 2.2, is the basis of the matching technique. The traditional matching approach estimates the expected non-treatment outcome for an IT user  $i$  with observable characteristics  $X$  by the fitted value of a nonparametric regression in the sample of IT non-users with identical  $X$ . This can be represented by the following formula:

$$ATT = \frac{1}{N_1} \sum_{i \in D_i=1} \{Y_{i1} - \sum_{i \in D_i=0} w_{N_0} Y_{i0}\}, \quad (15)$$

where  $N_1$  is the number of IT users and  $N_0$  is the number of IT non-users.  $w_{N_0}$  is a weight function that weights the observations of an IT non-user according to its similarity with respect to  $X$  to IT user  $i$ . The literature discusses different matching estimators, which differ with respect to the weight function. I will discuss the estimators used in this study at the end of this section. Beforehand, I shortly refer to the so-called “curse-of-dimensionality” problem that arises when the vector of observable characteristics  $X$  is high-dimensional (as it is the case in this study). As the dimension of the data increases and, hence, the information content increases, the complexity of the estimation problem increases exponentially. The “curse-of-dimensionality” arises because in high-dimensional settings, the data requirements needed to find counterfactuals that are similar along every dimension of  $X$  increase exponentially in the dimension of  $X$ .

Rosenbaum and Rubin (1983) introduced the propensity score as a means to overcome the dimensionality problem that arises when the set of conditioning variables  $X$  is large. They showed that when outcomes  $Y_0$  are independent of the treatment conditional on variables  $X$  (CIA assumption), then outcomes  $Y_0$  are also independent of the treatment conditional on the propensity score  $p(X)$ , defined as  $P(D_i = 1|X)$ , with  $0 < P(D_i = 1|X) < 1$ . As can be seen from this formula, the propensity score is the probability of treatment given the covariates  $X$ . In conditional independence notation, this is  $(Y_{1i}, Y_{0i}) \perp\!\!\!\perp D_i | p_i(X)$ . This property allows it to use the one-dimensional propensity score  $P(D_i = 1|X)$  instead of the high-dimensional vector  $X$  in the matching approach.  $P(D_i = 1|X)$  instead of  $X$  now determines the similarity of treated and non-treated employees. Dehejia and Wahba (1999, 2002), among others, have drawn attention to this class of estimators because their analyses suggest that propensity score matching is a powerful tool to resolve the selection bias problem inherent in non-experimental data.<sup>12</sup>

To implement this estimator, one first needs an estimate for the propensity score. I choose a flexible probit model that includes various covariates (described below), their quadratics

---

<sup>12</sup>Others also investigated the properties of matching methods in different settings (see, for example, Heckman, Ichimura and Todd, 1997, 1998a). However, Dehejia and Wahba (1999, 2002) explicitly make the case for matching on the propensity score in order to overcome LaLonde’s criticism of non-experimental estimators (LaLonde, 1986). Criticism of Dehejia and Wahba (1999, 2002) comes from Smith and Todd (2004).

and interactions. However, logit models as proposed by Rosenbaum and Rubin (1983) or non-parametric models may also be used (Powell, 1994, Heckman, Ichimura and Todd, 1997). I take account of the sampling variability in the estimated propensity scores by applying a bootstrap method to estimate the standard errors of the estimated treatment effects. The bootstrap results are based on 100 resamples.

Matching estimators differ with respect to the weights they attach to members of the non-treated group. The traditional pairwise matching method, nearest neighbor matching, weights the nearest non-treated employee with 1 and all others with 0, that is, only the closest member of the comparison group is used as the comparison level for the treated employee  $i$ . In order to guarantee that the “closest” neighbor is not too far away in terms of the propensity score, one usually sets a caliper. In the nearest neighbor matching estimations in this study the caliper is set to 0.001.

Extensions to this method are, for example, kernel or local linear matching estimators (Heckman, Ichimura and Todd, 1997 and Heckman, Ichimura, Smith and Todd, 1998), which have the advantage of reducing the asymptotic mean squared error. In this study, I present results using an Epanechnikov kernel.

## 2.3 Discussion of Methods

In this paper, I present two different approaches for estimating the *ATE* and *ATT*. They both rely on the CIA, but each of them uses different additional assumptions to overcome the omitted variable problem associated with the counterfactual framework. This has implications for the economic interpretation of estimated coefficients. Both methods are, however, similar with respect to the fact that they impute the missing potential outcomes. The data set is particularly suited for these approaches because it includes a large number of potential controls.

None of the approaches considers an instrument variable  $Z$  as a source of exogenous variation to approximate random assignment to IT use. This is the case because I did not find convincing candidates that are correlated with IT use but do not simultaneously determine wages. Therefore, selection models, which are otherwise powerful tools, cannot be applied in this study (see Heckman, 1979, and Heckman and Robb, 1985, 1986).<sup>13</sup> However, having an instrument at hand would have been helpful only in the case of homogeneous treatment effects (see, for example, Heckman and Robb, 1985, Blundell and Costa Dias, 2000, or Blundell, Dearden and Sianesi, 2003).<sup>14</sup> Assuming a homogeneous impact of treatment is always restrictive, but seems to be particularly cumbersome for the question of interest in this study, as previous research suggests that the effects of IT use do vary across individuals (see, for example, Entorf and Kramarz, 1997, 1998, Dolton

---

<sup>13</sup>In contrast to matching and control function approaches, which are considered “selection on observables,” selection models additionally account for selection on unobservables.

<sup>14</sup>An exception is the local average treatment effect interpretation of instrument estimates in the heterogeneous treatment effect case (see Imbens and Angrist, 1994).

and Makepeace, 2004). The fact that the productivity of IT depends on the occupational circumstances favors the argument that the effects of IT usage are heterogeneous across individuals. Employers select those employees into IT usage from which they expect the largest increases in profitability. These expectations rely on workplace characteristics, individual characteristics and company characteristics. Employees share in the gains of IT usage in terms of higher wages.

In the case of heterogeneous treatment effects, the composite error term is given by  $U_i + D_i(\alpha_i - \bar{\alpha})$ . To see this consider the outcome equation 1 (for  $t > s$ ):  $Y_i = \beta + D_i\alpha_i + U_i$  (abstracting from other regressors), where  $\alpha_i$  is the treatment impact of individual  $i$ . Define  $\bar{\alpha}$  as the population mean impact ( $E(\alpha_i) = \bar{\alpha} = ATE$ ),  $\epsilon_i$  as worker  $i$ 's deviation from the population impact, and then  $\alpha_i = \bar{\alpha} + \epsilon_i$ . The outcome equation may then be written as:  $Y_i = \beta + D_i\bar{\alpha} + [U_i + D_i\epsilon_i] = \beta + D_i\bar{\alpha} + [U_i + D_i(\alpha_i - \bar{\alpha})]$ . Therefore,  $U_i + D_i(\alpha_i - \bar{\alpha})$  is the composite error term. Even if the instrument  $Z$  is uncorrelated with  $U_i$ , it is, by assumption, not uncorrelated with  $U_i + D_i(\alpha_i - \bar{\alpha})$ .

Regression assumes a functional form that is linear in parameters. It imputes the missing potential outcome using the estimated regression function. For example, if  $D_i = 1$ , then  $Y_{1i}$  is observed and  $Y_{0i}$  is missing. The regression approach imputes  $Y_{0i}$  with a consistent estimator  $\hat{\mu}_0(X)$  for the conditional expectation. Simple OLS hinges crucially on the assumption  $E(U_{1i} - U_{0i}|X) = 0$  (see equation 9). The resulting estimate will not in general recover the *ATT*. The additional regression specification used in this study, which includes interactions between the observable variables  $X$  and the treatment dummy recovers the *ATT*. In addition, the interaction terms provide evidence of the presence and extent of heterogeneous effects, firstly, through the statistical significance of the estimated coefficients and, secondly, through the quantitative magnitude of effects for various subgroups that might be of particular interest.

Matching, on the other hand, does not assume that the functional form is linear in parameters. It is a non-parametric approach in which a comparison group among the non-treated is chosen, such that the selected group is as similar as possible to the treated group in terms of their observable characteristics. It is a very attractive approach in this study because the main interest lies on the average treatment effect for the treated (*ATT*)<sup>15</sup> and the data set has properties that have been identified as being favorable in the context of the evaluation of active labor market programs (Heckman, Ichimura and Todd, 1997). Matching estimators impute the potential outcome using only the outcome of the nearest neighbor of the opposite treatment group. However, various studies point to the importance of the choice of the matching variables (for example, Blundell, Dearden and Sianesi, 2003).

Both matching and OLS produce weighted averages of the treatment effects  $E(Y_{1i} - Y_{0i}|X)$ . Matching recovers the *ATT* by weighting the heterogeneous effects according to

---

<sup>15</sup>It is also possible to recover the *ATE* with the matching approach. This requires the estimation of the average treatment effect for the non-treated in addition to the average treatment effect on the treated. The *ATE* can then be calculated as a weighted average of the effect on the treated and the effect on the non-treated.

the proportion of treated at each value of  $X$  (that is, proportionally to the propensity score at  $X$ ). OLS, in contrast, weights the heterogeneous effects proportionally to the variance of treatment status at  $X$  (Angrist, 1998).

One issue is whether the support of the observable variables overlaps for the treated and the non-treated group (bias  $B_1$  in equation 5). This problem is particularly evident for matching. If there are regions where the support does not overlap, it is not possible to find a sufficiently comparable observation in the other group. Then matching is often performed only over the common support region.<sup>16</sup> The estimated effect must then be redefined as the mean treatment effect for those treated falling within the common support region, which, in the heterogeneous treatment case, might be quite different from the average treatment effect on the treated. This problem is less severe in the regression approach, because the parametric model can in general be used to predict the expected outcome even in regions outside the common support. However, the reliability of the estimated coefficients then hinges crucially on the ability of the model to predict outside its common support. Regression is not a transparent tool in this respect. Researchers are often not aware of the fact that the observables do not overlap much. The advantage of matching is that it quickly reveals the extent to which treated and non-treated groups overlap in terms of pre-treatment variables.

Neither approach overcomes the potential bias due to unobservables ( $B_3$  in equation 5). If IT use and wages are still positively correlated with some unobserved individual traits, such as ability, both  $ATT$  and  $ATE$  are upward biased. This situation is equivalent to arguing that the CIA does not hold in the present estimation problem. Given the large set of controls that are taken into account in the analysis, it is hard to imagine exactly what the remaining unobservables might be. For example, the specifications control for individual education, which takes into account that high ability persons choose higher levels of education. I therefore argue that the informational richness of the data set justifies the CIA.

### 3 Data and Empirical Framework

The analysis is based on the “Qualification and Career Survey”, which is a survey of employees carried out by the German Federal Institute for Vocational Training (Bundesinstitut für Berufsbildung, BIBB) and the Research Institute of the Federal Employment Service (Institut für Arbeitsmarkt- und Berufsforschung, IAB). I use the most recent cross-section, launched in 1998/99, in this study. It covers more than 30,000 individuals (men and women).<sup>17</sup>

---

<sup>16</sup>For a detailed discussion of this problem see Lechner (2000).

<sup>17</sup>I restrict the sample to West German residents with German nationality: in other words, East German residents and non-German employees are excluded from the sample. Moreover, the sample does not include self-employed, employees with agricultural occupations and employees working in the agricultural sector. Persons younger than 18 or older than 65 are also excluded from the sample.

In addition to the main variable of interest, IT use, three types of variables are considered in the analyses: individual characteristics, company characteristics and workplace characteristics. I include variables reflecting individual characteristics in order to account for the fact that employees may systematically differ with respect to characteristics that may affect both computer use and wages. As more highly educated workers are more likely to use computers at work and earn higher wages, I control for the level of formal education of employees, work experience and tenure with the current employer. As wages of civil servants are determined in a process that differs from the process for wages of employees in private companies, a dummy variable for civil servants is also included in the regressions.

One drawback of most estimates on individual-level data is that they generally do not provide information on the employers. Employer information may however be important if it determines systematic effects on wages and IT use. The data set used in this study allows me to take various company characteristics into account such as company size, industry affiliation, innovation strategy and company performance. This ability is a substantial improvement over other studies in this area. Based on previous empirical research, I expect, for example, that larger companies and innovative companies pay higher wages and that they are more likely to use IT intensively.

Another feature that distinguishes this data set from others is that it includes information on the task composition of occupations.<sup>18</sup> These tasks describe the occupational context in which IT is introduced. In addition, this information on occupational skill requirements allows me to further reduce unobserved heterogeneity.

The variables used in the estimates are constructed as follows (Summary statistics can be found in Table 1):

*Hourly Wages:* The survey contains information on monthly earnings, according to 18 categories. A midpoint is assigned for each category. These midpoints are then divided by the number of hours an individual usually spends at work.<sup>19</sup> Compared to other data sets that are usually used in comparable analyses such as the CPS for the U.S. or the IAB-S for Germany, this data set has the advantage that earnings of highly paid workers are not censored from above. In all estimates, the logarithm of wages is used as dependent variable. On average, employees in West Germany earned about 27 German Marks in 1998/99 (Table 1).

*IT equipment:* The data set includes detailed information on the tools and machines used by the employees in the workplace. The “IT use” variable is a dummy that takes the value 1 if the employee uses a computer, terminal or electronic data processing device on the job. As shown in Table 1, 57 percent of employees have used one or more of the above devices on the job in West Germany in 1998/99.

---

<sup>18</sup>For a detailed analysis of how the task composition of occupations changed in West Germany since 1979 see Spitz (2004).

<sup>19</sup>Comparable procedures are often used in literature, for example, by DiNardo and Pischke (1997) and by Entorf and Kramarz (1997).

*Individual characteristics:* I distinguish three levels of formal education attained by employees. Employees with a low level of education are those with no vocational training. Employees with medium levels of education have a vocational qualification either from an apprenticeship or they have graduated from a vocational college. Employees holding a degree from a university or a technical college are classified as having a high level of education. As shown in Table 1, the majority of the survey participants, 70 percent, has a medium qualification level, whereas 17 percent are highly qualified and only 12 percent have a low education level.

Table 1: SUMMARY STATISTICS

	Mean	Std. Deviation	Min.	Max.	Observations
IT	0.57	0.50	0	1	21816
Pencil	0.92	0.27	0	1	18775
(hourly) wages (in DM)	27.19	11.82	3.13	98.68	18561
<b>Qualification</b>					
high education level	0.17	0.37	0	1	21816
medium education level	0.71	0.46	0	1	21816
low education level	0.12	0.33	0	1	21816
experience	20.76	11.58	0	47	21816
tenure	11.75	9.84	0	47	21816
<b>Workplace Characteristics:</b>					
analytic task measure	14.01	23.80	0	100	18041
interactive task measure	30.26	28.18	0	100	21754
repetitive cognitive task measure	21.73	41.24	0	100	21813
repetitive manual task measure	17.43	30.83	0	100	21768
non-repetitive manual task measure	24.32	24.99	0	50	12319
<b>Company Characteristics</b>					
product innovation	0.37	0.48	0	1	20802
process innovation	0.51	0.50	0	1	20857
very good company performance	0.18	0.39	0	1	14450
good company performance	0.65	0.48	0	1	14450
rather bad company performance	0.14	0.35	0	1	14450
bad company performance	0.03	0.17	0	1	14450
<b>Other Controls</b>					
ever unemployed	0.30	0.46	0	1	22545
married	0.69	0.46	0	1	22677
civil servants	0.11	0.31	0	1	22677
born in East Germany	0.04	0.19	0	1	22677
woman	0.44	0.50	0	1	22677
lives in city	0.38	0.48	0	1	22677

The survey participants also indicate their first year of work. Based on these answers, I calculate the years of (potential) work experience (1999 - first year of work). In addition, employees indicate the year when they started to work for the current employer. This information is used to calculate tenure (1999 - first year with current employer).

The data set includes information about previous unemployment spells (dummy variable: "Have you ever been unemployed before?"), marital status, gender and whether survey participants were born in East Germany. It also contains information about whether

an employee is working as a civil servant. In addition, I constructed a dummy variable indicating whether employees live in a city (place of residence is larger than 100,000 inhabitants).

*Workplace Characteristics:* The analyses by Autor, Levy and Murnane (2003) and Spitz (2004) document how IT has changed the content of work towards analytical and interactive activities and away from manual and cognitive routine activities. The data set at hand is a cross-section, thus changes in the task composition of occupations cannot be taken into account. However, the data set allows me to consider task levels, capturing the content of jobs, and therefore, it gives a description of the context in which IT is used. Survey participants are asked what kind of activities they perform at the workplace. Based on these activities five categories are constructed, which classify the occupational skill requirements: analytic tasks, interactive tasks, repetitive cognitive tasks, repetitive manual tasks and non-repetitive manual tasks. Table 2 shows the list of activities that employees were asked for in the questionnaire and how the activities are classified in the five task categories. On the individual-level  $i$ , the task measures ( $Task_{ik}$ ) are defined as:

$$Task_{ik} = \frac{\text{number of activities in category } k \text{ performed by } i \text{ in 1998/99}}{\text{total number of activities in category } k \text{ in 1998/99}} * 100 \quad (16)$$

where

$$k = \begin{cases} 1 & : \text{ non-routine analytic tasks} \\ 2 & : \text{ non-routine interactive tasks} \\ 3 & : \text{ routine cognitive tasks} \\ 4 & : \text{ routine manual tasks} \\ 5 & : \text{ non-routine manual tasks.} \end{cases}$$

For example, if the analytical task category includes 4 activities and employee  $i$  performs 2 of them, the analytical task measure for employee  $i$  is 50.

The data set also contains information about the current occupation of the employees. Occupations are grouped according to the (2-digit) classification of occupational titles by the Federal Employment Bureau in 1999, leading to 78 occupational groups.

*Company characteristics:* Company size has been identified as an important component of wage determination in previous studies, finding that larger companies pay higher wages to employees with similar characteristics (see, for example, Brown and Medoff, 1989, Schmidt and Zimmermann, 1991). In addition, IT usage increases in company size (see Table 3). Company size measured as the number of employees is captured by 7 size classes. Companies with one to four employees are classified to belong to the first size bracket and companies with more than 1,000 employees to the last one. Based on these size classes,

Table 2: ASSIGNMENT OF ACTIVITIES

Classification	Tasks
non-routine analytic	researching, evaluating and planning, making plans, constructing, designing, sketching working out rules/regulations using and interpreting rules
non-routine interactive	negotiating, lobbying, coordinating, organizing teaching or training selling, buying, advising customers, advertising entertaining or presenting employing or managing personnel
routine cognitive	calculating, bookkeeping correcting of texts/data measuring of length/weight/temperature
routine manual	operating or controlling machines setting up machines
non-routine manual	repairing or renovation houses/apartments/machines/vehicles restoring art/monuments serving or accomodating

7 dummy variables are formed. Most of the survey participants, 28 percent, belong to companies with a size class from 10 up to 49 employees, followed by the size class from 100 up to 499 employees. Companies with more than 1000 employees are represented by 12 percent of the survey participants. About 20 percent of the interviewed employees belong to small companies with less than ten employees.

The data set also includes information about the performance of companies. The survey participants were asked whether the company was doing very good, good, rather bad or bad. For each of these categories, I constructed a dummy variable. Table 1 shows that 18 percent of employees report to work in companies that are doing very well and 65 percent work in companies that are doing well. 17 percent of employees work in companies that are either doing rather bad or bad.

Companies are classified according to 48 detailed industry codes. Based on these codes I group companies into three sectors: manufacturing, trade, and services.<sup>20</sup> The inclusion of these variables accounts for inter-industry wage differentials that are not already captured by the observed individual and company characteristics (see, for example, Krueger and Summers, 1987, Dickens and Katz, 1987, Gibbons and Katz, 1992 and Abowd, Kramarz and Margolis, 1999).

---

<sup>20</sup>I also ran regressions that included more detailed industry dummies. The results that I report in Section 4 are robust to this change in specification.

Table 3: IT USAGE BY COMPANY SIZE DISTRIBUTION

No. of employees	Freq.	Percent	IT usage (in percent)
1 to 4	1,707	7.8	42.9
5 to 9	2,716	12.3	45.3
10 to 49	6,186	28.1	50.1
50 to 99	2,790	12.7	57.1
100 to 499	4,457	20.2	63.8
500 to 999	1,447	6.6	68.3
1000 and more	2,721	12.4	72.1
Total	22,024		

## 4 Empirical Results

Table 4 displays the estimation results of the wage regressions. Unreported results show that the raw log wage differential for IT use in Germany is 0.275 (about 31 percent) in 1998/99. This figure is slightly lower than the raw log wage differential of 0.288 that DiNardo and Pischke (1997) report for Germany based on the 1991/92 cross-section of the BIBB/IAB data. Thus, in contrast to the period between 1979 and 1991/92, when the raw log wage differentials for computer use increased steadily (although at a declining pace), as shown in the paper by DiNardo and Pischke, this differential remained stable or even declined slightly in the 1990s.

Columns (1)-(5) show the results of specifications that are successively extended with additional controls. Including individual characteristics such as the level of formal education, work experience, tenure, gender, marital status or residence in a city reduces the coefficient of IT use by more than 30 percent (column 1). The coefficients of the controls have the expected sign, therefore I will not discuss them in detail. In column (2), workplace characteristics are included in the specification. The higher the measure is for non-routine cognitive activities, both analytical and interactive, the higher the wages are. By contrast, wages decrease in the measure for non-routine manual activities. Including the workplace characteristics additionally reduces the IT coefficient by 30 percent (compared to the coefficient in column 1). In column (3), company characteristics such as company size and information about the innovation strategy of the company are included in order to control, for example, for company size effects in remuneration. Industry dummies are also included to account for cross-sectoral differences in pay. In addition, dummies indicating company performance are included. Most interestingly, the inclusion of the company characteristics hardly affect the IT coefficient. By contrast, the returns to education decreased and the dummy for employees born in East Germany as well as the dummy for employees living in the city now are insignificant. Column (4) includes 10 dummies for West German states that control for cross-state differences in wage levels owing to, for example, differing economic conditions. However, these variables neither affect the IT coefficient nor the coefficients of the other controls. The specification in

column (5) includes 77 two-digit occupation dummies. The occupation dummies have a large impact on the estimated IT wage differential. From column (1) to column (5), the IT coefficient drops by more than 70 percent, indicating that the largest part of the raw logarithm wage differential for IT users has been due to observable differences that would have resulted in employees earning different wages even in the absence of IT. The results indicate that observable workplace characteristics such as workplace tasks and occupational affiliation account for the largest proportion of the bias in the raw logarithm wage differential. However, conditional on all the controls, the results still suggest that employees who use IT on the job earn around 8 percent higher wages.

Ideal control variables are only those that are attributes of the assignment to IT use and the earnings process but are unaffected by the treatment itself (for example, time-invariant individual characteristics such as gender and place of birth). Some of the controls used in this study, such as work experience or the incidence of previous periods of unemployment, might be affected by the treatment. IT users are, for example, less likely to become unemployed. Therefore, the treatment effect estimated here does not capture the indirect effects of IT use on wages (for example, through productivity).

Up to now, the coefficient of IT use has been constrained to be homogeneous on average conditional on the observable variables. I now extend the specification by adding a control function that includes interactions between all covariates  $X$  (previously demeaned using sample averages) and the IT use dummy  $D_i$ . Although the above specifications already include a large number of controls, the remaining 8 percent wage markup for IT users may still be due to characteristics that are not observable in the data set at hand if these unobservables are positively correlated with both IT use and wages. The aim of control functions is to purge the specification from the remaining covariance between unobservables and IT use (see Section 2.2.1). The results in column (6) show that including the control function in the specification reduces the estimated coefficient of IT use to 6 percent (=  $ATE$ ).<sup>21</sup>

The interaction terms are jointly significant ( $F_{(112,8666)} = 8.34, p = 0.0000$ ) and therefore provide evidence of the presence of heterogeneous effects. Table 5 shows the results for selected interaction terms. The empirical evidence suggests, for example, that IT users with a university degree benefit particularly in terms of wages. Differences in gender, work experience or workplace tasks, by contrast, do not have a significant effect on the gains from computer use.

The estimated  $ATT$  is 0.083, which is significantly different to the  $ATE$ . On average, treatment effect heterogeneity seems to be important. The result that the  $ATT$  is higher than the  $ATE$  suggests that there has been selection into treatment based on expected returns. If IT non-user had started to use IT instead of those who actually did, they would have enjoyed a substantial lower benefit than the group of actual IT users.

---

<sup>21</sup>The standard errors in column (6) are estimated using a design-matrix bootstrap approach in order to account for the generated regressors in the specification. Because of the large number of dummy variables in the specification the resamples are chosen to be twice as large as the sample in order to guarantee that all coefficients can be estimated. The estimated covariance matrix then is doubled in accordance with the rate of convergence of the estimator.

Table 4: OLS REGRESSIONS FOR THE EFFECT OF IT ON WAGES

Dependent Variable: Log(Hourly Wages)						
	(1)	(2)	(3)	(4)	(5)	(6)
IT	0.182*** (0.005)	0.123*** (0.006)	0.122*** (0.008)	0.122*** (0.008)	0.076*** (0.010)	0.061*** (0.024)
<b>Individ. Characteristics</b>						
high educ. level	0.437*** (0.011)	0.379*** (0.014)	0.330*** (0.018)	0.329*** (0.018)	0.222*** (0.020)	0.089 (0.068)
medium educ. level	0.126*** (0.009)	0.102*** (0.011)	0.083*** (0.014)	0.082*** (0.014)	0.052*** (0.014)	0.042** (0.024)
experience	0.017*** (0.001)	0.018*** (0.001)	0.018*** (0.001)	0.017*** (0.001)	0.017*** (0.001)	0.014*** (0.003)
experience <sup>2</sup> *(1/100)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000 (0.000)
tenure	0.009*** (0.000)	0.008*** (0.000)	0.006*** (0.000)	0.006*** (0.000)	0.005*** (0.000)	0.005*** (0.001)
woman	-0.092*** (0.009)	-0.083*** (0.010)	-0.098*** (0.013)	-0.098*** (0.014)	-0.088*** (0.014)	-0.110*** (0.037)
married	0.105*** (0.008)	0.097*** (0.008)	0.088*** (0.010)	0.088*** (0.010)	0.083*** (0.010)	0.082*** (0.020)
woman*married	-0.120*** (0.011)	-0.107*** (0.012)	-0.104*** (0.012)	-0.103*** (0.016)	-0.105*** (0.015)	-0.120*** (0.032)
ever unemployed	-0.022*** (0.006)	-0.024*** (0.007)	-0.037*** (0.008)	-0.037*** (0.008)	-0.036*** (0.008)	-0.008 (0.020)
born in East Germany	-0.045*** (0.013)	-0.031** (0.015)	-0.021 (0.018)	-0.019 (0.018)	-0.015 (0.014)	0.004 (0.037)
civil servant	-0.064*** (0.009)	-0.076*** (0.009)	-0.094** (0.041)	-0.094** (0.041)	-0.107*** (0.045)	0.003 (0.103)
lives in city	0.016*** (0.005)	0.014** (0.006)	0.008 (0.007)	0.005 (0.008)	0.002 (0.008)	-0.000 (0.020)

< Table continues on next page >

<Continued from previous page>

Dependent Variable: Log(Hourly Wages)							
	(1)	(2)	(3)	(4)	(5)	(6)	
<b>Company Characteristics</b>							
product innovation			0.004 (0.008)	0.004 (0.008)	0.008 (0.007)	0.016 (0.016)	
process innovation			0.036*** (0.008)	0.034*** (0.008)	0.025*** (0.008)	0.018 (0.016)	
very good company performance			0.044** (0.026)	0.038* (0.021)	0.045** (0.022)	-0.017 (0.046)	
good company performance			0.007 (0.025)	0.003 (0.020)	0.010 (0.020)	-0.025 (0.046)	
rather bad company performance			-0.012 (0.021)	-0.016 (0.021)	-0.011 (0.021)	-0.027 (0.046)	
<b>Workplace Characteristics: Measure of...</b>							
non-routine analytic tasks			0.087*** (0.012)	0.120*** (0.016)	0.120*** (0.016)	0.077*** (0.016)	0.048 (0.060)
non-routine interactive tasks			0.154*** (0.011)	0.154*** (0.014)	0.155*** (0.014)	0.150*** (0.015)	0.179 (0.036)
routine cognitive tasks			0.031*** (0.012)	0.005 (0.014)	0.004 (0.014)	-0.001 (0.014)	-0.002 (0.028)
routine manual tasks			-0.006 (0.017)	-0.040** (0.018)	-0.037** (0.018)	-0.033* (0.020)	-0.043 (0.041)
non-routine manual tasks			-0.119*** (0.011)	-0.096*** (0.014)	-0.095*** (0.014)	-0.083*** (0.014)	-0.041 (0.036)
77 occupation dummies	No	No	No	No	Yes	Yes	
7 company size dummies	No	No	Yes	Yes	Yes	Yes	
dummies for manufacturing & trade	No	No	Yes	Yes	Yes	Yes	
10 dummies for West German states	No	No	No	Yes	Yes	Yes	
$IT * [x - E(x)]$	No	No	No	No	No	Yes	
R <sup>2</sup>	0.348	0.340	0.399	0.401	0.443	0.455	
Number of observations	18547	15266	8936	8936	8897	8897	

Employees with low levels of education working in large companies in the services sector are the base category. Heteroscedasticity-consistent standard errors are in parentheses in column (1)-(5). Column (6):  $IT * [x - E(x)]$  means that the specification includes interaction terms between IT use and all the level variables  $X$ , where all the  $X$  had previously been demeaned by the sample averages. Bootstrapped standard errors using 100 resamples. \*\*\*, \*\*, \*-indicate significance at the 1, 5, 10 percent level.

Table 5: SELECTED INTERACTION TERMS

Dependent Variable: Log(Hourly Wages)	
high educ. level * IT	0.160*** (0.077)
medium educ. level * IT	0.023 (0.035)
woman * IT	0.040 (0.040)
non-routine analytic tasks * IT	0.035 (0.062)
non-routine interactive tasks * IT	-0.030 (0.048)
routine cognitive tasks * IT	0.007 (0.044)
routine manual tasks * IT	0.021 (0.064)
non-routine manual tasks * IT	-0.059 (0.046)
experience * IT	0.005 (0.004)
experience <sup>2</sup> * IT	-0.000 (0.000)
civil servant * IT	-0.154 (0.108)
tenure * IT	-0.000 (0.002)
born in East Germany * IT	-0.026 (0.052)
married * IT	0.007 (0.026)
very good company performance * IT	0.091 (0.084)
good company performance * IT	0.053 (0.088)
rather bad company performance * IT	0.021 (0.080)

One of the most powerful criticisms of the “returns on computer use” literature comes from DiNardo and Pischke (1997) who show that there is also a considerable wage effect of the use of pencils (and other “white-collar” tools) in their cross-section estimates. They convincingly argue that if we don’t believe that pencils changed the wage structure, why should we believe that computers did. Hence, they attribute the estimated wage differentials for IT use to unobserved skills.

One part of their analysis deserves special attention. Table 6 reports parts of their Table III. Panel A reports the coefficients for computer use when separate regressions are performed for each workplace tool. Panel B shows the results when all tools are included together in the specification. Controlling for the different workplace tools attenuates the coefficient for computer use in each period, but the effect declines over time. Including the tools reduces the estimated coefficient on computer use by around 40 percent in 1979, by around 33 percent in 1985/86 and by 26 percent in 1991/92. In 1991/92, the coefficient of computer use has a magnitude of 13 percent even after controlling for all the tools. I reproduced their estimates for the most recent wave of the data set (last column). Relative to the result in Panel A, the estimated coefficient for computer use decreases by around 30 percent due to the inclusion of the dummies for the different workplace tools (Panel B). The coefficient of computer use still has a magnitude of around 15 percent. In contrast to the previous years, in which the different tools have always had a significant positive impact on wages, only the coefficients for the use of calculators and working while sitting remained significantly positive in 1998/99. The coefficient for using a telephone at work is now even significantly negative.

These results suggest that the changes in computer technology over time has also altered its role in the workplace. Computer technology has decreased the costs of accessing and processing information in the 1980s and early 1990s, whereas in the late 1990s it additionally reduced the costs of communication. The convergence of information technology and communication technology, once two distinct technological areas, has been the central feature of technological change in the 1990s.

I take up the DiNardo and Pischke (1997) idea and estimate regressions that include dummies indicating the use of various workplace tools instead of the computer dummy. The results are shown exemplarily for pencils in Table 7. Unreported results show that the first-order relationship between pencil use and wages is 5.7 percent (significant at the 1-percent level). Similar to the specifications in Table 4, I successively augment the specifications with additional controls. In contrast to the IT effect, the estimated pencil effect disappears as soon as controls for individual and workplace characteristics are included in the specification (column 2). The variables that have had only attenuating effects on the coefficient of IT use, now result in the pencil effect to disappear. Column (2) shows the results of the regression specification including the minimum number of controls that are necessary to eliminate the wage effect of pencil use. It is interesting to notice that the coefficient in the comparable IT equation (Table 4, column 2) still is about 12 percent. This result also suggests that changes in computer technology over time altered its function in the workplace.

Table 6: OLS REGRESSIONS FOR THE EFFECT OF DIFFERENT TOOLS ON PAY

Dependent Variable: Log(Hourly Wages)				
	Germany 1979	Germany 1985–86	Germany 1991–92	Germany 1998–99
A. Tools entered separately				
Computer	0.112 (0.010)	0.157 (0.007)	0.171 (0.006)	0.204 (0.006)
B. Tools entered together				
Computer	0.066 (0.010)	0.105 (0.008)	0.126 (0.007)	0.146 (0.007)
Calculator	0.017 (0.008)	0.053 (0.007)	0.044 (0.007)	0.051 (0.006)
Telephone	0.072 (0.007)	0.043 (0.008)	0.045 (0.008)	-0.019 (0.006)
Pen/Pencil	0.062 (0.007)	0.031 (0.008)	0.035 (0.008)	0.003 (0.011)
Work while sitting	0.058 (0.007)	0.050 (0.007)	– –	0.065 (0.006)

Standard errors are in parentheses. Source of columns 1-3: DiNardo and Pischke (1997, p. 299, Table III, 2nd part). Column 4: own regressions. Similar to the specification in DiNardo and Pischke education, experience, experience squared, dummies for part-time, city, female, married, female\*married, and for civil servants are included in the regressions.

Table 7: OLS REGRESSIONS FOR THE EFFECT OF PENCIL USE ON WAGES

Dependent Variable: Log(Hourly Wages)					
	(1)	(2)	(3)	(4)	(5)
Pencil	0.033*** (0.010)	0.016 (0.011)	0.016 (0.013)	0.016 (0.013)	0.005 (0.013)
<b>Individ. Characteristics</b>					
high educ. level	0.495*** (0.013)	0.398*** (0.015)	0.358*** (0.020)	0.357*** (0.020)	0.231*** (0.021)
medium educ. level	0.151*** (0.011)	0.112*** (0.013)	0.096*** (0.016)	0.095*** (0.016)	0.057*** (0.016)
experience	0.018*** (0.001)	0.019*** (0.001)	0.017*** (0.001)	0.017*** (0.001)	0.018*** (0.001)
experience <sup>2</sup> *(1/100)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)
tenure	0.009*** (0.000)	0.008*** (0.000)	0.005*** (0.001)	0.006*** (0.001)	0.005*** (0.001)
woman	-0.079*** (0.010)	-0.075*** (0.011)	-0.088*** (0.014)	-0.089*** (0.014)	-0.095*** (0.015)
married	0.120*** (0.009)	0.103*** (0.009)	0.091*** (0.011)	0.091*** (0.011)	0.086*** (0.011)
woman*married	-0.119*** (0.012)	-0.101*** (0.013)	-0.095*** (0.017)	-0.093*** (0.017)	-0.098*** (0.016)
ever unemployed	-0.036*** (0.007)	-0.033*** (0.007)	-0.046*** (0.009)	-0.046*** (0.009)	-0.044*** (0.008)
born in East Germany	-0.056*** (0.015)	-0.035*** (0.015)	-0.024 (0.020)	-0.023 (0.020)	-0.010 (0.020)
civil servant	-0.065*** (0.009)	-0.081*** (0.009)	-0.105*** (0.040)	-0.105*** (0.040)	-0.107*** (0.045)
lives in city	0.023*** (0.006)	0.017*** (0.006)	0.011 (0.008)	0.007 (0.009)	0.002 (0.008)

< Table continues on next page >

<Continued from previous page>

Dependent Variable: Log(Hourly Wages)					
	(1)	(2)	(3)	(4)	(5)
<b>Company Characteristics</b>					
product innovation			0.002 (0.008)	0.002 (0.008)	0.005 (0.008)
process innovation			0.057*** (0.008)	0.056*** (0.009)	0.036*** (0.008)
very good company performance			0.053** (0.023)	0.048** (0.022)	0.055** (0.023)
good company performance			0.008 (0.021)	0.004 (0.021)	0.013 (0.021)
rather bad company performance			-0.012 (0.023)	-0.016 (0.023)	-0.009 (0.022)
<b>Workplace Characteristics: Measure of...</b>					
analytic tasks		0.110*** (0.012)	0.139*** (0.017)	0.139*** (0.017)	0.082*** (0.017)
interactive tasks		0.185*** (0.011)	0.185*** (0.014)	0.186*** (0.014)	0.158*** (0.015)
routine cognitive tasks		0.047*** (0.014)	0.023 (0.016)	0.022 (0.016)	0.008 (0.016)
routine manual tasks		-0.032* (0.019)	-0.067*** (0.020)	-0.064*** (0.020)	-0.038* (0.022)
non-routine manual tasks		-0.171*** (0.012)	-0.135*** (0.015)	-0.134*** (0.015)	-0.095*** (0.015)
77 occupation dummies	No	No	No	No	Yes
7 company size dummies	No	No	Yes	Yes	Yes
dummies for manufacturing & trade	No	No	Yes	Yes	Yes
10 dummies for West German states	No	No	No	Yes	Yes
R <sup>2</sup>	0.300	0.323	0.383	0.385	0.437
Number of observations	15951	13986	8037	8037	8001

Employees with low levels of education working in large companies in the services sector are the base category. Heteroscedasticity-consistent standard errors are in parentheses. \*\*\*, \*\*, \*-indicate significance at the 1, 5, 10 percent level.

Table 8: RESULTS OF THE PROPENSITY SCORE MATCHING

Dependent Variable: Log(Hourly Wages)				
	(1)	(2)	(3)	(4)
IT	0.078*** (0.032)	0.077*** (0.021)	0.062*** (0.025)	0.082*** (0.022)
Number of Treated	5,379	5,379	5,076	5,076
Number of Controls	3,461	3,461	3,461	3,461

(1) shows the ATT of IT use using nearest neighbor matching (random draw version, bootstrapped standard errors using 100 replications). (2) shows the ATT using kernel matching (Epanechnikov kernel, bootstrapped standard errors using 100 replications). (3) shows the ATT of IT use using nearest neighbor matching (random draw version, bootstrapped standard errors using 100 replications) with the additional restriction that treated and controls have the same level of education. (4) shows the ATT using kernel matching (Epanechnikov kernel, bootstrapped standard errors using 100 replications) with the additional restriction that treated and controls have the same level of education. Only observations that are on the common support are used. The caliper is set to 0.001. The propensity score is estimated using the level of formal education, age, age<sup>2</sup>, work experience, work experience<sup>2</sup>, interaction between work experience and education, workplace tasks, born in East Germany, ever unemployed, living in a city, woman, married, married woman, 8 company size dummies, 39 industry dummies, 3 dummies reflecting company performance, 79 occupation dummies, 10 dummies for West German states and a constant as regressors.

Table 8 displays the results of the matching estimations. Column (1) shows that, using nearest neighbor matching, the average wage effect of IT use for employees who use IT is 7.8 percent. The coefficient is significant on the 1-percent level. In the matching, 3,461 controls are used to estimate the potential missing outcome of IT users had they not adopted computer technology.<sup>22</sup> Owing to the restriction that only observations within the region of common support are used, there are 400 IT users who are not considered in the analysis. Column (2) shows the results when an Epanechnikov kernel is used in the matching process. This barely changes the magnitude of the ATT, but increases the precision of the estimate.

Table 9, last column, shows the means of the main individual characteristics of the IT non-users that are used as controls in the matching approach. It is evident that, although

<sup>22</sup>Owing to missing values in the matching variables, the sample reduces to 9,240 individuals (5,779 IT user and 3,461 IT non-user) in the matching specification.

Table 9: MEAN COMPARISON FOR IT USERS AND IT NON-USERS

	IT user	IT non-user prior to matching	selected controls <sup>§</sup>
high education level	0.25	0.07*	0.17*
medium education level	0.69	0.72*	0.74*
low education level	0.06	0.21*	0.09*
age	39.99	40.29*	40.20
experience	20.04	21.52*	21.22*
tenure	12.37	10.82*	12.33
ever unemployed	0.26	0.34*	0.29
married	0.70	0.65*	0.70
woman	0.44	0.44	0.44
civil servants	0.15	0.05*	0.13
born in East Germany	0.03	0.05*	0.04

\* indicates that the means differ with statistical significance of 5 percent in a two-tailed t-test between IT user and either IT non-user prior to matching (column 3) or the selected IT non-user (last column).

<sup>§</sup> IT non-users who are selected by the matching procedure.

there is a convergence between treated and controls with respect to most of the observable characteristics, the difference in educational level is still significant. Therefore, columns (3) and (4) of Table 8 show estimates that, in addition to having close propensity scores, restrict matches to be within groups of employees with equal levels of education. This extension increases the number of IT users that are off the common support to 703. The estimate of the nearest neighbor matching declines to 6.2 percent, whereas the coefficient of the estimate that uses an Epanechnikov kernel increases to 8.2 percent. Both estimates are highly significant.

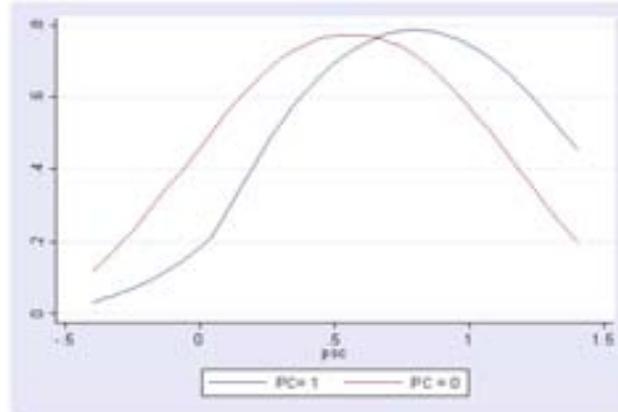
As for the *ATT*, the estimate using RBM is very close to the one estimated by the propensity score matching procedure. This may be the case because (i) there is no common support problem, (ii) there is little heterogeneity in treatment effects or all the propensity scores are small, and (iii) there is no serious mis-specification in the no-treatment outcome (see Blundell, Dearden and Sianesi, 2003).

Figure 1 shows the kernel density estimates of the distribution of propensity scores for IT users and IT non-users. The two distributions greatly overlap. However, the results of the matching procedure reveal that 400 (703) of the 5,779 IT users are outside the region of common support. In this application imputing the values that are outside the common support by relying on a functional form assumption or discarding the IT users without similar counterfactuals from the analysis does not lead to different results.

The difference in the *ATE* and *ATT* in the RBM approach suggests that heterogenous treatment effects are important. This notion is also supported by the joint significance of the interaction terms in the RBM specification, although selected interaction terms only point to weak effects.

The propensity scores in the analysis assume values between 0 and 1. The mean is 0.64 and the median is 0.73. This information does not suggest that propensity scores are particularly clustered at certain points in the distribution.

Figure 1: KERNEL DENSITY ESTIMATIONS OF PROPENSITY SCORES FOR IT USERS AND IT NON-USERS



## 5 Conclusions

This paper deals with the question of whether IT users earn higher wages than employees who do not use IT on the job. In addition to the average treatment effect of IT use, which was the focus of previous studies, I estimate the average treatment effect for the treated. The *ATT* is particularly interesting for the question of IT wage differentials because the implementation of IT cannot be analyzed without considering the occupational context. Employers choose employees to use IT on the basis of expected profitability increases owing to the implementation of IT, which depends on occupational characteristics, individual characteristics and company characteristics. The prevailing circumstances greatly influence the impact of IT on productivity and thus individual wages.

The analysis rests upon the advantage of the data set that includes a large number of controls that have been identified as important determinants of the IT use-wage relationship in previous studies. This feature of the data set allows me to use methods that assume that, conditional on the controls, the differences in wages of IT users and IT non-users are attributable to the use of IT.

I find a robust average treatment effect for the treated of around 8 percent. This indicates that IT users would be worse off had they not started to use IT in the workplace. I also find evidence for the presence of heterogeneous treatment effects. A comparison of the *ATE* and *ATT* reveals that IT non-users would have witnessed lower benefits in terms of wages had they started to use IT instead of those who actually have. However, the difference in *ATT* and *ATE* is only about 2 percentage points.

An 8 percent wage markup appears to be small, in particular in comparison with the 10-15 percent that have typically been found in cross-section analyses in the 1990s. The size of the effect is rather similar to results previously found in panel analyses. The data set does not include unemployed. I therefore argue that the 8 percent wage markup for IT users represents a lower bound. Analyses of Entorf et al. (1999), for example, suggest that people who do not use computers have a higher probability of becoming unemployed. Therefore, the IT non-users observed in my data set are probably already a positive selection.

## 6 Appendix

Intermediate steps between equation 12 and 13:

$$\begin{aligned}
 E(Y_i|D_i, X) &= \mu_0 + \alpha D_i + g_0(X) + D_i[g_1(X) - g_0(X)] \\
 g_0(X) &= \eta_0 + X\beta_0 \quad E g_0(X) = 0 \\
 g_1(X) &= \eta_1 + X\beta_1 \quad E g_1(X) = 0 \\
 E(Y_i|D_i, X) &= \mu_0 + \alpha D_i + \eta_0 + X\beta_0 + D_i[\eta_1 + X\beta_1 - \eta_0 - X\beta_0] \\
 &= \mu_0 + \eta_0 + \alpha D_i + X\beta_0 + D_i[\eta_1 - \eta_0 + X(\beta_1 - \beta_0)] \\
 &= \underbrace{\mu_0 + \eta_0}_{\gamma_0} + \alpha D_i + X\beta_0 + D_i\left[\frac{\eta_1 - \eta_0}{\beta_1 - \beta_0} + X\right] \underbrace{(\beta_1 - \beta_0)}_{\delta}
 \end{aligned}$$

To be shown:  $\frac{\eta_1 - \eta_0}{\beta_1 - \beta_0} = -E(X)$

$$\begin{aligned}
 \frac{\eta_1 - \eta_0}{\beta_1 - \beta_0} &= \frac{1}{\beta_1 - \beta_0}(g_1(X) - X\beta_1 - g_0(X) + X\beta_0) \\
 \eta_1 - \eta_0 &= g_1(X) - g_0(X) + X(\beta_0 - \beta_1) \\
 X &= \frac{1}{\beta_1 - \beta_0}(g_1(X) - g_0(X) - \eta_1 + \eta_0) \\
 E(X) &= \frac{1}{\beta_1 - \beta_0}(\underbrace{E g_1(X)}_{=0} - \underbrace{E g_0(X)}_{=0} - \eta_1 + \eta_0) \\
 &= -\frac{\eta_1 - \eta_0}{\beta_1 - \beta_0}
 \end{aligned}$$

## References

- Abowd, J., Kramarz, F. and Margolis, D. (1999). High Wage Workers and High Wage Firms, *Econometrica* **67**(2): 251–333.
- Acemoglu, D. (2002). Technical Change, Inequality, and the Labor Market, *Journal of Economic Literature* **40**: 7–72.
- Angrist, J. (1998). Estimating the Labour Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants, *Econometrica* **66**(2): 249–288.
- Angrist, J., Imbens, G. and Rubin, D. (1996). Identification of Causal Effects Using Instrumental Variables, *Journal of the American Statistical Association* **91**(434): 444–455.
- Autor, D., Levy, F. and Murnane, R. (2003). The Skill Content of Recent Technological Change: An Empirical Exploration, *Quarterly Journal of Economics* **118**(4): 1279–1333.
- Bell, B. (1996). Skill-Biased Technical Change and Wages: Evidence from a Longitudinal Data Set, *Technical report*, Nuffield College.
- Blundell, R. and Costa Dias, M. (2000). Evaluation Methods for Non-Experimental Data, *Fiscal Studies* **21**(4): 427–468.
- Blundell, R., Dearden, L. and Sianesi, B. (2003). Evaluating the Impact of Education on Earnings in the UK: Models, Methods and Results from the NCDS, *Working Paper 03–20*, IFS.
- Borghans, L. and ter Weel, B. (2004). The Diffusion of Computers and the Distribution of Wages, *Working paper*, ROA Maastricht University.
- Brown, C. and Medoff, J. (1989). The Employer-Size Wage Effect, *Journal of Political Economy* **97**(5): 1027–1059.
- Card, D. and DiNardo, J. (2002). Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles, *Journal of Labor Economics* **20**(4): 733–783.
- Chennells, L. and van Reenen, J. (2002). Technical change and the structure of employment and wages: A survey of the microeconomic evidence, in N. Greenan, Y. L’Horty and J. Mairesse (eds), *Productivity, Inequality and the Digital Economy*, MIT Press, Cambridge, MA, pp. 175–223.
- Dehejia, R. and Wahba, S. (1999). Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs, *Journal of the American Statistical Association* **94**: 1053–1062.
- Dehejia, R. and Wahba, S. (2002). Propensity Score Matching for Nonexperimental Causal Studies, *Review of Economics and Statistics* **84**: 151–161.
- Dickens, W. and Katz, L. (1987). Inter-Industry Wage Differences and Industry Characteristics, in K. Lang and J. Leonard (eds), *Unemployment and the Structure of the Labor Market*, Basil Blackwell, New York, pp. 48–89.
- DiNardo, J. and Pischke, J. (1997). The Returns to Computer Use Revisited: Have Pencils Changed the Wage Structure Too?, *Quarterly Journal of Economics* **112**: 291–303.

- Dolton, P. and Makepeace, G. (2004). Computer Use and Earnings in Britain, *Economic Journal* **114**: C117–C129.
- Entorf, H. and Kramarz, F. (1997). Does Unmeasured Ability Explain the Higher Wages of New Technology Workers?, *European Economic Review* **41**: 1489–1509.
- Entorf, H. and Kramarz, F. (1998). The Impact of New Technologies on Wages: Lessons from Matching Panels on Employees and on Their Firms, *Economics of Innovation and New Technology* **5**: 165–197.
- Entorf, H., Gollac, M. and Kramarz, F. (1999). New Technologies, Wages, and Worker Selection, *Journal of Labor Economics* **17**(3): 464–491.
- Gibbons, R. and Katz, L. (1992). Does Unmeasured Ability Explain Inter-Industry Differentials, *Review of Economic Studies* **59**(3): 515–535.
- Heckman, J. (1979). Sample Selection Bias as a Specification Error, *Econometrica* **47**: 153–161.
- Heckman, J. and Robb, R. (1985). Alternative Methods for Evaluating the Impact of Interventions, in J. Heckman and B. Singer (eds), *Longitudinal Analysis of Labor Market Data*, Cambridge University Press, Cambridge, pp. 156–245.
- Heckman, J. and Robb, R. (1986). Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes, in H. Wainer (ed.), *Drawing Inferences from Self-Selected Samples*, Springer, Berlin, pp. 63–107.
- Heckman, J., Ichimura, H. and Todd, P. (1997). Matching as a Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme, *Review of Economic Studies* **64**(4): 605–654.
- Heckman, J., Ichimura, H. and Todd, P. (1998a). Matching as a Econometric Evaluation Estimator, *Review of Economic Studies* **65**(2): 261–294.
- Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1998b). Characterizing Selection Bias Using Experimental Data, *Econometrica* **66**(5): 1017–1098.
- Imbens, G. and Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effects, *Econometrica* **62**(2): 467–475.
- Katz, L. and Autor, D. (1999). Changes in the Wage Structure and Earnings Inequality, in O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Elsevier Science, Amsterdam, pp. 1463–1555.
- Krueger, A. (1993). How Computer have Changed the Wage Structure: Evidence from Microdata, 1984-1989, *Quarterly Journal of Economics* **108**(1): 33–60.
- Krueger, A. and Summers, L. (1987). Reflections on the Inter-Industry Wage Structure, in K. Lang and J. Leonard (eds), *Unemployment and the Structure of the Labor Market*, Basil Blackwell, New York, pp. 17–47.
- LaLonde, R. (1986). Evaluating the Econometric Evaluations of Training Programs with Experimental Data, *American Economic Review* **76**(4): 604–620.
- Lechner, M. (2000). A Note on the Common Support Problem in Applied Evaluation Studies, *mimeo*, University of St.Gallen.

- Machin, S. (2001). The Changing Nature of Labour Demand in the New Economy and Skill-Biased Technological Change, *Oxford Bulletin of Economics and Statistics* **63**(Special Issue): 753–776.
- Rosenbaum, P. and Rubin, D. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika* **70**(1): 41–55.
- Rubin, D. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, *Journal of Educational Psychology* **66**(5): 688–701.
- Schmidt, C. and Zimmermann, K. (1991). Work Characteristics, Firm Size and Wages, *Review of Economic Studies* **73**(4): 705–710.
- Smith, J. and Todd, P. (2004). Does Matching Overcome Lalonde’s Critique of Nonexperimental Estimators?, *Journal of Econometrics*, *forthcoming*.
- Spitz, A. (2004). Are Skill Requirements in the Workplace Rising? Stylized Facts and Evidence on Skill-Biased Technological Change, *Discussion Paper 04-33*, ZEW Mannheim.
- Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, Massachusetts.