RUHR
ECONOMIC PAPERS

Janina Reinkowski
Björn Alecke
Timo Mitze
Gerhard Untiedt

# Effectiveness of Public R&D Subsidies in East Germany

## Is it a Matter of Firm Size?

RUHR UNIVERSITÄT BOCHUM RUB rwi

#204

# Imprint

## Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

## Editors



## Editorial Office



## Ruhr Economic Papers #204

Responsible Editor: Christoph M. Schmidt



ISSN 1864-4872 (online) – ISBN 978-3-86788-233-0

Ruhr Economic Papers #204

Janina Reinkowski, Björn Alecke, Timo Mitze, and Gerhard Untiedt

# Effectiveness of Public R&D Subsidies in East Germany

## Is it a Matter of Firm Size?

## Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

ISSN 1864-4872 (online)
ISBN 978-3-86788-233-0

Janina Reinkowski, Björn Alecke, Timo Mitze, and Gerhard Untiedt[1]

# Effectiveness of Public R&D Subsidies in East Germany – Is it a Matter of Firm Size?

## Abstract

*This paper analyses the impact of public subsidies on private sector research and development (R&D) activity for East German firms. Using propensity score matching, our empirical results indicate that subsidized firms indeed show a higher level of R&D intensity and a higher probability for patent application compared to non-subsidized firms for our sample year 2003. On average we find an increase in the R&D intensity of about 3.7 percentage points relative to non-subsidized firms. The probability for patent applications rises by 21 percentage points. These results closely match earlier empirical results for East Germany. Given the fact that the East German innovation system is particularly driven by small and medium sized enterprises (SME), we put a special focus on the effectiveness of the R&D subsidies for this latter subgroup. Here no previous empirical evidence is available so far. Our findings indicate that policy effectiveness also holds for private R&D activity of SMEs, where the highest increase in terms of R&D intensity is estimated for micro businesses with up to 10 employees.*

*JEL Classification: C14, C21, O32, O38*

*Keywords: Propensity score matching; R&D subsidies; East Germany; SME*

*September 2010*

---

*1 Janina Reinkowski, RWI and ifo Munich; Björn Alecke, Gesellschaft für Finanz- und Regionalanalysen (GEFRA) Münster; Timo Mitze, RWI and Ruhr-Universität Bochum; Gerhard Untiedt, GEFRA Münster & Technical-University Clausthal.* *– Earlier versions of this paper have been presented at the DRUID Winter conference 2010, Aalborg, the XII. Spring Meeting of Young Economists 2010, Luxembourg, the NAKE Research Day 2010, Utrecht, the RWI Therapy Winter Seminar 2010 and the 50. ERSA Congress 2010, Jönköping. The authors wish to thank Michela Bia, Judith Niehues, Joel Stiebale, Paul Windrum and further participants of the above events for helpful comments and advices. Special thanks also to Oliver Falck, Dirk Engel and Christoph M. Schmidt, whose suggestions helped to signicantly improve the structure of the paper.* *– All correspondence to Janina Reinkowski, RWI, Hohenzollernstr. 1-3, 45128 Essen, Germany, E-Mail: janina.reinkowski@rwi-essen.de.*

# 1 Introduction

Two decades after German re-unification the East German economy still faces several challenges on its convergence path towards the economic structures of its Western counterpart. Although the regional economy and especially its manufacturing sector have made considerable progress throughout the last two decades, the economy still shows some distinct structural weaknesses such as a by far lower innovation activity compared to the Western average (see e.g. Guenther et al., 2010, and Alecke et al., 2010). That is, expenditures on R&D by the private sector were about half as high in East Germany compared to West Germany in 2006 (BERD of 1% compared to 1.9% as a share of regional GDP) and similar also for patent applications per capita (70% of the Western average). The importance of the region's absorptive capacity in building up stocks of knowledge and R&D as key determinants of the region's technology level and subsequently long-run economic development has recently been stressed by the 'new' growth theory (see e.g. Barro & Sala-i-Martin, 2004).

Consequently, there have been several political attempts for East Germany to foster income growth and convergence, which also increasingly accounted for the role played by knowledge creation through R&D. As a result, about 60% of all innovating firms in East Germany received public fundings troughout the mid-1990s compared to only 10% o in West Germany (see Ebling et al., 1999). This dependency has not been reduced much throughout the last years (see e.g. Rammer & Czarnitzki, 2003, as well as Czarnitzki & Licht, 2006). It is thus a matter of primary interest from the policy perspective, to look more carefully at the effectiveness of these spending schemes given their rather large absolute volumes (roughly 1.1 billion spent per year).

What makes East Germany a particularly interesting case study, is its regional innovation system, which is dominated by small and medium-sized firms. Compared to West Germany, the share of small and medium-sized enterprises (SME) engaged in continuous R&D activity is far higher (36% of all SME relative to 9%). The same holds for the share of R&D employment in total employment for firms up to 50 employees (see Guenther et al., 2010). As earlier work has shown, the regional knowledge production function transforming knowledge and R&D inputs into innovative outputs may substantially differ from innovation systems driven by large innovative firms (see e.g. Conte & Vivarelli, 2005). This naturally points to the question in how far the publically financed R&D factor inputs can be absorbed by SMEs so that public policies effectively foster private R&D activity rather than crowding-out private spendings.

One the hand one may argue that SMEs are highly affected by barriers to innovation such as financial bottlenecks, shortage and hindered access to qualified personnal as well as limited internal and missing market know-how, which negatively affects their innovative performance. This may particularly be true for SMEs in pheripherical regions such as East Germany (see e.g. McAdam et al., 2006). On the other hand, SMEs are likely to be more flexible and may easily adjust their knowledge production function to changes in changes of the surrounding innovation system. With respect to R&D grants this could imply that SMEs are not able to increase their innovation process over a certain maximum, so that with an increase in the level of subsidy also increased crowing-out effects can be observed. However, at the same time R&D grants may also significantly contribute to overcome barriers to innovation for SME in East Germany. Thus, from a theoretical point of view there are both arguments advoacting as well as challenging the effectiveness of R&D subsidies. In the following, we put special emphasis on answering the latter point using a large firm level dataset for East German regions and applying propensity score based matching estimation.

Compared to the huge international evidence on R&D policy effects, there are rather few empirical references for the East German case.[1] Almus and Czarnitzki (2003) as well as Czarnitzki and Licht (2006) report that public R&D support has a significant positive effect on private sector R&D intensity. Both studies report an increase in the innovation activity of East German firms receiving public funding of about 4 percentage points relative to the non-subsidized comparison group. Additionally, Czarnitzki and Licht (2006) analyse the marginal effects of R&D subsidies on patent activity, which is estimated to be about 22 %. For both the R&D intensity as well as patent activity, the latter authors also identify a higher effect in East Germany compared to the Western states throughout the period 1994 to 2000. However, no empirical evidence is given for the important subgroup of SMEs in the regional innovation system of East Germany. This papers aims to fill this gap in the empirical literature.

The remainder of the paper is organized as follows: In the next section we briefly introduce the micro dataset used for the empirical exercise. Section 3 then sketches the Kernel matching approach and discusses the empirical results for the full sample as well as three subsamples for small firms. Section 4 reports the results for robustness checks based on a variety of alternative matching algorithms as well as a sample restriction to only those firms, who permanently perform R&D. Section 5 concludes.

[1] Literature surveys summarizing the effects found at the international levels are for instance given by David et al. (2000), Hall & van Reenen (2000), as well as Garcia-Quevedo (2004). On average public subsidies are found to exert a positive effect on different private sector outcome variables such as R&D expenditures and R&D employment (as factor inputs) as well as innovation activity and patent application (as output indicator in the process of knowledge creation).

## 2 Data and Variable Definition

We assess the impact of direct R&D support measures using micro data based on the GEFRA Business survey (GEFRA et al., 2004 and 2005). This survey was conducted in the evaluation process of two direct enterprise support schemes, namely the "Joint Task for the Promotion of Industry and Trade" and the "Promotion of Joint Research Projects" on behalf of the Thuringian Ministry of Economics and the Thuringian Ministry of Science, Research and Arts. For the survey a total of 6.861 firms in the manufacturing and production-oriented service sector have been contacted. The return rate was about 21%, so the survey contains a total of 1.484 firms of which 284 firms received public R&D grants. The questionnaire refers to firm-specific data for the year 2003. Since earlier evidence on the effectiveness of R&D policies in East Germany was exclusively based on data from the *Mannheimer Innovation Panel* (MIP), our results may serve as the first cross-check for the robustness of these results.[2]

The questionnaire was designed to incorporate data on various factor inputs (labour, intermediate inputs, human and physical capital) as well as economic outcome variables such as sales, exports and labour productivity. With respect to variables representing R&D activity at the firm level the dataset includes a binary dummy variable for general patenting activity and the firm's R&D intensity as the ratio of R&D expenditure to sales. Firms were asked whether they received funding by any R&D support programmes of the federal government, the federal states or the European Union. Since all possible R&D programmes launched by public authorities are covered by the survey, this study is not restricted to a particular policy measure, but reflects the joined effect of the available set of public R&D policies. Many studies deal with only one specific public R&D program and cannot control for possible effects of other sources of public R&D funding (for a discussion see e.g. Almus and Czarnitzki, 2003). In contrast, our approach is able to construct a treatment group consisting of those firms that received subsidies at the regional, national or EU level.

A large set of firm specific control variables is necessary to appropriately isolate the causal effect of R&D subsidies. Basic variables are firm size in terms of total employment and firm age. We further use capital-intensity, defined as tangible assets per employee, to control for the technology used in the production process and also test for the effect of the investment intensity, defined as total investments per sales. The skill structure of a firm's workforce is used as a determinant of research activity and the ability to attract

[2]For details on the MIP see e.g. Janz et al. (2001).

public funding in a significant way (see Kaiser, 2004). We thus include the share of highly educated employees, i.e. those who hold a university degree (including universities of applied sciences). To account for the role of firm competitiveness we include the export ratio in our model.

Additionally, we use binary dummy variables to indicate the firm's legal form and the affiliation to a parent company either in West, East Germany or abroad respectively. We also take into account whether a firm is performing research on a regular basis and/or is running an own R&D department.[3] This latter variable should reflect the absorptive capacity and R&D experience for a specific firm. A detailed list of variables covered in this analysis together with their descriptive statistics is given in Table 1 and Table 2. Before applying the data for empirical analysis, we tested for sample representativeness, see Appendix A.

## 3 Matching Estimation

In order to identify the causal impact of public subsidies on private sector R&D activity we apply nonparametric matching, which is by now a common approach for microeconometric programme evaluation (see e.g. Heckman et al., 1997). The approach is nonparametric in nature and compares the sample average of outcome variables for firms that exhibit a treatment with those firms that are similar in terms of a predefined set of variables, but are not subject to the treatment. Our estimation strategy rests on a two-stage process, which estimates a probit model for the receipt of programme participation in a first step, and then uses the obtained propensity score in order to control for further determinants of private R&D activity beside the public subsidies, when comparing the means of treated and comparison firms in a second step.

It is worth noting, that the matching approach rests on two crucial assumptions. First, we assume that for individuals (firms) with identical characteristics our outcome measure (R&D intensity, patent applications) is independent from the treatment. This conditional independence assumption (CIA) is expected to be fulfilled if all variables influencing the outcome are at our disposal. Although the CIA cannot be tested, our sample contains a rich set of firm-level information which is likely to make the CIA being fulfilled. A second

[3]Concerning the question of the frequency of R&D being conducted, the GEFRA Business Survey asks whether a firm is either 'regularly', 'sporadically' or 'not at all' engaged in own R&D activities. When using this variable for propensity score estimation there is of course the potential risk of endogeneity, since the decision to apply for R&D subsidies may depend on the frequency of R&D activities. However, here we assume that the impact of a subsidy is not that big that a firm would actually change its structural behavior. This is even more reasonable for the case of an R&D department. We thus take the latter variable as default proxy for regular R&D activity at the firm level.

**Table 1: Variable definition**

| Variable | | Description |
|---|---|---|
| **R&D activity** | | |
| *Patent activity* | =1 | if firm has applied for patent registration during the years 2001 to 2003 |
| *R&D intensity* | | R&D intensity defined as R&D expenditures relative to total turnover (net of intermediate inputs) in 2003 |
| **Treatment variable** | | |
| *R&D subsidies* | =1 | if firm received subsidies either from the federal state Thuringia, national or EU wide programmes or combinations; 0 otherwise |
| **Standard firm specific control variables and skill structure** | | |
| *Size* | | Firm size in terms of total employment |
| *Age* | | Number of years since firm was created, relative to 2004 |
| *Capital* | | Capital intensity defined as total capital stock per employees |
| *Investment* | | Investment intensity defined as total investment per sales in 2003 |
| *Human capital high* | | Share of high skilled employees as share of total employment |
| *Human capital low* | | Share of low skilled employees as share of total employment |
| **Internationalization and regional input-output relations** | | |
| *Regional sales* | | Sales within the core region (30km) relative to total sales, in % |
| *East German sales* | | Sales within East Germany relative to total sales, in % |
| *West German sales* | | Sales within West Germany relative to total sales, in % |
| *Exports* | | Export share in percent, defined as total exports relative to sales, in % |
| *Regional inputs* | | Input from suppliers within the core region (30km) relative to total inputs, in % |
| *East German inputs* | | Input from suppliers from East Germany relative to total inputs, in % |
| *West German inputs* | | Input from suppliers from West Germany relative to total inputs, in % |
| *Imports* | | Import share defined as imports relative to total inputs, in % |
| **Binary dummy variables** | | |
| *Liability* | =1 | if firm owner has full legal liability, 0 for limited liability |
| *West German ownership* | =1 | if firm belongs to a parent company in West Germany |
| *East German ownership* | =1 | if firm belongs to a parent company in East Germany |
| *Foreign ownership* | =1 | if firm belongs to a parent company abroad |
| *Small* | =1 | if firm has less than five employees |
| *R&D department* | =1 | if firm has R&D employees within a fixed R&D department |

**Table 2: Descriptive Statistics**

| Variable | N | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| *Patent activity* | 1412 | 0.141 | 0.348 | 0 | 1 |
| *R&D Intensity* | 1265 | 0.047 | 0.24 | 0 | 6.6 |
| *R&D subsidies* | 1267 | 0.223 | 0.417 | 0 | 1 |
| *Size* | 1431 | 56.67 | 148.03 | 1 | 2947 |
| *Age* | 1279 | 11.07 | 7.79 | 1 | 73 |
| *Capital* | 1238 | 70.97 | 178.26 | 0 | 3875 |
| *Investment* | 986 | 32.05 | 364.80 | 0 | 10831 |
| *Human capital high* | 1385 | 0.19 | 0.24 | 0 | 1 |
| *Human capital low* | 1431 | 0.07 | 0.39 | 0 | 13.79 |
| *Regional sales* | 1382 | 23.64 | 30.69 | 0 | 100 |
| *East German sales* | 1387 | 11.50 | 16.89 | 0 | 100 |
| *West German sales* | 1391 | 39.45 | 31.50 | 0 | 100 |
| *Exports* | 1365 | 12.70 | 22.21 | 0 | 100 |
| *Regional inputs* | 1261 | 23.42 | 28.79 | 0 | 100 |
| *East German inputs* | 1265 | 11.83 | 17.14 | 0 | 100 |
| *West German inputs* | 1269 | 39.76 | 29.68 | 0 | 100 |
| *Imports* | 1242 | 9.82 | 18.66 | 0 | 100 |
| *Liability* | 1426 | 0.275 | 0.447 | 0 | 1 |
| *West German ownership* | 1164 | 0.140 | 0.347 | 0 | 1 |
| *East German ownership* | 1164 | 0.043 | 0.203 | 0 | 1 |
| *Foreign ownership* | 1164 | 0.049 | 0.218 | 0 | 1 |
| *R&D department* | 1362 | 0.277 | 0.448 | 0 | 1 |

assumption that has to be fulfilled is the so called stable unit treatment value assumption (SUTVA). This assumption states that individual causal effects may not be influenced by the participation status of other firms (Angrist, Imbens and Rubin, 1996). In the latter case we rely on the argument raised by Almus & Czarnitzki (2003) that the presence of such distorting effects, which may alter the relative price for R&D factor inputs, is not very likely for East Germany. The main reason is that the pricing mechanisms for R&D factor inputs are assumed to be driven by national and international rather than regional factors. R&D input prices should largely be determined by market forces and rather be independent from policy distortions such as R&D subsidies to the East German economy.

Building on these assumptions a general matching model to measure the effect of the subsidy $\theta$ by the outcome-difference between treated ($D$) and comparison firms ($C$) can be written as

$$\hat{\theta} = \frac{1}{N} \sum_{i \in D} Y_{1i} - \sum_{i \in C} W(i,j) Y_{0j} \tag{1}$$

where $N$ is the total number of treated firms $i$, $Y_1$ and $Y_0$ denote outcome values for treated and comparison firms $j$. Then, the match of each treated firm is constructed as a weighted average over the outcomes of non-treated, where the weights $W(i,j)$ depend on the 'distance' between $i$ and $j$.

For empirical application a decision about the matching routine has to be made. There are several types of matching criteria available, which differ by the specific weighting function $W(i,j)$ used. Most criteria match treated firms only with a certain fraction of the comparison group, where selection is based on the obtained propensity score from the first-step probit model. Prominent examples are nearest-, $k$-nearest neighbor and stratification matching. Other routines like Kernel and Mahalanobis matching use a weighted matching approach based on averaging procedures of the outcomes for all comparison units. As Caliendo & Kopeinig (2005) point out, one major advantage of these latter approaches is the lower variance, which is achieved because more information is used. In the following we use the Kernel matching algorithm, which was introduced by Heckman et al. (1997), as default weighting function since it has also shown to have good finite sample properties e.g. compared to the standard $k$-nearest neighbor matching function (see Fröhlich, 2004).[4] The weighting function in the Kernel based matching algorithm can be defined as:

$$W(i,j) = \frac{K[(P_j(X) - P_i(X))/h]}{\sum_{k \in C} K[(P_k(X) - P_i(X))/h]}, \tag{2}$$

where $K$ is the Kernel function and $h$ is the bandwidth parameter to be chosen.[5] $P(X)$ denotes the propensity score for variable vector $X$. Since we are dealing with rather small numbers of observations we compute bootstrapped standard errors.[6] We also apply a common support restriction to our Kernel matching routine in order to minimize the risk of bad matches.[7]

Table 3 shows the results of the first-step probit estimation. We report marginal effects for different model specifications: In model I ('full') a fairly general specification is estimated. The regression results in table 3 show that only few variables turn out to be significant, while the loss of observations due to missing values is large. In model II we therefore try to reduce the number of variables used in a stepwise regression approach, that starts from a slim model only containing industry dummies and subsequently adds

[4]For estimation we use software codes for Stata by Becker & Ichino (2002) as well as Leuven & Sianesi (2003).

[5]We use the Epanechnikow kernel as default and apply different values for the bandwidth, which are commonly proposed in the literature (see e.g. Jones et al., 1996).

[6]We applied both bootstrapping for just the second step of the matching estimator as well as simultaneously bootstrapping for both steps including the first-step probit estimation. We set the number of bootstrap replications equal to 500. Compared to the results with standard errors based on asymptotically normal statistical inference the bootstrapped SE are somewhat more restrictive and may thus be seen as the more conservative benchmark in evaluating programme effectiveness. For a discussion of bootstrapping standard errors for matching estimators see e.g. Caliendo & Kopeinig (2005).

[7]The underlying idea is that the matching procedure is restricted to those treated observations for which also firms in the comparison group are found with similar values for the propensity score, while treated units whose propensity score value is larger than the largest value in the non-treated pool are left unmatched (for details see e.g. Lechner, 2000).

further variables that are found to be significant at the 10 % significance level. As table 3 shows filtering by statistical significance already leads to a considerable reduction in model size. However, since we allow for a large set of candidate variables, model II still drops many observations.

We thus take model II as the starting point to take out those variables, which show a large number of missings and have been excluded from model II based on statistical significance tests. We nevertheless keep those variables in the candidate set that were tested insignificant in model II but do not restrict the overall sample size. We then take this reduced set of candidate variables and perform a further stepwise regression. The resulting model specification III employs a much higher number of observations. Most variables the were already found to be significant in model II are also selected for model III. Moreover, some variables like firm size and the share of sales to West Germany even turn out significant now.

As the results for model III additionally show, the McFadden $R^2$ is only slightly reduced compared to the full model and even higher than the first stepwise regression result. Moreover, testing for model III being nested in I is clearly rejected. Given the large increase in the numbers of observation versus the rather small loss in terms of model fit, we choose model III as empirical basis for our propensity score calculation. This point was also raised by Augurzky & Schmidt (2001), arguing that the discard of many sample observations may lead to a non-representative matched sample. The obtained fitted values can then be used as weights in order to match pairs of treated and comparison firms.

According to table 3, the probability to receive an R&D subsidy is thus most importantly driven by the skill composition of the firm's employees as well as the presence of an own R&D department. For both variables we get significant and positive marginal effects. The same holds for firm size, while firm's age has a significantly positive influence on the probability to receive subsidies. That is, the younger a firm is, the lower is its probability to receive a subsidy. Finally, regarding regional input-output linkages, a high share of regional sales as well as sales to West German has a negative impact on programme participation. Though being tested significant in model specifications I and II, the dummy for West German ownership turns out insignificant in our final specification and was dropped.

As a first check for the appropriateness of the propensity score method we then perform a mean comparison for our explanatory variables. This allows us to see whether the obtained propensity score is successful in terms of balancing differences for the set of covariates. The results are shown in table 4. They indicate that in general the null hypothesis of identical means for treated and non-treated has to be rejected for the unmatched

**Table 3: Probit estimation for receiving R&D subsidies**

| Dep. Var.: *R&D subsidies* | I Full | II Stepwise | III Stepwise |
|---|---|---|---|
| $log(Size)$ | 0.024 | | 0.064*** |
| | (0.049) | | (0.024) |
| $log(1/Age)$ | -0.295* | -0.106*** | -0.192** |
| | (0.166) | (0.036) | (0.085) |
| $log(1/Age)^2$ | -0.052 | | -0.035* |
| | (0.037) | | (0.019) |
| *Capital* | -0.016 | | |
| | (0.017) | | |
| *Investment* | -0.001 | | |
| | (0.001) | | |
| *Human capital high* | 0.562*** | 0.591*** | 0.356*** |
| | (0.125) | (0.112) | (0.059) |
| *Human capital low* | -0.257 | -0.307* | |
| | (0.176) | (0.174) | |
| *Regional sales* | -0.003* | -0.002** | -0.002*** |
| | (0.0015) | (0.0009) | (0.0005) |
| *East German sales* | -0.001 | | |
| | (0.001) | | |
| *West German sales* | -0.001 | | -0.001* |
| | (0.001) | | (0.0004) |
| *Exports* | -0.001 | | |
| | (0.001) | | |
| *Regional inputs* | 0.003** | | |
| | (0.001) | | |
| *East German inputs* | 0.003* | | |
| | (0.001) | | |
| *West German inputs* | 0.001 | -0.001* | |
| | (0.001) | (0.0007) | |
| *Imports* | 0.002* | | |
| | (0.001) | | |
| *Liability* | 0.064 | | |
| | (0.099) | | |
| *West German ownership* | -0.106** | -0.108** | |
| | (0.039) | (0.039) | |
| *East German ownership* | -0.038 | | |
| | (0.091) | | |
| *Foreign ownership* | -0.061 | | |
| | (0.061) | | |
| *R&D department* | 0.337*** | 0.338*** | 0.345*** |
| | (0.053) | (0.046) | (0.059) |
| *Obs.* | 529 | 529 | 1023 |
| *Industry dummies* | yes | yes | yes |
| *McFadden* $R^2$ | 0.35 | 0.33 | 0.34 |
| *Interaction terms* $\chi^2$ | 17.74 | | |
| (P-$Value$) | (0.17) | | |
| $Non-nested\ test$ $\chi^2$ | | 13.24 | 296.7 |
| (P-$Value$) | | (0.42) | (0.00) |

*Note:* ***, **, * = denote significance levels at the 1%, 5% and 10% level respectively. For the owernship dummy the reference case is being an independent Thuringian firm. All report coefficients are marginal effects. Descriptions of sizes classes and industry dummies are given in the appendix.

sample. After applying propensity score matching the differences vanish as the results in Table 4 show. So, propensity score matching is a successful way to select or weight (depending on the matching procedure) the comparison group in a way that matches similar characteristics.[8]

**Table 4: Mean comparison for treated and comparison group before and after matching**

| Mean of: | **Treated** | **Comparison unmatched** | **Comparison matched** |
|---|---|---|---|
| $log(Size)$ | 3.35 | 2.92*** | 3.29 |
| $t-stat.$ | | (4.52) | (0.27) |
| $log(1/Age)$ | -2.22 | -2.19 | -2.23 |
| $t-stat.$ | | (-0.64) | (-0.26) |
| $log(1/Age)^2$ | 5.24 | 5.32 | 5.23 |
| $t-stat.$ | | (-0.42) | (-0.04) |
| $Human\ capital\ high$ | 0.36 | 0.15*** | 0.36 |
| $t-stat.$ | | (14.38) | (0.18) |
| $Regional\ sales$ | 13.08 | 26.38*** | 12.267 |
| $t-stat.$ | | (-6.35) | (0.52) |
| $West\ German\ sales$ | 43.16 | 38.45** | 40.80 |
| $t-stat.$ | | (2.18) | (-0.45) |
| $R\&D\ department$ | 0.72 | 0.16*** | 0.73 |
| $t-stat.$ | | (21.33) | (-0.17) |
| **Treated** | | 283 | 224 |
| **Comparison** | | 984 | 792 |

*Note:* ***, **, * = denote significance levels at the 1%, 5% and 10% level respectively. Statistical significance was tested in a two-tailed t-test between the supported firms and either firms from the total of controls or from the propensity score based selected comparison group (t-values in brackets).

Results for the Kernel matching procedure on the private R&D outcome variables (R&D intensity and patent activity) are shown in table 5. First, we look at the effect for the full sample of all firms. The results show a significantly positive difference between the R&D activity of treated and comparison firms. This is a first important indication that R&D subsidies do not fully lead to substitution effects, but instead increase the level of R&D intensity and of the probability for patent applications. Regarding the R&D intensity, the results show that the average R&D activity of treated firms relative to the comparison group is around 2.4 to 2.8 percentage points higher (depending on the chosen bandwidth parameter of the Kernel function). Since the difference in the R&D intensity is measured in terms of logarithms, the latter effect can be calculated from $\hat{\theta}$ in table 5 as follows: Taking the estimated coefficient of 0.93 for all firms based on Kernel matching

[8]Since the definition of the treated and comparison group changes for different outcome variables and matching designs, we did run a specific test in each case. The results nevertheless hold uniformly. Detailed testing outputs can be obtained from the authors upon request.

with a bandwidth of 0.06, the difference between the treated and comparison group can be calculated as $0.93 * exp(0.93) \approx 2.4$ percentage points. Alternatively to R&D intensity as private sector outcome variable, we also use the binary dummy variable for patent applications. The results in table 5 show for the sample of all firms that the probability for a treated firm to apply for a patent is about 20 percentage points higher relative to the comparison group.

**Table 5: Kernel matching for private R&D activity**

| | **bandwidth = 0.06** | | **bandwidth = 0.12** | |
|---|---|---|---|---|
| | **log(R&D intensity)** | **Patent activity** | **log(R&D intensity)** | **Patent activity** |
| | | **All Firms** Treated = 207, Comparison = 745 | | |
| $\hat{\theta}$ | 0.933*** | 0.199*** | 1.014*** | 0.197*** |
| | (0.181) | (0.048) | (0.195) | (0.047) |
| | | **Medium** ($50 < size \leq 250$) Treated = 66, Comparison = 160 | | |
| $\hat{\theta}$ | 0.769*** | 0.248*** | 0.813*** | 0.229** |
| | (0.295) | (0.091) | (0.308) | (0.094) |
| | | **Small** ($20 < size \leq 50$) Treated = 83, Comparison = 308 | | |
| $\hat{\theta}$ | 1.043*** | 0.242*** | 1.043*** | 0.246*** |
| | (0.321) | (0.069) | (0.285) | (0.071) |
| | | **Micro** ($1 \leq size \leq 20$) Treated = 50, Comparison = 253 | | |
| $\hat{\theta}$ | 1.696*** | 0.085 | 1.467*** | 0.062 |
| | (0.608) | (0.102) | (0.527) | (0.099) |

*Note:* ***, **, * = denote significance levels at the 1%, 5% and 10% level respectively. Bootstrapped standard errors (in brackets) are calculated based on 500 repetitions.

We then estimate the causal impact of funding for three different subgroups of SMEs. Starting with medium sized firms, measured in terms of 51 to 250 employees, table 5 shows that the effect on private R&D intensity is also significantly positive, but smaller compared to the average of all firms. We estimate a difference in the R&D intensity of about 1.7 to 1.8 percentage points. The increase in the probability for patent application is slightly bigger compared to the average of all firms (around 23 to 24 percentage points). While the effect for small firms with 11 up to 50 employees is found to be in line with the average estimate for all firms (around 2.9 percentage points for R&D intensity, 24 percentage points for patent activity), the results for the subgroup of micro firms with up to 10 employees are rather specific: Here we estimate a much higher increase in the R&D intensity (more than 6 percentage points), while we do not observe any significant difference in the patent activity relative to the comparison group of non-subsidized micro firms.

The latter insignificant result for very small firms may reflect differences in the innovation strategies between small and large firms. Conte & Vivarelli (2005), for instance, find for Italian micro data between 1998 and 2000 that large firms heavily rely on own R&D innovative effort, while small firms more actively participate in cooperation agreements and business groups including the acquisition of external technology in the process of developing product and process innovations rather than choosing costly and time consuming patent strategies. Ball & Kesan (2009) argue that due to high transaction costs of litigation, small firms may not be able to effectively monitor their patent rights and thus consequently choose not to apply for patents at all. Both observations give a qualitative argument, why input and output related R&D activity measures react differently to the policy stimulus by firm size.[9] One should thus be careful, in using the latter result as an argument in favour for policy ineffectiveness of R&D subsidies. The strong increase in the R&D intensity for micro firms rather hints at a large leverage effect in terms of crowing-in private R&D inputs by the grant schemes. This latter high outcome difference in turn is likely to be driven by firms that start R&D activities for the first time (induced by public fundings), which is likely to be associated with a certain threshold effect.

## 4 Robustness Checks

To check the sensitivity of the results with respect to the chosen matching routine, we use a broad set of alternative weighting schemes for the matching procedure. The results for R&D intensity are shown in table 6 and for patent activity in table 7. We apply stratification matching, $k = 5$ nearest-neighbor matching with an additional caliper restriction in terms of one fourth of the standard error, as well as Mahalanobis metric distance matching. The latter algorithm allows including additional matching information besides the propensity score (such as the 2-digit industry classification and firm size categories). All the procedures are again subject to the common support restriction.

Table 6 and Table 7 show that the results are very stable for different weighting schemes. With respect to R&D subsidies we find a statistically significant result for combinations of sample design and chosen matching routine. Depending on the weighting scheme, the overall effect for the sample of all firms varies from 2.4 to 5 percentage points (with the highest effect found by the 5-nearest neighbor routine). Moreover, as highlighted in figure 1, all matching routines uniformly estimate the effect for micro firms to be higher compared to the samples of all, medium sized and small firms. For patent activity the

[9]Additionally, rather long time lags in the transmission from R&D inputs to R&D outputs have to be considered, which are not captured by the sample data.

estimated outcome difference for all firms varies between 20 and 27 percentage points and is also highly significant for all matching routines. Also, all algorithms report insignificant results for the sample of micro firms.

**Table 6: Sensitivity analysis for** *log*(*R&D intensity*)

| **Dep. Var.:** *R&D intensity* | **Stratification** *blocks* = 7 | **5-NN Caliper** $\eta = 0.25 \times \sigma_{PS}$ | **Mahalanobis** *PS*, *Industry* | *PS*, *Ind*, *Size* |
|---|---|---|---|---|
| | | **All Firms** | | |
| $\hat{\theta}$ | 1.173*** | 1.360*** | 1.222*** | 1.186*** |
| | (0.215) | (0.198) | (0.225) | (0.218) |
| | | **Medium** ($50 < size \leq 250$) | | |
| $\hat{\theta}$ | 0.891*** | 0.927*** | 0.765* | 0.974** |
| | (0.294) | (0.266) | (0.416) | (0.392) |
| | | **Small** ($10 < size \leq 50$) | | |
| $\hat{\theta}$ | 1.220*** | 1.305*** | 1.254*** | 1.064*** |
| | (0.380) | (0.291) | (0.350) | (0.359) |
| | | **Micro** ($1 \leq size \leq 10$) | | |
| $\hat{\theta}$ | 1.192** | 1.909*** | 1.434** | 1.435** |
| | (0.582) | (0.479) | (0.618) | (0.649) |

*Note:* ***, **, * = denote significance levels at the 1%, 5% and 10% level respectively. For the kernel matching we use the Epanechnikow kernel as default. *PS* denotes Propensity Score. Standard errors in brackets.

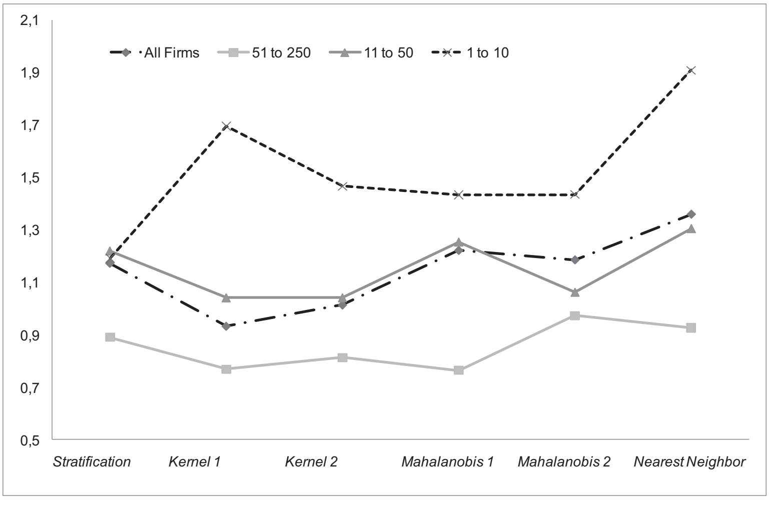
**Table 7: Sensitivity analysis for** *Patent activity*

| **Dep. Var.:** *Patent activity* | **Stratification** *blocks* = 7 | **5-NN Caliper** $\eta = 0.25 \times \sigma_{PS}$ | **Mahalanobis** *PS*, *Industry* | *PS*, *Ind*, *Size* |
|---|---|---|---|---|
| | | **All Firms** | | |
| $\hat{\theta}$ | 0.189*** | 0.209*** | 0.210*** | 0.272*** |
| | (0.046) | (0.045) | (0.061) | (0.065) |
| | | **Medium** ($50 < size \leq 250$) | | |
| $\hat{\theta}$ | 0.295*** | 0.258*** | 0.302** | 0.349*** |
| | (0.080) | (0.082) | (0.116) | (0.113) |
| | | **Small** ($10 < size \leq 50$) | | |
| $\hat{\theta}$ | 0.240*** | 0.233*** | 0.312*** | 0.287*** |
| | (0.075) | (0.068) | (0.078) | (0.087) |
| | | **Micro** ($1 \leq size \leq 10$) | | |
| $\hat{\theta}$ | 0.023 | 0.086 | 0.01 | 0.01 |
| | (0.107) | (0.0083) | (0.111) | (0.111) |

*Note:* ***, **, * = denote significance levels at the 1%, 5% and 10% level respectively. For the kernel matching we use the Epanechnikow kernel as default. *PS* denotes Propensity Score. Standard errors in brackets.

Since the econometric literature does not offer clear cut guidance with respect to the preferable matching estimator in empirical applications, we finally calculate a mean estimator defined as equally weighted average over all matching algorithms according to

**Figure 1: Estimated $\hat{\theta}$ for R&D intensity by matching algorithms and size groups**




*Note:* For details see table 6.

$\hat{\theta}_{mean} = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_i$ with $i$ as number of matching estimators (in our case $i = 6$). The results, together with their t-statistics, are reported in table 8. For the sample of all firms we get a mean estimate for the R&D intensity of about 3.7 percentage points, which comes pretty close to the findings in Almus & Czarnitzki (2003) as well as Czarnitzki & Licht (2006), who find an increase of roughly 4 percentage points for the R&D intensity in East Germany between 1996–1998 and 1994–2000 respectively. The estimated mean effect for patent activity is 21.6 percentage points and also comes close to the marginal effect of 22 % reported in Czarnitzki & Licht (2006). The mean effect for micro firms is with 7.4 percentage points the highest, for small firms with 11 up to 50 employees we get an increase of 3.7 percentage points, for medium sized firms up to 250 employees a moderate effect of 2 percentage points. The effect on patent activity varies between 21 and 28 percent for small, medium and all firms. Only for micro firms the effect is estimated to be statistically insignificant. The mean estimation results thus qualitatively support the findings from the Kernel estimation shown above.

Using the estimated outcome effect from table 8, we derive a rough proxy for the leverage effect of the public funding on private sector R&D inputs. To do so, we take the mean R&D intensity for non-subsidized firms as benchmark level and compute the leverage effect as ratio of outcome effect (net of public co-financing) relative to the benchmark level. Since we are lacking information on the absolute amounts of public expenditures, we assume by 'rule-of-thumb' that the latter is about 50 % of the outcome effect. This approximation is a very conservative estimate, since it assumes that there is a strict linear relationship between private and public spendings in the determination of the firm's R&D intensity. That is, any second round effects of public subsidies on private inputs are ignored. For the full sample of all firms and small firms with size between 11 and 50 employes the results show that a doubling of public spendings also leads to a equiproportional doubling in private inputs. The impact is the highest for micro firms, indicating that a doubling of R&D subsidies, leads to a leverage effect of 250%Ėven if we reasonably assume that the share of public spendings for this subgroup is actually higher, let's say a 'rule-of-thumb' of 80 % public co-finaning, this would still yield a leverage effect for private inputs of roughly 100 %. The effect is the smallest for medium-sized firms up to 250 employees. However, as pointed out, these results should only be interpreted carefully as lower bounds of the actual leverage effect.

**Table 8: Mean estimates of policy effect for private R&D activity (in percentage points)**

| | **R&D intensity** | **Patent activity** |
|---|---|---|
| **All Firms** | 3.71*** | 21.26*** |
| $t-stat.$ | (5.59) | (4.08) |
| **Medium** ($50 < size \leq 250$) | 2.03*** | 28.01*** |
| $t-stat.$ | (2.61) | (2.91) |
| **Small** ($10 < size \leq 50$) | 3.72*** | 26.00*** |
| $t-stat.$ | (3.48) | (3.48) |
| **Micro** ($1 \leq size \leq 10$) | 7.40*** | 4.3 |
| $t-stat.$ | (2.64) | (0.42) |

*Note:* ***, **, * = denote significance levels at the 1%, 5% and 10% level respectively. Coefficients and t-values (in brackets) are calculated as mean estimates according to $\hat{\theta}_{mean} = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_i$ with $i$ as number of matching estimators (in our case $i = 6$) and similar for the standard error.

As a final robustness check, we also restrict our sample only to those firms, who are permanently performing R&D activity proxied by running an own R&D department. The motivation for this restriction is simply to 'raise the bar'. That is, since firms with permanent R&D activity should have an on average higher R&D intensity and probability for patent application, this will lead to a tighter selection of the comparison group and thus may shed additional light on the question of R&D policy effectiveness. As Czartnitzki & Licht (2006) argue, the latter setup is important since it compares treated and comparison firms that are more equal in their structural characteristics, nevertheless it is also likely to underestimate the effect of R&D promotion, as it is implicitly assumed that the subsidy alone is not able to motivate firms to start R&D activities.

Our results show that – both for R&D intensity as well as patent activity – we still estimate a significant positive impact of R&D subsidies supporting the general effectiveness of the analyzed grant schemes for the case of East Germany. As expected, the effect for R&D intensity is smaller compared to the case of the unrestricted comparison group. Whether this smaller increase may indicate a partial crowding-out of private spendings or simply underestimates the total effect, cannot be distinguished at this point. The estimated effect for the probability of patent applications is within the range of the outcome differences found for the full sample of all firms. Since both treatment and comparison group are rather small for permanently R&D active firms, a further disaggregation into SME subaggregates is not feasible.

- **Outcome difference for firms permanently performing R&D activity**

  1. *R&D Intensity*
     - $\hat{\theta}_{mean} = 0.644$, $SE_{\hat{\theta}_{mean}} = 0.258$, $t - value = 2.49$
     - Treated = 116, Comparison = 153
  2. *Patent Activity*
     - $\hat{\theta}_{mean} = 0.190$, $SE_{\hat{\theta}_{mean}} = 0.067$, $t - value = 2.81$
     - Treated = 116, Comparison = 153

## 5 Conclusion

We have analysed the impact of public support schemes on private R&D activity for a large cross-section of East German firms in 2003. Using different matching routines to identify the causal effect of public funding, we estimate a significant positive effect of public support on private R&D activity. Our mean estimates show that subsidized firms have on average about 3.7 percentage points higher R&D intensity than non-subsidized firms. This result comes close to earlier findings reported in Almus & Czarnitzki (2003) as well as Czarnitzki & Licht (2006), who find a roughly 4 percentage point higher R&D intensity for East German firms throughout the second half of the 1990s. Also, the estimated increase of 22 percentage points in the probability for patent application is similar to the result reported in Czarnitzki & Licht (2006).

While our results may thus be seen as a backup and robustness check for earlier empirical evidence on East Germany based on a different dataset, so far no empirical evidence is reported with respect to the effectiveness of R&D subsidies for small and medium-sized firms. Here our study fills an important gap in the empirical literature. The focus on SMEs is particularly relevant for East Germany, since its regional innovation system is largely driven by the latter group. Our results show, that the positive effects found for the total sample of firms also hold for three subgroups of SMEs (micro firms with 1 up to 10 firms, small firms with 21 and up to 50 employees, as well as medium sized firms with 51 and up to 250 employees). Regarding the R&D intensity as outcome variable, the biggest increase is found for micro firms. Here the policy schemes seem to be particularly successful in activating firms to start R&D activity for the first time. For small and medium-sized firms the effect is also found to be positive but of smaller size.

With respect to patent activity, treated small and medium sized firms show a significant increase in the probability of patent application, while we do not find any significant effect for micro firms. The latter insignificant result may reflect differences in the innovation strategies between small and large firms, with very small firms choosing different innovation strategies rather than patenting (e.g. due to high transaction costs of litigation small firms may not be able to effectively monitor their patent rights and thus consequently choose not to apply for patents at all). It does not seem reasonable to take this result as an indicator for policy ineffectiveness.

We have also investigated whether subsidies add to private R&D inputs, when only those firms regularly performing R&D activity are analysed. Though the effect is smaller compared to the total sample of all firms, the significant positive findings hold both for the R&D intensity as well as patent activity. Finally, we aimed at proxying the leverage effect of public subsidies on private inputs. Using the mean of non-subsidized firms as a benchmark level and a 'rule-of-thumb' for public co-financing, we get that a doubling of public subsidizes also leads to an equiproportional increase in private R&D spendings. This effect is even higher for micro firms. Taken together, our results hint at the complementary nature of public subsidies to private sector R&D activity in the East German regional innovation system, which is mainly driven by innovative small and medium-sized firms.

## References


[1] **Alecke, B.; Mitze, T.; Untiedt, G. (2010):** 20 Jahre deutsche Einheit - Ist die ostdeutsche Wirtschaft immer noch 'krank'? In: List Forum fuer Wirtschafts- und Finanzpolitik, Vol. 36, pp. 1-26.

[2] **Almus, M.; Czarnitzki, D. (2003):** The Effects of Public R&D Subsidies on Firms' Innovation Activities: The Case of Eastern Germany. In: Journal of Business  Economic Statistics, Jg. 21, H. 2, pp. 226-236.

[3] **Angrist, J.; Imbens, G.; Rubin, D. (1996):** Identification of Causal Effects Using Instrumental Variables. In: Journal of Econometrics, Vol. 71, No. 1-2, pp. 145-160.

[4] **Augurzky, B.; Schmidt, C. (2001):** The Propensity Score: A Means to An End. IZA Discussion Paper No. 271.

[5] **Ball, G.; Kesan, J. (2009):** Transaction Costs and Trolls: Strategic Behavior by Individual Inventors, Small Firms and Entrepreneurs in Patent Litigation. Illinois Law and Economics Papers Series Research Papers Series No. LE09-005.

[6] **Barro, R.; Sala-i-Martin, X. (2004):** Economic growth. 2nd. edition, Cambridge.

[7] **Becker, S.; Ichino, A. (2002):** Estimation of Average Treatment Effects Based on Propensity Scores. The State Journal, Vol. 2(4), pp.358-377.

[8] **Caliendo, M.; Kopeining, S. (2005):** Some Practical Guidance for the Implementation of Propensity Score Matching. IZA Discussion Paper No. 1588.

[9] **Conte, A.; Vivarelli, M. (2005):** One or Many Knowledge Production Functions? Mapping Innovative Activity Using Microdata. IZA Discussion Paper No. 1878.

[10] **Czarnitzki, D.; Licht, G. (2006):** Additionality of public R&D grants in a transition economy. In: Economics of Transition, Jg. 14, H. 1, pp. 101-131.

[11] **David, P.; Hall, B.; Toole, A. (2000):** Is public R&D a complement or substitute for private R&D? A review of econometric evidence. In: Research Policy, Jg. 29, pp. 497-429.

[12] **Ebling, G.; Hipp, C.; Janz, N.; Licht, G.; Niggeman, H. (1999):** Innovationsaktivitaeten im Dienstleistungssektor: Ergebnisse der Innovationserhebung 1997. In: Janz, N.; Licht, G. (Eds.) (1999): Innovationsaktivitaeten in der deutschen Wirtschaft, Baden-Baden.

[13] **Fröhlich, M. (2004):** Finite-Sample Properties of Propensity-Score Matching and Weighting Estimators. In: The Review of Economics and Statistics, Vol. 86(1), pp. 77-90.

[14] **Garcia-Quevedo, J. (2004):** Do Public Subsidies Complement Business R&D? A Meta-Analysis of the Econometric Evidence. In: Kyklos, Vol. 57, pp. 87-102.

[15] **GEFRA, MR, TraSt (2004):** Ergebnisse der Unternehmensbefragung Investitionsfoerderung und Regionale Wirtschaftsfoerderung' im Rahmen der Gemeinschaftsaufgabe Verbesserung der Regionalen Wirtschaftsstruktur' 1997 - 2003 in Thueringen, Muenster.

[16] **GEFRA, GEWIMAR, TraSt (2005):** Evaluierung der Foerderung von Verbundprojekten im Bereich Forschung und Entwicklung im Freistaat Thueringen 1997 - 2003, Muenster.

[17] **Guenther, J.; Wilde, K.; Sunder, M.; Titze, M. (2010):** 20 Jahre nach dem Mauerfall: Strken, Schwchen und Herausforderunge des ostdeutschen Innovationssystems heute. Studien zum deutschen Innovationssystem Nr. 17-2010, IWH.

[18] **Hall, B. H.; van Reenen, J. (2000):** How effective are fiscal incentives for R&D? A review of the evidence. In: Research Policy, Vol. 29(4-5), pp. 449-470,

[19] **Heckman, J.; Ichimura, H.; Todd, P. (1997):** Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. In: Review of Economic Studies, Vol. 64, pp. 605-654.

[20] **Janz, N.; Ebling, G.; Gottschalk, S.; Niggemann, H. (2001):** The Mannheim innovation panels (MIP and MIP-S) of the Centre for European Economic Research (ZEW). In: Schmollers Jahrbuch, 121, S.123-9

[21] **Jones, M.; Marron, J.; Sheather, S. (1996):** A Brief Survey of Bandwidth Selection for Density Estimation. In: Journal of the American Statistical Association, Vol. 91, No. 433, pp. 401-407.

[22] **Kaiser, U. (2004):** Private R&D and Public R&D subsidies: Microeconometric Evidence from Denmark. CEBR-Discussion Paper No. 19.

[23] **Lechner, M. (2000):** A note on common support problem in applied evauation studies. Economics Discucssion Paper 2001-01, University of St. Gallen.

[24] **Leuven, E.; Sianesi, B. (2003):** PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing and Covariate Imbalancing Testing. Boston College Software Code s43200.

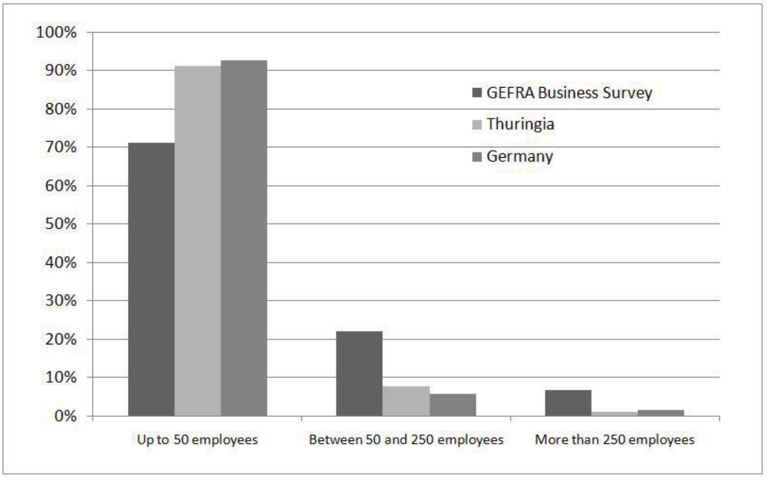
[25] **McAdam, R.; McConvery, T.; Armstrong, G. (2004):** Barriers to innovation within small firms in a peripherical location. In: International Journal of Entrepreneurial Behaviour & Research, Vol. 10, No. 3, pp. 206-221.

[26] **Rammer, C.; Czarnitzki, D. (2003):** Innovationen und Gruendungen in Ostdeutschland. Studien zum deutschen Innovationssystem 15-03, Mannheim.

# A Testing for Sample Representativeness

Testing sample representativeness is an important pre-condition for empirical application. We use both graphical and statistical approaches to check for the representativeness of the GEFRA Business Survey. To start with, figure A.1 plots the proportion of firms for the size classes "up to 50 employees", "between 50 and 250 employees" as well as "more than 250 employees" in the GEFRA Business Survey as well as the total population of firms in the Manufacturing Sector in Thuringia for the year 2003. As the figure shows, though the response rate of small firms in the survey is somewhat smaller compared to the actual population of firms in Thuringia and Germany, however the GEFRA Business Survey still covers a very large number of firms with up to 50 employees. Additionally, figure A.2 plots the proportion of firm for each 2-digit manufacturing subsector relative to the total manufacturing sector for both the sample distribution and the total population of all firms in Thuringia for the year 2003. As the figure shows the sample distribution closely follows the pattern of the total population of Thuringian employees in descending order of their relative importance.

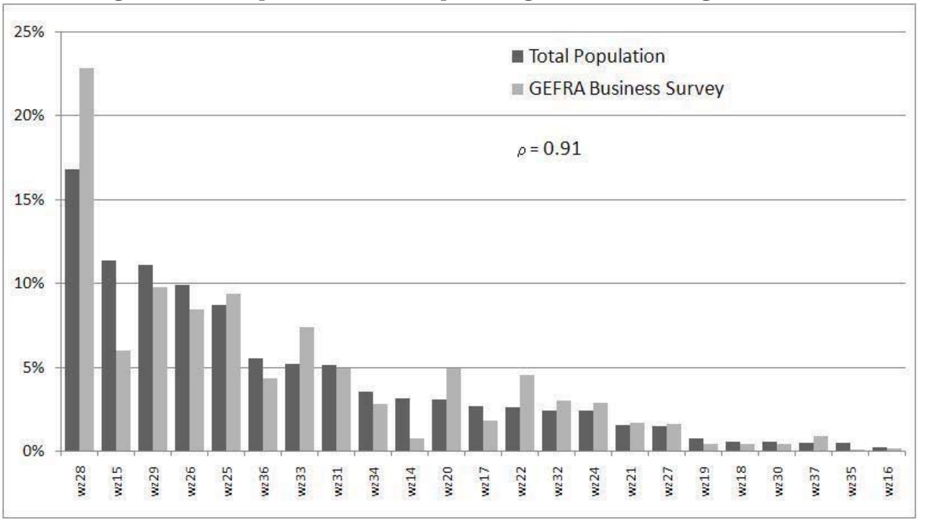
**Figure A.1: Proportions of Firms by Size Class**




*Source:* Data from the German Statistical Office, GEFRA-Business Survey 2004.

Next to the above described descriptive statistics we also test for sample deviations from the total population using a standard Z-statistic based test, which compares the

**Figure A.2: Proportions of firms per 2-digit Manufacturing subsectors**




*Source:* Data from the German Statistical Office, GEFRA Business-Survey 2004.
*Note:* Based on 2-digit Manufacturing Sector data for the year 2003.

sample distribution relative to the overall distribution of the population. Since the total population of Thuringian Manufacturing Sector firms in 2003 is known the proportion based test is given by:

$$Z = \frac{\rho - P}{\sigma_\rho}, \tag{3}$$

where $\rho$ is the sample population proportion, $P$ is the population proportion and $\sigma_\rho$ is the standard error of the proportion given by $\sigma_\rho = \sqrt{\frac{P(1-P)}{n}}$, where $n$ is the number of observations in the respective (sub-)sample. We apply the test for for sample and population proportions of each 2-digit industry. Table A.1 plots the respective sector shares together with the standard error of the sample proportion and the corresponding $Z$-statistic. As the results show, the $Z$-statistic indicates a statistically significant deviation of the sample distribution from the overall population only for one single sub-sector, WZ28 "Manufacture of fabricated metal products, except machinery and equipment" being overrepresented in the GEFRA Business-Survey. However, taken together the results support the representativeness of our survey data.

**Table A.1: Test for Sample Representativeness based on Manufacturing Sector Firm Data**

| WZ Code | Pop. | Share Pop. | Sample | Share Sample | $\sigma_s$ | Z-Statistic |
|---|---|---|---|---|---|---|
| 14 | 59 | 3,15 % | 10 | 0,76 % | 0,06 | -0,43 |
| 15 | 213 | 11,38 % | 79 | 6,04 % | 0,04 | -1,50 |
| 16 | 4 | 0,21 % | 2 | 0,15 % | 0,03 | -0,02 |
| 17 | 51 | 2,73 % | 24 | 1,83 % | 0,03 | -0,27 |
| 18 | 11 | 0,59 % | 6 | 0,46 % | 0,03 | -0,04 |
| 19 | 14 | 0,75 % | 6 | 0,46 % | 0,04 | -0,08 |
| 20 | 58 | 3,10 % | 65 | 4,97 % | 0,02 | 0,87 |
| 21 | 29 | 1,55 % | 22 | 1,68 % | 0,03 | 0,05 |
| 22 | 49 | 2,62 % | 60 | 4,58 % | 0,02 | 0,95 |
| 23 | 0 | 0,00 % | 0 | 0,00 % | 0,00 | 0,00 |
| 24 | 45 | 2,41 % | 38 | 2,90 % | 0,02 | 0,20 |
| 25 | 163 | 8,71 % | 123 | 9,40 % | 0,03 | 0,27 |
| 26 | 186 | 9,94 % | 111 | 8,48 % | 0,03 | -0,51 |
| 27 | 28 | 1,50 % | 21 | 1,60 % | 0,03 | 0,04 |
| 28 | 314 | 16,78 % | 299 | 22,84 % | 0,02 | 2,80*** |
| 29 | 208 | 11,12 % | 128 | 9,78 % | 0,03 | -0,48 |
| 30 | 11 | 0,59 % | 6 | 0,46 % | 0,03 | -0,04 |
| 31 | 96 | 5,13 % | 65 | 4,97 % | 0,03 | -0,06 |
| 32 | 46 | 2,46 % | 40 | 3,06 % | 0,02 | 0,24 |
| 33 | 97 | 5,18 % | 97 | 7,41 % | 0,02 | 0,99 |
| 34 | 66 | 3,53 % | 37 | 2,83 % | 0,03 | -0,23 |
| 35 | 9 | 0,48 % | 1 | 0,08 % | 0,07 | -0,06 |
| 36 | 104 | 5,56 % | 57 | 4,35 % | 0,03 | -0,40 |
| 37 | 10 | 0,53 % | 12 | 0,92 % | 0,02 | 0,18 |

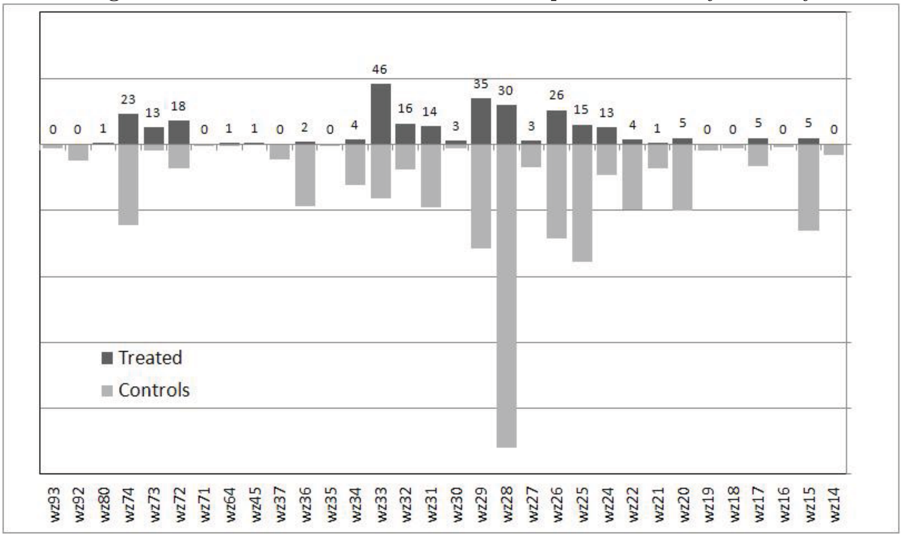
*Note:* ***, **, * = denote significance levels at the 1%, 5% and 10% level respectively. The test is based on the sectoral propotion of employees in the total population and sample proportion. For details about the applied Z-statistic see text.

## B Supplementary Descriptive Statistics

**Table A.2: Variable description for size classes and 2-digit industries**

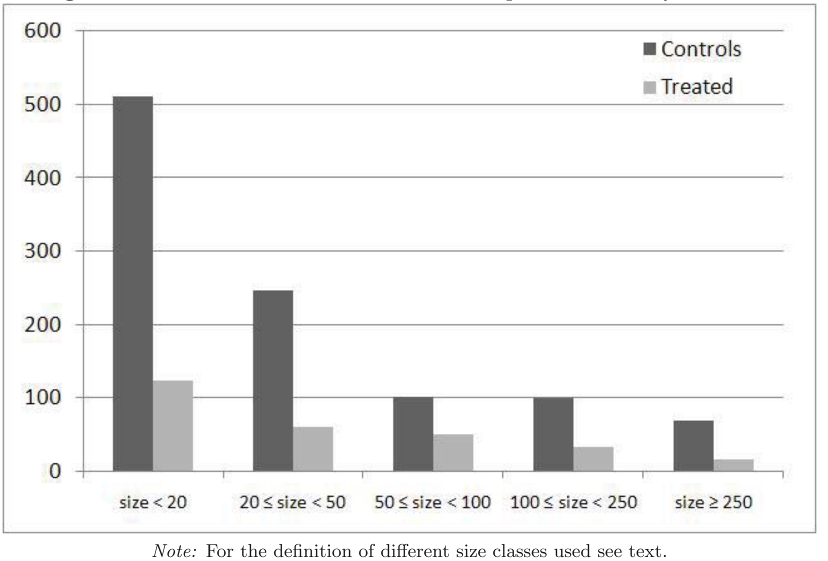
| Variable | | Description |
|---|---|---|
| **Sizes Classes** | | |
| $Size < 20$ | =1 | if the firm has less than 20 employees; 0 otherwise |
| $Size < 50$ | =1 | if the firm has between 20 and 50 employees; 0 otherwise |
| $Size < 100$ | =1 | if the firm has between 50 and 100 employees; 0 otherwise |
| $Size < 250$ | =1 | if the firm has between 100 and 250 employees; 0 otherwise |
| $Size > 250$ | =1 | if the firm has 250 or more employees; 0 otherwise |
| **2-digit Industry dummies** | | |
| $Ind15/16$ | =1 | if firm belongs to industry 15 and 16 according to the German classification of Economic Activities WZ2008 (Manufacture of food products and beverages and tobacco products) |
| $Ind17/19$ | =1 | if firm belongs to industry 17 to 19 Manufacture of textiles, of wearing apparel; dressing and dyeing of fur; Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear) |
| $Ind20$ | =1 | if firm belongs to industry 20 (Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials) |
| $Ind21/22$ | =1 | if firm belongs to industry 21 and 22 (Manufacture of pulp, paper and paper products; Publishing, printing and reproduction of recorded media) |
| $Ind24$ | =1 | if firm belongs to industry 24 (Manufacture of chemicals and chemical products) |
| $Ind25$ | =1 | if firm belongs to industry 25 (Manufacture of rubber and plastic products) |
| $Ind26$ | =1 | if firm belongs to industry 26 (Manufacture of other non-metallic mineral products) |
| $Ind27/28$ | =1 | if firm belongs to industry 27 and 28 (Manufacture of basic metals; of fabricated metal products, except machinery and equipment) |
| $Ind29$ | =1 | if firm belongs to industry 29 (Manufacture of machinery and equipment n.e.c.) |
| $Ind30/33$ | =1 | if firm belongs to industry 30 to 33 (Manufacture of office machinery and computers; of electrical machinery and apparatus n.e.c.; of radio, television and communication equipment and apparatus; and of medical, precision and optical instruments, watches and clocks) |
| $Ind34/35$ | =1 | if firm belongs to industry 34 and 35 (Manufacture of motor vehicles, trailers and semi-trailers; of other transport equipment) |
| $Ind36/37$ | =1 | if firm belongs to industry 36 and 37 (Manufacture of furniture; manufacturing n.e.c.; recycling) |
| $Ind72/74$ | =1 | if firm belongs to industry 72 to 74 (Computer and related activities; Research and development; Other business activities) |

**Figure A.3: Distribution of treated and comparison firms by industry**




*Note:* For the definition of different industry dummies used see table 2.

**Figure A.4: Distribution of treated and comparison firms by size class**




*Note:* For the definition of different size classes used see text.