Employment Effects of Offshoring and FDI - Comparing Measures and Methods

Christoph Moser*
ETH Zurich

Steffen Sirries†
RWTH Aachen University

Dieter M. Urban‡
RWTH Aachen University and CESifo

Preliminary Version

Abstract

This paper investigates the employment effects of offshoring or FDI on German establishment data to explore systematically in a unified framework why some previous studies found positive and others negative employment effects. We compare different measures for offshoring or FDI, different estimation methods, and different sets of control or selection variables. We find positive employment effects from FDI, market-seeking FDI, and even from cost-saving FDI, but negative employment effects from relocation abroad. Hence, the choice of treatment variable rather than the estimation method, or the choice of control or selection variables is responsible for diverse results in the previous literature.

JEL-Classification: F16, F21, F23
Keywords: offshoring, relocation, FDI, matching estimator, employment

*KOF Swiss Economic Institute, ETH Zurich, Weinbergstrasse 35, 8092 Zurich, Switzerland. E-mail: moser@kof.ethz.ch

†Faculty of Business and Economics, RWTH Aachen University, Templergraben 64 (IV), 52062 Aachen, Germany. E-mail: steffen.sirries@rwth-aachen.de

‡Faculty of Business and Economics, RWTH Aachen University, Templergraben 64 (IV), 52062 Aachen, Germany. E-mail: dieter.urban@wiwi.rwth-aachen.de

1 Introduction

The recent two decades of globalization have been vividly debated in the media and in politics (e.g. Mankiw & Swagel (2006)), fearing the exodus of production jobs in highly developed countries. But was the economic impact really as large as the intensity of the debate on it? At the very heart of the recent wave of globalization (Baldwin (2006)) was the unbundling of the production process, a shift of production steps to locations with lower costs.

Contrary to the strong media perception, academic research on employment effects of offshoring or correlated measures of FDI are ambiguous.[1] A significant number of studies even finds positive employment effects contrary to the perception of the public and contrary to theory. A potential explanation is that offshoring has not been systematically measured by statistical offices and researchers have to refer to proxy variables, which may erraneously include many events other than offshoring.[2]

This study investigates on establishment data for Germany whether ambiguous employment effects of offshoring arise from using different proxy variables of offshoring and FDI, or whether different estimation methods, different samples, or different control/selection variables are responsible for the ambiguity.

Considering only employment studies with microdata and firm-level measures of offshoring or FDI, these studies differ by the choice of the employed measure of FDI or offshoring: new investments abroad (Barba Navaretti & Castellani (2004), Mattes (2010)), expansion of employment in foreign affiliates (Becker & Muendler (2008)), increase in intermediate input purchases from abroad (Biscourp & Kramarz (2007), Moser et al. (2009)), increase in usage of intermediate inputs combined with domestic plant restructuring (Moser et al. (2009)), relocation (Wagner (2009)).

These studies differ also by the countries on which data were drawn: Italy (Barba Navaretti & Castellani (2004)), France (Biscourp & Kramarz (2007)), and Germany (all other above mentioned studies). They differ further by the estimation technique: OLS, dynamic panel data, or, in most studies, matching estimators, where control or selection variables again differ almost in each study.

As large as the range of choices in study design are, so large is the range of results,

[1] Crinó (2009) for a survey.

[2] For example, the survey Statistisches-Bundesamt (2008) by the German Statistical Office of was a one-time event to fill the gap.

with strong positive employment effects from foreign employment expansion (Becker & Muendler (2008)) on one end, and slightly negative employment effects in some sample subgroup (Biscourp & Kramarz (2007)), or when not excluding outliers (Wagner (2009)), or when interacting offshoring treatment with partial plant closure events (Moser et al. (2009)) on the other end.

We employ a unified framework on a unique dataset of German establishments that experienced offshoring during the time period 2004-2006 and compare several offshoring measures. On one hand, we apply FDI, market-seeking FDI and cost-saving FDI measures, which were previously studied by Mattes (2010), albeit using different control variables and different estimation methods. On the other hand, we apply a measure of relocation abroad, which is similar to a measure used by Wagner (2009), albeit the data period is different, and the data differ by the observational unit (firm vs. establishment), their coverage, and their data quality (response rates and missing values). To compare methods, we apply both OLS estimators and matching techniques. To keep results comparable, we use three different sets of selection variables to determine the probability of FDI or relocation abroad of a plant for the matching methods.

Despite a unified data framework, we find significant positive employment effects from FDI and negative significant employment effects from relocation abroad. The latter result is the first robust evidence for significant employment losses from one type of foreign activity, relocation abroad. Moreover, the disparity of results on the two types of measures of offshoring or FDI does neither hinge on differences in estimation methods (OLS vs. matching), nor on the choices of selection or control variables. We explain this disparity of results by the variety of activities that are captured by these different measures. None of the measures captures only one single type of FDI. For example, the FDI measure may comprise horizontal FDI, vertical FDI, export platform FDI, etc.; relocation abroad may also consist of horizontal or vertical FDI. Some of these activities may occur in the vein of a general expansion of a firm both abroad, but also at home. This may explain why most FDI measures, even those of cost-saving FDI may go hand in hand with domestic employment expansion. Only in some cases, an expansion abroad substitutes for domestic production, and sheds off domestic labor. Most FDI activities abroad either stimulate domestic activities, or are concomitant to a general expansion of a multinational firm. This result is also in line with the study of Moser et al. (2009), which also finds either positive employment effects from offshoring, or negative ones from offshoring if accompanied with partial plant closures.

The paper is organized as follows: the next section gives a framework for a plant-level analysis of offshoring including a comparison of different empirical measures linked to theoretical concepts. Section 3 outlines the empirical methodology; section 4 discusses briefly the IAB data set; section 5 provides the estimation of the propensity score of offshoring and reports various auxiliary tests; section 6 presents the main estimations of the average treatment effect on the treated of offshoring or FDI on employment. The last section concludes.

2 Employment Effects of Offshoring and FDI

To explore why employment effects differ across various studies on FDI/offshoring, we need to understand first how these studies differ in data, measurement, and methodology. We focus in this section on a comparison of measures of FDI/offshoring and ask what types of FDI or outsourcing are captured by each of those measures and what employment effects are expected from each type of FDI.

For example, Becker & Muendler (2008) use the measure expansion of employment in foreign affiliates, which may capture both an incremental increase in horizontal and vertical FDI. If foreign markets grow fast and FDI is of the horizontal type, then foreign affiliates increase and employment at home will not be affected if horizontal FDI is literally replicating the domestic production process abroad. If it is instead of the horizontal type according to Venables (1999), the first stage of the production process may take place at home, and a second one abroad, while the product is always sold abroad in equilibrium. An expansion abroad will then go along with a positive employment effect at home. If the investment is of the vertical type according to Venables (1999), the foreign affiliate produces intermediate inputs for assembly and sales at home. Expansion abroad will occur, because there is increased demand at home, increasing the demand for intermediate inputs from the foreign affiliates. Again, a positive employment effect at home is expected. A negative employment effect may arise, instead, if some production steps, undertaken previously at home, suddenly are shifted abroad. Even then, a firm's relocation of domestic production steps abroad may help to save costs, increase its competitiveness, and subsequently augment its world market share, which in turn may stimulate the activities related to production steps that remain at home.

A similar argumentation may result if the measure is a dummy variable for a domestic plant having a new foreign affiliate (Barba Navaretti & Castellani (2004), Buch & Lipponer (2010), Mattes (2010)). Again this may be horizontal or vertical in nature, yielding ambiguous effects on employment in dependence on which of the above mentioned cases are taking place. On top of the previous cases, some of the new investments may even be mergers & acquisitions which may be completely detached from the domestic production process and domestic employment effects are absent.

Such an FDI measure may be further specified by the motivation to undertake it. Mattes (2010) distinguishes FDI that is undertaken for the purpose to seek new markets and FDI that is seeking to reduce costs according to self-assessments of firms. While the market seeking motive is rather associated with horizontal FDI, cost reduction is typically associated with vertical FDI. Still, also horizontal FDI is driven by cost savings (Markusen (2002)). Again, both positive or negative employment effects may arise for both these specifications for reasons outlined above.

A fourth measure of FDI or offshoring is imported intermediate input demand (Biscourp & Kramarz (2007), Moser et al. (2009)). While this measure is excluding horizontal FDI, but focusses on vertical FDI and international outsourcing, instead, a domestic plant may substitute domestic suppliers for foreign suppliers, leaving employment in its own domestic plant possibly unaffected. Alternatively, cost savings through offshoring render the firm more competitive on world markets and stimulate domestic employment. Should the increase in intermediate inputs, instead, go along with a substitution of domestic production, then there may be an employment decline in the domestic plant.

A fifth measure is relocation of domestic production to a plant abroad Wagner (2009). This measure may again capture both horizontal or vertical FDI or international outsourcing. However, it excludes foreign expansions of operations, which are detached from domestic operations and excludes substitution of domestic for foreign suppliers, too. Still, the closure of a part of a plant may go along with a change in the specialization pattern, giving up some tasks, but expanding instead others. For example, certain low-skilled production activities may be shifted outside of the home country, while high-skilled intensive headquarter services are extended at home.

In summary, all types of FDI or offshoring have ambiguous employment effects in theory, and all measures available in existing data capture several types of FDI. If one wants to pin down employment effects unambiguously, then positive employment effects at home will arise, if a firm is expanding both at home and abroad. Instead, a negative employment effect at home is to be expected, if domestic production is substituted for production abroad, keeping the overall level of activity constant. While the literature, so far, has used mostly one of these measures at a time, and research designs have been different with respect to the data, the estimation method, and the control or selection variables, the previous results are hard to compare. For this reason, we investigate the above mentioned measures in a unified estimation design on the same data set to investigate systematically why studies differ in their empirical results so strongly.

3 Empirical Methodology

We follow Moser et al. (2009), and Wagner (2009) and combine a difference-in-differences estimator with a propensity score matching technique to investigate the relationship between FDI or offshoring and plant-level employment. The econometric problem is one of the missing counterfactual, i.e. what would have happened if plants had not undergone treatment offshoring/FDI. Matching techniques address this problem by *statistically designing a counterfactual*. To do this in the simplest possible way, a non-treated observation is assigned to each treated one that had ex ante the same probability of obtaining treatment than its treated twin. Treatment is then purely random conditional on the selection variables x, which determine the probability of treatment, $P(D = 1 \mid x)$, where D is a binary variable with value 1 if an observation obtained treatment.

The coefficient of interest is therefore the average treatment effect on the treated (ATT) of plant employment on the treatment FDI/offshoring. The ATT measures the average difference between the outcome of the treated observations and the hypothetical outcome without treatment.

To apply matching methods, three core-assumptions of matching must be fulfilled:

1. *Conditional-Mean-Independence assumption* (CMIA):

$$E[y_1 \mid D = 0, x] = E[y_1 \mid D = 1, x] = E[y_1 \mid x],$$

$$E[y_0 \mid D = 0, x] = E[y_0 \mid D = 1, x] = E[y_0 \mid x],$$

   where y_1 is the employment outcome of an average plant under treatment and y_0 the outcome if the same plant does not experience treatment. This assumption ensures that the assignment to the treatment group is random conditional on observable characteristics, i.e. self-selection into treatment is allowed, but only conditional on observable characteristics of the observation. This implies that the mean of observations outcomes with the same observable characteristics without treatment would be the same.[3]

2. *Overlap Assumption*:
$$0 < P(D = 1 \mid x) < 1$$
This assumption ensures that observations with probability zero or one are excluded from the matching process, because their assignment is not random by definition.

3. *Stable Unit Treatment Value assumption* (SUTVA):

   SUTVA means there exist no interdependencies between the two matching groups. Under this assumption the treatment only affects the treated observation itself. Thus, the effects on the treated have no impacts on the non-treated observations; general equilibrium effects are not allowed. (Rosenbaum & Rubin (1983)).

The combination of the matching estimator with the difference-in-difference approach somewhat relaxes a part of the CMIA. The measurement of the outcome variables in differences allows to eliminate time invariant, constant time trends based on unobservables like a first difference estimator or a fixed effect model. Indeed, a varying different time trend between treated and non-treated observations remains and is excluded by assumption.

Observations that are off the overlapping support region are not a problem in our analysis. No overlapping observations would pose a problem if many observations would be lost by controlling for this assumption. However, in our specifications we exclude the observations of the treatment group that have a lower propensity score than the lowest of the non-treatment group or non-treated observations that have a higher propensity score as the highest of the treatment observations vice versa. The results are constrained to this sample.

The third assumption could be a problem for the offshoring framework which is outlined in the previous section. Program evaluation methods are typical used to investigate the effects of *small* treatments that have no general equilibrium effects.[4] Consider a job-training program only for a small number of unemployed people that does not change the

[3] This is the motivation for the Pre-Test which is described and performed subsequently.

[4] An introduction to matching methods is given in Imbens & Wooldridge (2009). For a useful textbook section see Cameron & Trivedi (2005). Angrist & Pischke (2009) follow a new approach to teach these methods and compare them to standard econometrics expediently. A general implementation guide is Caliendo & Kopeinig (2008) and specific problems are discussed for instance in Abadie (2005), Abadie & Imbens (2006), Angrist & Hahn (2004), Dehejia & Wahba (2002), Dehejia (2005), Heckman, Ichimura, Smith & Todd (1998), Heckman, Ichimura & Todd (1998) or Smith & Todd (2005*b*). Holland (1986) discusses general causal inference based on the potential outcome model and Rosenbaum & Rubin (1983) concentrate on the propensity score.

overall skill of all unemployed people and thus not changing the labor demand at all. We argue somehow with the increasing competitiveness of the offshoring firms as the main driver for the employment gains. But increasing competitiveness due to higher domestic (and foreign) market shares hurts the non-general-equilibrium assumption heavily. Market switching effects are the core arguments in the offshoring framework in micro-analysis. Also the substitution of a domestic supplier can have a general effect on the domestic labor market and bias our estimations.

Wagner (2009) argues that the amount of offhoring units in his sample is too small to have such general impacts. from a technical point of view this is somehow imprecise. Moser et al. (2009) cope with this problem by modifying their econometric model and do not exclude these effects by assumption. We follow them in this analysis. By conditioning on a new dimension we can allow for a special case of a general equilibrium. Supposing that the observations belong to the same competitive price-market only the aggregate share of firms that deciding in the period before treatment to offshore is reasonable for the equilibrium employment. Hereafter, this is the vector M_0 (see Moser et al. (2009)). Note that then the ATT cannot be interpreted as usual. Here we must explicitly allow an impact on the non-treated through the treatment to handle the SUTVA. The reported ATTs in this study must be interpreted as a relative effect.

The general data generating process for the outcome y_{it} of a plant i at time $t = \{0, 1\}$, where 0 denotes the period before and 1 denotes the period after each case of offshoring, is described subsequently. The main outcome variable is the total employment at plant-level:

$$y_{it}^{T} = g(x_{i0})t + f^{T}(x_{i0}, M_0)t + \delta_{it}^{T}(M_0)t + \gamma_i + U_{it}t + \varepsilon_{it} \tag{1}$$

$$y_{it}^{NT} = g(x_{i0})t + f^{NT}(x_{i0}, M_0)t + \delta_{it}^{NT}(M_0)t + \gamma_i + U_{it}t + \varepsilon_{it} \tag{2}$$

Equation (1) with y_{it}^{T} as outcome describes the data generating process for the offshoring plants and (2) with y_{it}^{NT} as outcome describes it for the non-treated plants. $g(x_{i0})t$ is the function of the growth trend depending on observables x_{i0} before treatment which is independent of the treatment. $f^{NT}(x_{i0}, M_0)$ captures the causal impact of offshoring also depending on the observable characteristics x_{i0} and on the introduced aggregate vector M_0; this is allowed to be heterogeneous. The unobservable heterogeneous causal impact of the treatment is $\delta_{it}^{NT}(M_0)$ which the plants also include in their decision and is dependent on the mass of offshoring firms M_0. γ_i are the time invariant attributes that affect the outcome, both observable and/or unobservable. $U_{it}t$ varies over time and is not observable, but also affects the outcome.

Assuming we could observe the same plant's outcome first in the offshoring situation and then in the non-offshoring situation. Then, $g(x_{i0})t$, γ_i and $U_{it}t$ cancel out and the difference conditional on mass of offshoring plants M_0 and on the observables x_{it} could be measured as:

$$f^T(x_{i0}, M_0) + \delta_{i1}^T(M_0) - f^{NT}(x_{i0}, M_0) - \delta_{i1}^{NT}(M_0).$$

But this difference is hypothetical. We cannot observe the counterfactual of a plant's outcome. Therefore we have to design a counterfactual outcome conditional on the observables and on M_0 for every plant and estimate the average difference in these outcomes over all observations. As mentioned above we concentrate on the ATT which can be formalized as:

$$E[y_{i1}^T - y_{i1}^{NT} \mid D_{i1} = 1, M_0] = E[f^T(x_{i0}, M_0) + \delta_{i1}^T(M_0) - f^{NT}(x_{i0}, M_0) - \delta_{i1}^{NT}(M_0) \mid D_{i1} = 1, M_0],$$

where D_{it} is an indicator variable with value of one for the treatment group in period one and zero if there is no offshoring event in the first period. Remember that this average causal effect is a relative measure in our case. $E[f^{NT}(x_{i0}, M_0) + \delta_{i1}^{NT}(M_0) \mid D_{i1} = 1, M_0]$ is the part we have to construct where the methods from program evaluation are used for.

The difference-in-difference estimator

$$\Delta y_{i1} = \beta_0 + \beta_1 x_{i0} + \beta_2 D_{i1} + \varepsilon_i$$

with D_{it} as treatment indicator needs four assumptions to estimate the ATT consistently: no heterogeneous treatment effects based on observables, x_{i0} are exogenous time trend determinants, a linear functional form for the time trend and the time trend on observables x_{i0} has a common average for treated and non-treated. The last assumption implies that for a consistent estimator there are no self-selection effects into offshoring for the plants on unobservables, $E[U_{i1} \mid D_{i1} = 1, x_{i0}, M_0] = 0$, and no heterogeneous causal effects on unobservables $E[\delta_{i1}^T(M_0) - \delta_{it}^{NT}(M_0) \mid D_{i1} = 1, x_{i0}, M_0] = 0$.[5]

In combination with the matching estimator the *difference-in-difference-matching* approach relaxes the first three assumptions as described above (see for this approach Heckman et al. (1997)). The ATT under the remaining assumption of conditional mean

[5] We use this estimator as a robustness check for our results by programming a twofold differentiated standard OLS estimator. For an useful introduction to difference-in-difference technique we can refer to Angrist & Pischke (2009)

independence is given as

$$E[\Delta y_{i1} \mid x_{i0}, D_{i1} = 0, M_0] = E[\Delta y_{i1} \mid x_{i0}, D_{i1} = 1, M_0] = E[\Delta y_{i1} \mid x_{i0}, M_0]$$

and the ATT in the population is

$$ATT = E[\delta_x \mid D_{i1} = 1, M_0]$$

with

$$\delta_x \equiv E[\Delta y_{i1} \mid x_{i0}, D_{i1} = 1, M_0] - E[\Delta y_{i1} \mid x_{i0}, D_{i1} = 0, M_0].$$

Obviously, if we try to match the observations by x_{i0} or if we try to condition on x_{i0} respectively, there is a problem of dimensionality. Consider the case of some continuous variables or a large set of categorical variables or any combination of these two as determinants of the treatment. Hence, exact matching is not useful or practicable; we prefer to match on the propensity score. The propensity score is the conditional probability of getting treated of a plant i, $P(D_{i1} = 1) = P(x_{i0}) \equiv p_i$. Rosenbaum & Rubin (1983) can show that conditioning on the propensity score p_i instead of conditioning on x_{i0} is a valid approach. The central idea is that if the outcome is independent of the selection into treatment D_{i1} conditional on x_{i0}, the same is valid conditional on $P(x_{i0})$:

$$y_{it}^{T}, y_{it}^{NT} \perp D_{it} \mid x_{i0} \Rightarrow y_{it}^{T}, y_{it}^{NT} \perp D_{it} \mid P(x_{i0})$$

The propensity score has to be estimated. Typically, a binary model is used for that purpose. We choose a multinominal logit model to estimate the propensity score for the plants to offshore (McFadden (1974)).

The idea of the propensity score matching estimator is now to find for any treated observation another non-treated observation with the same estimated probability of treatment ($\hat{p}_i$) as for the treated one and compare their outcomes. But the propensity score is also a continuous variable and to find a matching partner with the same estimated $\hat{p}_i$ has zero probability in a random sample; we have to include similar observations instead of (non-existent) identical one to compare the outcomes. Several various matching estimators exist to tackle this problem. They vary in their idea of defining the "right" set of matching partners or control observations, their measurement of the distance between or in weighting issues. Note that every deviation from the identical propensity score matching makes the estimated coefficient potentially biased.

In this study two different, but intuitive matching strategies are employed: The kernel-estimator and the k-nearest neighbor approach.[6] They differ in the number of observations included and in their underlying non-parametric weighting function $g(.)$ of the included control observations. To formalize this we follow the difference-in-difference matching ATT formulation of Heckman et al. (1997):

$$\hat{\delta} = \sum_i D_{i1} \left[\Delta y_{i1} - \sum_j (((1 - D_{j1}) g(p_i, p_j) \Delta y_{j1})) \right],$$

where for this estimated $A\hat{T}T$ the expected value is replaced by sample mean due to conditions like finite higher-order moments, independent draws from a population and the law of large numbers. The weighting function for the kernel-estimator can then be formalized as:

$$g(p_i, p_j) = \frac{K((p_j - p_i)/h)}{\sum_{j \in A(i)} K((p_j - p_i)/h)}.$$

$A(i) = (j \mid\mid p_i - p_j \mid < h)$ is the set of control group observations and $K(.)$ is the Epanechnikov Kernel function which defines the weight in particular.[7] h is a parameter that defines the bandwidth around the treated observation where the potential control observations are located. The bandwidth allows to vary the number of control observations that are included for calculating the $A\hat{T}T$ and the Epanechnikov Kernel function allows to weigh the more distant observation less in the calculation. Heckman, Ichimura & Todd (1998) have shown that this approach generates consistent estimates of the ATT under common assumptions.

The second estimator in this analysis is the k-nearest neighbor estimator. We use it for some variation and robustness checks and to employ the necessary *balancing tests*. It substitutes the function $g(.)$ of $\hat{\delta}$ with:

$$g(p_i, p_j) = \begin{cases} 1, & \text{if } j = \arg\min \mid p_i - p_j \mid \\ 0, & \text{else} \end{cases}$$

This function uses the k-nearest non-treated neighbor observations of the treatment observation by the propensity score and weighs them with factor one. If there is only one neighbor the outcome of this one non-treated observation is compared to one treatment observation.

The choice of the bandwidth h for the Kernel approach or the number of neighbors for

[6] Caliendo & Kopeinig (2008) present other matching estimators.

[7] There are several other kernel functions available aside from the Epanechnikov function; I also use a gaussian kernel, but it did not matter for any coefficient or specification.

the k-nearest neighbors approach is a trade-off. On the one hand a bigger set of neighbors or a bigger parameter h for the bandwidth go along with a bias in the estimator; every match which is not a perfect match biases the estimator and if the bandwidth of this potential matching partners increases - i.e. h increases - also the bias increases potentially.[8] The same is intuitive for the k-nearest neighbor approach. The more neighbors are included the less is the quality of the matches by propensity score. Put differently: a distant neighbor is distant because its characteristics x_{i0} are different from the treatment observation.

But on the other hand every single observation added increases the efficiency of the estimator, because the variance decreases. This is reasonable for both approaches. We use the variation of the parameter h or k as a robustness check subsequently.

One remaining problem of matching is to size the standard errors to enable inference. A general approach to get such missing standard errors is the bootstrapping method. This seems to be useful for matching estimators, too (see Caliendo & Kopeinig (2008)). But Abadie & Imbens (2008) proof formally that bootstrapping is not valid for the nearest neighbor approach with replacement that we employ. On the other hand they suggest that bootstrapping is valid for the kernel matching estimator. Hence, for the kernel estimator we provide the bootstrapped standard errors and for the standard errors of the nearest neighbor results the analytical asymptotic standard errors of Abadie & Imbens (2006) is provided. The practical implementation is done in STATA version 10.1. The point results and standard errors for the kernel matching stem from the PSMATCH2 package Leuven & Sinaesi (2003) with 500 bootstrap iterations, where the propensity score estimation is repeated, too, in every iteration. The standard errors for the nearest neighbor approach stem from the NNMATCH package (Abadie et al. (2004)) which uses the calculation of Abadie & Imbens (2006) for valid standard errors. A practical guide to implement these matching estimators is given by Abadie et al. (2004) and by the help-file of the PSMATCH2 package.

As mentioned above one robustness check we perform is to vary the two different matching estimators by their parameters for the bandwidth h and different numbers of neighbors. As a second robustness check we use different logit specifications, stemming from (Wagner (2009)) and (Moser et al. (2009)).

The crucial assumption of the matching approach is the CMIA. Becker & Muendler (2008) describes this as that the selection into treatment must be exhaustively determined by

[8]Except for the case where there are no other observations within the bandwidth.

observables to get consistent ATTs. Regrettably, there is no formal test of this assumption. The only test of the CMIA is a pre-test, following Heckman & Hotz (1989), Imbens (2004), and Smith & Todd (2005*a*). But this test can only reject the CMIA if it is not fulfilled in the data or at most it can indicate some plausibility of the CMIA. The idea is to perform the matching estimator for the same observations but before the treatment period. If there would be a significant difference of the ATTs without treatment conditional on the same x_{i0} the CMIA is not valid. But if there is no difference a self-selection effect into treatment is less plausible.

The last methodological topic are the tests of the balancing property between the matching partners. The balancing of the covariates or the offshoring determinants is the core idea of the matching estimator. Balancing in the population is not a problem (see Rosenbaum & Rubin (1983)). But there are three possible reasons why the balancing is not fulfilled in the sample: First, the estimated propensity score is different from the real propensity due to a misspecification of the binary model; consider common problems of such an estimation like endogenous variables. Second, as mentioned above the matching is not an exact procedure. And third, even if the propensity score estimation is correct and the matching is exact - i.e. identical propensity scores for treated and matched-control can be found - the balancing property could be invalid due to an "unlucky" sample draw (see Rosenbaum & Rubin (1985)).

Hence, we have to test the balancing of the covariates between the two matching groups after the propensity score estimation. The literature offers a set of balancing tests. We decide to perform three typical balancing tests in our analysis: the standardized-difference test between the treatment group and the matched-control group according to Rosenbaum & Rubin (1985), a t-test of mean-difference between these groups and a t-squared Hotelling-Test by propensity score quantiles. The first two tests check the balancing of the covariates separately. The big advantage of the Hotelling-Test is that the selection variables of every matching group are tested jointly.

All balancing tests are provided for the simplest case of nearest neighbor matching with one neighbor, because there are no statistical problems stemming from the weighting function in this case. The matching partner is defined definitely just in this case.

4 Data

We employ data from the recent waves of the IAB Establishment Panel.[9] This is a stratified annual survey on behalf of the Institute for Employment Research (IAB) from 1993 onwards on West German establishments and from 1996 onwards on East German establishments.

The sample is drawn from a nationwide population consisting of about two million establishments. There is no size cut-off in the panel, thus, every establishment with at least one employee who is liable to the German social security system is included. Such are all sectors, subdivided into 17 industries. The stratification occurs along the dimensions of region (Bundesland), establishment size class in terms of employees, and industry, oversampling establishments of large size located in small regions, and belonging to industries with few establishments. But within the 170 cells of the stratification matrix, the sampling is random. Establishments that refuse to answer are replaced by randomly drawn establishments of the same strata. The number of establishments

The high quality of the data and high response rates are secured by attributes of the survey like professional face-to-face interviews (response rates up to 84%), elaborated questionnaire designs with pre-tests and a complex editing process after the field phase with comprehensive plausabilty and consistency checks.[10]

The questionnaire consists of several topic blocks like employment, business policy, investments, wages and salaries, and so on. The main interest of the survey is to collect labor market information. Furthermore, it consists of regular and irregular questions. The former are asked every year. The latter are dependent on actual developments or policy interests and on experiences of previous questionnaires and therefore asked only once or a few times. Unfortunately, our treatment variables of offshoring and FDI belong to the irregular questions, constraining our analysis to various treatment periods in between the years 2004 and 2006 (see details below). To apply the matching estimator, we employ three types of variables, treatment, outcome, and selection variables, which are described next.

[9]For this general data description of the IAB Establishment Panel we refer to Fischer et al. (2008). For a condensation of this detailed *Methodenreport* see Fischer et al. (2009).

[10]Implausabilities in the data are cleared up for instance with individual telephone calls with the interviewee; highly erroneous or implausable questionnaires are excluded from the data.

4.1 Treatment variables

We use four offshoring measures: FDI, market-seeking FDI, cost-saving FDI, and relocation.

4.1.1 FDI

We call a treatment in the following "*FDI*" if an establishment has invested abroad in the two previous business years, i.e. for the normal case in legal years 2004 and/or 2005 according to the 2006 IAB Establishment Panel questionnaire. If an establishment answered to this question with "yes", it is defined to belong to the treatment group, if it answered with "no" it potentially belongs to the control group. Alltogether, 170 out of 5759 establishments engaged in new FDI during the years 2004 or 2005. This measure was used for Germany by Mattes (2010), albeit applying a different estimation technique (dynamic panel data analysis) and different control variables. This treatment was also used by Barba Navaretti & Castellani (2004) on Italian firm data, using different selection variables.

4.1.2 Market-seeking FDI

The FDI treatment is further refined by the main objective or motive according to which an establishment has made its decision on its most important foreign investment. The questionnaire of 2006 offers seven motives: penetrate new markets/protect market share, procurement options for intermediate inputs, lower costs, taxes and contributions, lower labor costs, fewer administrative regulations, option of public funding, and other motive.

If one motive for the most important new foreign investment was to penetrate new markets or to protect foreign market shares, then we call this market-seeking FDI, which may capture horizontal or export platform FDI. There are 126 such types of FDI during the years 2004 or 2005. This measure was previously used in Mattes (2010), applying different methodology and control variables.

4.1.3 Cost-saving FDI

Likewise, we call FDI cost-saving if one motive for the most important new foreign investment was to save labor cost. While labor cost savings is associated foremost with vertical FDI, it is also relevant for horizontal FDI (Markusen (2002), Braconier et al. (2005)). There are 67 such cases of FDI during the years 2004 or 2005. This measure was also applied in the empirical study of Mattes (2010), albeit methodology and control variables were different.

4.1.4 Relocation

The last measure indicates an offshoring event where a domestic in-house activity is displaced by a foreign one. This variable belongs to the questionaire of the year 2007. The establishments were asked whether they closed down a part of their activities in the period from July 1, 2006 to June 30, 2007, and reopened abroad either in way of a cross-border spin-off or spun-off.[11] We consider a positive answer to one or both of these questions a "*relocation*" event. Establishments that did not close down any division or closed down a division but displaced it only domestically belong potentially to the control group for this treatment variable. There were altogether 48 relocations abroad among 7347 establishments.

4.2 Outcome variable

Our aim is to estimate the offshoring effect on a German plant's total employment. To capture the impact of treatment, difference-in-differences estimation compares employment before and after treatment of treated plants with appropriately chosen plants not treated during the same time period. To allow for some adjustment period, we take the difference in log employment before treatment with log employment up to one year after the treatment period. The treatment period of the FDI variables covers the years 2004 and 2005. Total plant-level employment is counted at June 30 in each year. Hence, we take the difference in log employment from June 30, 2003 to June 30, 2006, if treatment is one of the FDI variables. Likewise, relocation abroad occurs during July 1, 2006 and

[11] The 2008th survey of the German Federal Statistical Office (Statistisches-Bundesamt (2008)) supports these two cases to be typical offshoring events. The biggest part of German cross national displacements are represented by foundations of new establishments within the network of the firm (spun-off: 50,6%), or by displacing the domestic activity to an organisationally aligned firm that already exists (spin-off: 38%).

June 30, 2007. Hence, we take the difference in log employment during the period June 30, 2006, and June 30, 2008.

Total employment is the most reliable variable of the IAB Establishment Panel. First, it stems from the social security register, the reporting of which is obligatory by law. Second, it is checked, before the current interview in the establishment starts. Before the interview of the next year starts, the last year's employment is checked again; at last it is equalized during the editing process.

4.3 Selection variables

The last type of variables are the covariates that are necessary to estimate the propensity score for every plant. Selection variables are thus the decision criteria according to which management may have decided upon FDI or relocation. Hence, we include the lags of the time varying variables in the propensity score estimation and only the time invariant or persistent selection variables are included with their value contemporaneous to treatment, in order to loose as few observations as possible. Concerning the FDI treatments, selection variables date back to the period of the year 2003 or of June 30, 2004 - still before treatment starts on July 1, 2004. Concerning the relocation treatment, selection variables date from the period of the year 2005 or June 30, 2006 - still before treatment starts on July 1, 2006.

To take into account sample stratification, we always include the stratification variables among the selection variables, i.e. 16 regional dummies and 17 industry dummies, and firm size. Moreover, we take into account the knowledge from descriptive statistics Statistisches-Bundesamt (2008) that relocation activity varies by industries, federal state, and firm size.

To make the results of previous studies comparable, we investigate whether different choices of further selection variables matter for the outcome. In particular, we choose the selection variables previously used in the studies of Moser et al. (2009), selection variables MUW, henceforth, and of Wagner (2009), selection variables Wagner, henceforth. Moreover, we use an additional specification that explains the probability of the treatment relocation abroad better than the previous two specifications, selection variables S, henceforth.

Selection variables MUW

For the FDI variables we use the same selection variables that Moser et al. (2009) do. Their logit-estimates show that offshoring is the more likely according to their measure the larger is an establishment in terms of full-time employees, the more advanced its technology, the higher average wage costs, and the larger the share of high-skilled workers. Moreover, foreign-owned plants have also a larger probability of offshoring.[12] These selection variables are measured as follows:

- *log total employment*: logarithm of total employment at a plant before treatment;
- *log wage per employee*: logarithm of total wage cost per employee at a plant during the year before treatment;
- *high technology*: dummy variable taking value of one if a plant is self-assessed by the interviewee to employ a technology, which is above average or state-of-the-art during the year before treatment;
- *high-skilled*: percentage share of high-skilled employees at a plant during the year before treatment;
- *foreign ownership*: dummy variable taking value of one if majority of the plant is held by a foreign investor;

Selection variables Wagner

Additionally we provide the same variables as Wagner (2009) for all offshoring measures as a robustness check for our results. These are the number of employees, its square, and its cubic term, sales per employee, average wage costs, export share, and the change in employment during the last year before treatment. These selection variables are measured as follows:

[12]Similar sets of selection variables are applied, for instance, by Becker & Muendler (2008) or Barba Navaretti & Castellani (2004). According to the first study the firms that displace their activities internationally, stem from the high technology (manufacturing) sectors and are larger in terms of employment. Additionally, Barba Navaretti & Castellani (2004) find the size of a firm and its productivity and profitability to be adequate covariates for the treatment of international investments. An increase of all these variables increases the probability to be one of the firms that offshore. Becker & Muendler (2008) also identify the domestic employment and the plant's average wage costs per employee to be significant selection variables for their measure of foreign employment expansion. Furthermore, they employ variables that describe the skill composition of the workforce at the plant.

- *total employment*: total employment at a plant previous to treatment;
- *total employment squared*: total employment squared;
- *total employment cubic*: cubic term in total employment;
- *sales per employee*: sales per employee at a plant before treatment;
- *wage per employee*: wage per employee at plant i at time t-1;
- *export share*: share of total exports of total sales at plant i at time t-1 (percentage);
- *employment change*: change of total employment at plant i from time t-2 to t-1.

Selection variables MSU

This specification adjusts the estimation of the propensity score to fit it better to the relocation case. Apart from the industry and region dummies and the firm size variable in terms of number of employees, we include also the export share as in the Wagner (2009) specification and the technology variable as in the MUW specification. In addition, we include an indicator for an establishment that belongs to a corporate group, and an indicator for whether an establishment has a work council.

Affiliates of a corporate group may be more likely to be relocated, because these are often production units, intensive in production workers, which may be cheaper elsewhere. Instead, headquarters are intensive in high-skilled labor, which is fairly cheap in Germany. Moreover, single plant corporations are often too small to finance foreign investments, or lack the managerial experience of supervising affiliates.

Plants with more than five employees are eligible in Germany to have a works council if there exist employees who desire to have one. In fact, many, even large, firms do not have a works council. The decision to close an in-house activity and to dismiss employees is the prototype of a situation a works council takes part in. Because it is in the interest of the works council to secure domestic employment, and works councils can increase the cost of relocation (if not block it), its presence is likely to reduce the probability of relocation.

5 Propensity Score - Determinants of Offshoring

Following the propensity matching procedure the first estimates we depict stem from the binary model, which predicts the conditional probability for every establishment to be an offshoring plant. We split up the auxiliary estimates into two tables.

Table 1 presents the results from our logit specification for the different FDI measures. Column (1) presents the MUW selection variable specification for the FDI variable. The same specification is used in columns (3) and (4) for the cost-saving FDI treatment (column (3)) and for the market-seeking FDI variable (column 4). Column (2) provides the estimations from the Wagner specification as a robustness check.

As expected we find in our baseline the logarithm of the number of employees at a firm as measure for firm size with a positive sign and highly significant. The same holds for the logarithm of wage per employee, the high technology measure and the skill composition of the establishment. All these coefficients indicate the expected signs and are significant at the highest level. The foreign ownership dummy is significant as well, but shows a counterintuitive sign at first glance. We have expected a positive sign for foreign owned firms. To explain the negative sign we have to keep in mind that we observe single establishments instead of firms headquarters. If we observe an establishment that is foreign owned it is likely that this establishment is part of a multinational. Hence, it might be just a subsidiary. If we look at a foreign direct investment decision, as we do here, it is fair to say that this decision is undertaken more likely by the (foreign) headquarter. Hence, it might not be surprising that we find a negative sign.

If we compare the coefficients of the covariates of FDI in general to our cost-saving FDI or market-seeking FDI measure we find no major differences. Just the wage covariate looses its power of explanation for the cost-saving FDI treatment. We suspect here a problem of multicollinearity with the skill-structure covariate.

If we look at the covariates Wagner (2009) uses, we find no counterintuitive results. Moreover we find the same signs for every covariate as Wagner (2009) does and mostly no differences in the significance level to his trimmed baseline specification. Additionally, we find no important differences in the explanatory power across all specifications presented in 1.

Table 2 presents the impacts of covariates on the relocation decision. We provide three specifications. First, column (1) shows the coefficients of our baseline selection variable motivated above. Column (2) and (3) serve as robustness checks as before; therefore we use the MUW and the Wagner (2009) selection variables.

The baseline shows the expected positive and significant signs for size, export share and the affiliate dummy. Works councils are found to have a significant negative impact on the probability to offshore, too. Contrary to the FDI cases, establishments that relocate abroad self-assess to be further away from their technology frontier than plants that do not relocate. In specifications (2) and (3), only the size and the export-share variables remain significant with the expected signs.

Tables 5 and 6 (Appendix) provide the balancing tests for the general FDI indicator between the treatment and matched-control observations. Tables 8 and 9 (Appendix) do so for the relocation variable. As mentioned above the tests are employed for the nearest neighbor estimator with one neighbor. Unfortunately there is no analytical measure for the standardized difference test but a percent bias below 20 is mentioned by Rosenbaum & Rubin (1985) to be sufficient to state the covariates to be balanced.

None of the remaining percent biases after the matching process reaches this critical value. Also the mean difference t-test in column five does not reject the null hypothesis. Also the mean difference t-test in column five does not reject the null hypothesis. All p-values are far away from indicating an unbalanced variable. The last balancing test of Hotelling is performed over three quantiles and the hypotheses of an unbalanced composition in treatment and matched-control group is clearly rejected.

The only (indirect) way to check the CMIA assumption is the Pre-Test of Heckman & Hotz (1989), according to which a significant difference in outcomes of the treated and control observations before the treatment indicates a possible self-selection effect on unobservables, rejecting the CMIA assumption. Tables 7 and 10 (Appendix) provide this test for our treatment variables FDI and relocation. The first column compares only the baseline estimates of the matching procedure with a standard difference-in-differences approach which employs the OLS estimator on a differentiated estimation equation.[13] According to the idea of this test, all outcomes stem from the last and the second last period before treatment, respectively. None of the ATTs show a significant difference before treatment for the same matching partners as in the "main" matching period with treatment. Hence, we do not find an indication of violation of the CMIA assumption.

[13]This standard approach is reported in all ATT output tables in the following.

6 Results

We present our results of the average treatment effect on the treated of FDI and relocation abroad on employment divided into two tables, table 3 covering specifications FDI treatment variables, fitting to the specifications columns before. We present the ATTs for different bandwidths of kernel matching and different number of neighbors for k-nearest-neighbor matching. Table 4 does so for the relocation variable, but reports for the baseline specification of covariates an additional column that covers one more outcome period after the treatment.

To sum up the results for our FDI measures, we can state a consistent positive relative treatment effect for the employment of an establishment that invests abroad. Additionally, we cannot state a significant difference in the point estimates between the different measures of FDI. Hence, employment effects of FDI in general, market-seeking FDI, or cost-saving FDI are mostly similar. Hence, our results are in line with the most findings of previous studies.

The picture looks quite different if we look on the results for the relocation measure, table 4. Here, all point estimates are negative and mostly significant at the common levels. These effects are qualitatively comparable to the relocation effects of Wagner (2009). That means, if we look on a different internationalization strategy, we find opposite effects on the employment.

Additionally, we do not find qualitatively different results for the OLS difference-in-difference technique. The point estimates somehow differ in size - what is expected through a self selection of establishments into internationalization - but not by their sign. Thus, the previous results resist.

Our results are in line with Moser et al. (2009), using also data on German establishments, but covering the period 1998 until 2004, as compared to the years 2004 until 2006 in this study, and using different treatment variables. They find positive employment effects from their treatment, increase in intermediate input purchases, but negative employment effects from their treatment intermediate input purchases simultaneous to partial plant closure.

Hence, relocation abroad is distinct from FDI expansion. In most cases, FDI expansion - independently of the type of FDI - creates jobs abroad and at home, or occurs in firms that expand both at home and abroad. Only in cases, when domestic production is substituted for foreign production while the firm stagnates, negative employment effects show up.

7 Conclusion

Empirical studies on employment effects of offshoring or FDI obtain opposing results. To understand why results differ so much, we have been investigating how different measures of offshoring or FDI impact on domestic employment in German establishments, using additionally different estimation techniques, and control or selection variables. While neither estimation techniques, nor the choice of variables is decisive for opposing employment effects, positive employment effects arise from FDI, market-seeking FDI, and even cost-saving FDI. Instead, negative employment effects derive from relocation abroad. We explain this disparity of results by the different types of FDI that are captured with the various measures. In most cases, FDI expansion may occur in the vein of a general expansion of a multinational firm, creating jobs both at home or abroad. In other cases, expansion abroad may even stimulate activities at home. Yet, in other cases, foreign activities may substitute for domestic activities while the firm as a whole stagnates. Different measures of offshoring or FDI capture those cases in different proportions.

Tables

Table 1: Propensity Score Estimation logit - FDI

	FDI MUW (1)	FDI Wagner (2)	FDI cost (3)	FDI market (4)
log total employment (t-1)	0.724***		0.877***	0.785***
	(0.065)		(0.104)	(0.076)
log wage per employee (t-1)	0.682***		-0.093	0.708**
	(0.266)		(0.400)	(0.318)
high technology (t-1)	0.797***		0.903**	0.887***
	(0.253)		(0.402)	(0.305)
high-skilled (t-1)	1.918***		1.579**	2.383***
	(0.406)		(0.653)	(0.468)
foreign ownership	-1.268***		-1.713***	-1.374***
	(0.40)		(0.662)	(0.456)
total employment (t-1)		0.0007658***		
		(0.0001459)		
total employment squared (t-1)		7.42e-08**		
		(2.99e-08)		
total employment cubic (t-1)		1.73e-12		
		1.27e-12		
sales per employee (t-1)		-2.61e-08		
		(8.14e-08)		
wage per employee (t-1)		0.0002301***		
		(0.0000489)		
export share (t-1)		0.0149916***		
		(0.0017389)		
employment change (t-1)		-0.81088***		
		(0.298539)		
17 industry dummies	yes	yes	yes	yes
16 regional dummies	yes	yes	yes	yes
Pseudo R^2	0.3322	0.3261	0.3441	0.3451
Number of Obs.	5759	4972	3979	5109

+++ Notes: dependent variable: investment abroad in the business years 2004 and/or 2005; in (3) interacted with main investment motive *labor cost savings*; in (4) interacted with main investment motive *market seeking*; standard errors in parenthesis; *** 99% significance level, ** 95% significace level, * 90% significance level.

Table 2: Propensity Score Estimation logit - Relocation

	Relocation MSU (1)	Relocation MUW (2)	Relocation Wagner (3)
log total employment (t-1)	0.396*** (0.121)	0.228** (0.101)	
high technology (t-1)	-0.570* (0.330)	-0.419 (0.309)	
export share (t-1)	0.023*** (0.006)		0.009*** (0.003)
affiliate	0.782*** (0.365)		
works council	-1.049*** (0.460)		
log wage per employee (t-1)		-0.086 (0.335)	
high-skilled (t-1)		0.147 (0.592)	
foreign ownership		0.783 (0.415)	
total employment (t-1)			0.0003868*** (0.0001469)
total employment squared (t-1)			-3.19e-08* 1.81e-08
total employment cubic (t-1)			4.81e-13 3.98e-13
sales per employee (t-1)			-1.13e-07 (3.79e-07)
wage per employee (t-1)			0.0000269 (0.0000796)
employment change			-0.1746347 (0.3732809)
17 industry dummies	yes	yes	yes
16 regional dummies	yes	yes	yes
Pseudo R^2	0.1259	0.0819	0.1262
Number of Obs.	6496	7347	5271

+++ Notes: dependent variable: displacement of an in-house activity to a foreign country in period 01.07.2006 to 30.06.2007; standard errors in parenthesis; *** 99% significance level, ** 95% significace level, * 90% significance level.

Table 3: ATTs - FDI

	FDI MUW	FDI Wagner	FDI cost saving	FDI market seeking
	(1)	(2)	(3)	(4)
OLS DiD	0.047	0.033	0.038	0.056**
	(0.029)	(0.022)	(0.026)	(0.025)
kernel 0.01	0.087***	0.064*	0.063	0.097***
	(0.028)	(0.033)	(0.033)	(0.033)
	[0.032;0.147]	[-0.008;0.126]	[0.008;0.140]	[0.050;0.184]
kernel 0.03	0.083***	0.047	0.074**	0.095***
	(0.027)	(0.031)	(0.029)	(0.030)
	[0.035;0.134]	[-0.009;0.113]	[0.012;0.135]	[0.053;0.165]
kernel 0.05	0.083***	0.047	0.072***	0.096***
	(0.026)	(0.029)	(0.027)	(0.029)
	[0.032;0.131]	[-0.000;0.111]	[0.021;0.129]	[0.054;0.164]
NN 1	0.095***	0.087***	0.041	0.119**
	(0.035)	(0.042)	(0.041)	(0.047)
NN 2	0.081***	0.062*	0.025	0.129***
	(0.028)	(0.034)	(0.032)	(0.034)
NN 3	0.074***	0.065*	0.060**	0.121***
	(0.025)	(0.034)	0.030	(0.033)
treated Obs.	170	148	67	126

+++ Notes: treatment variable: investment abroad in the business years 2004 and/or 2005; in (3) interacted with main investment motive *labor cost savings*; in (4) interacted with main investment motive *market seeking*; standard errors in parenthesis; bootstrapped 95%-confidence interval in squared brackets; *** 99% significance level, ** 95% significace level, * 90% significance level; OLS DiD: Difference-in-Difference estimator with robust standard errors; matching estimator of Leuven & Sinaesi (2003); kernel-matching: epanechnikov kernel; standard errors are generated via bootstrapping with 500 replications; NN-matching: no caliper; standard errors stem from Abadie & Imbens (2008) via NNMATCH (Abadie et al. (2004)).

Table 4: ATTs - Relocation

	Relocation MSU	Relocation MSU t+1	Relocation MUW	Relocation Wagner
	(1)	(2)	(3)	(4)
OLS DiD	-0.148*	-0.177*	-0.326*	-0.043**
	(0.079)	(0.097)	(0.191)	(0.020)
kernel 0.01	-0.325*	-0.391**	-0.310*	-0.356
	(0.170)	(0.183)	(0.180)	(0.221)
	[-0.658;-0.043]	[-0.701;-0.023]	[-0.722;-0.051]	[-0.800;-0.015]
kernel 0.03	-0.328*	-0.393**	-0.310*	-0.346
	(0.177)	(0.194)	(0.179)	(0.225)
	[-0.660;-0.043]	[-0.748;-0.030]	[-0.723;-0.053]	[-0.859;-0.025]
kernel 0.05	-0.330*	-0.395**	-0.310*	-0.344
	(0.178)	(0.200)	(0.179)	(0.223)
	[-0.727;-0.044]	[-0.755;-0.019]	[-0.724;-0.053]	[-0.859;-0.025]
NN1	-0.365**	-0.412**	-0.287	-0.068
	(0.146)	(0.192)	(0.189)	(0.168)
NN2	-0.362***	-0.399**	-0.265*	-0.339
	(0.134)	(0.167)	(0.160)	(0.236)
NN3	-0.348	-0.375	-0.307*	-0.361
	(0.188)	(0.256)	(0.163)	(0.288)
treated Obs.	43	36	48	37

+++ Notes: treatment variable: displacement of an in-house activity to a foreign country in period 01.07.2006 to 30.06.2007; standard errors in parenthesis; bootstrapped 95%-confidence interval in squared brackets; *** 99% significance level, ** 95% significace level, * 90% significance level; OLS DiD: Difference-in-Difference estimator with robust standard errors; matching estimator of Leuven & Sinaesi (2003); kernel-matching: epanechnikov kernel; standard errors are generated via bootstrapping with 500 replications; NN-matching: no caliper; standard errors stem from Abadie & Imbens (2008) via NNMATCH (Abadie et al. (2004)).

References

Abadie, A. (2005), 'Semiparametric Difference-in-Differences Estimators', *Review of Economic Studies* **72**, 1–19.

Abadie, A., Drukker, D., Herr, J. L. & Imbens, G. W. (2004), 'Implementing matching estimators for average treatment effects in Stata', *Stata Journal* **4**(3), 290–311.

Abadie, A. & Imbens, G. W. (2006), 'Large sample properties of matching estimators for average treatment effects', *Econometrica* **74**(1), 235–267.

Abadie, A. & Imbens, G. W. (2008), 'On the Failure of the Bootstrap for Matching Estimators', *Econometrica* **76**(6), 1537–1557.

Angrist, D. J. & Pischke, J.-S. (2009), *Mostly Harmless Econometrics - An Empiricist's Companion*, 1 edn, Princeton University Press, Princeton.

Angrist, J. & Hahn, J. (2004), 'When to control for covariates? Panel asymptotics for estimates of treatment effects', *The Review of Economics and Statistics* **86**(1), 58–72.

Baldwin, R. (2006), *Globalisation: the great unbundling(s)*, Secretariat of the Economic Council, Finnish Prime Minister's Office, Helsinki, chapter 1.

Barba Navaretti, G. & Castellani, D. (2004), 'Investments Abroad and Performance at Home Evidence from Italian Multinationals', *CEPR Discussion Papers* **4283**.

Becker, S. O. & Muendler, M.-A. (2008), 'The Effect of FDI on Job Security', *The B.E. Journal of Economic Analysis & Policy* **8**(1).

Biscourp, P. & Kramarz, F. (2007), 'Employment, skill structure and international trade: Firm-level evidence for France', *Journal of International Economics* **72**(1), 22–51.

Braconier, H., Norbäck, P.-J. & Urban, D. M. (2005), 'Multinational enterprises and wage costs: vertical FDI revisited', *Journal of International Economics* **67**(2), 446–470.

Buch, C. M. & Lipponer, A. (2010), 'Volatile multinationals ? Evidence from the labor demand of German firms', *Labour Economics* **17**(2), 345–355.

Caliendo, M. & Kopeinig, S. (2008), 'Some Practical Guidance for the Implementation of Propensity Score Matching', *Journal of Economic Surveys* **22**, 31–72.

Cameron, A. C. & Trivedi, P. K. (2005), *Microeconometrics - Methods and Applications*, 1 edn, Cambridge University Press, Cambridge.

Crinó, R. (2009), 'Offshoring, Multinationals and Labour Market: A Review of the Empirical Literature', *Journal of Economic Surveys* **23**(2), 197–249.

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

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

Fischer, G., Janik, F., Müller, D. & Schmucker, A. (2008), 'The IAB Establishment Panel - from Sample to Survey to Prediction', *FDZ Methodenrteport* .

Fischer, G., Janik, F., Müller, D. & Schmucker, A. (2009), 'The IAB Establishment Panel - Things Users Should Know', *Schmollers Jahrbuch* **129**, 133–148.

Heckman, J. J. & Hotz, V. J. (1989), 'Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training', *Journal of the American Statistical Association* **84**(408), 862– 874.

Heckman, J. J., Ichimura, H., Smith, J. & Todd, P. E. (1998), 'Characterizing Selection Bias Using Experimental Data', *Econometrica* **66**(5), 1017–1098.

Heckman, J. J., Ichimura, H. & Todd, P. (1998), 'Matching As An Econometric Evaluation Estimator', *Review of Economic Studies* **65**(2), 261–294.

Heckman, J. J., Ichimura, H. & Todd, P. E. (1997), 'Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme', *The Review of Economic Studies* **64**(4), 605.

Holland, P. W. (1986), 'Statistics and Causal Inference', *Journal of the American Statistical Association* **81**(396), 945– 960.

Imbens, G. W. (2004), 'Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review', *The Review of Economics and Statistics* **86**(1), 4–29.

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

Leuven, E. & Sinaesi, B. (2003), '"PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing", Software', *http://ideas.repec.org/c/boc/bocode/s2432001.html* .

Mankiw, N. G. & Swagel, P. (2006), 'The politics and economics of offshore outsourcing', *Journal of Monetary Economics* **53**(5), 1027–1056.

Markusen, J. R. (2002), *Multinational firms and the theory of international trade*, MIT Press, Cambridge MA, London.

Mattes, A. (2010), 'The Impact of Horizontal and Vertical FDI on Labor Demand for Different Skill Groups', *IAW Discussion Papers* **59**.

McFadden, D. L. (1974), 'Conditional Logit Analysis of Qualitative Choice Behavior', *Frontiers in Econometrics* pp. 105–143.

Moser, C., Urban, D. M. & Weder Di Mauro, B. (2009), 'Offshoring, Firm Performance and Plant-level Employment: Identifying Productivity and Downsizing Effects', *CEPR Discussion Papers* **7455**.

Rosenbaum, P. R. & Rubin, D. B. (1983), 'The Central Role of the Propensity Score in Obsevational Studies for Causal Effects', *Biometrika* **70**(1), 41–55.

Rosenbaum, P. R. & Rubin, D. B. (1985), 'Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score', *The American Statistician* **39**(1), 33–38.

Smith, J. A. & Todd, P. E. (2005*a*), 'Does matching overcome LaLonde's critique of nonexperimental estimators?', *Journal of Econometrics* **125**(1-2), 305–353.

Smith, J. & Todd, P. E. (2005*b*), 'Rejoinder', *Journal of Econometrics* **125**, 365–375.

Statistisches-Bundesamt (2008), 'Verlagerung Wirtschaftlicher Aktivitäten - Ergebnisse der Piloterhebung'.

Venables, A. J. (1999), 'Fragmentation and multinational production', *European Economic Review* **43**(4-6), 935–945.

Wagner, J. (2009), 'Offshoring and Firm Performance: Self-Selection, Effects on Performance, or Both?', *IZA Discussion Papers* **4605**.

A Appendix

Table 5: Balancing Tests from Nearest-Neighbor-Matching - FDI

Covariate	Mean treatment group	Mean matched control group	Percent bias	Percent bias reduction	Mean difference test
log total employment	5.3857	5.4260	-2.5	98.3	-0.22(0.83)
log wage per employee	7.8654	7.8651	0.1	99.9	0.01(0.99)
high technology	0.8765	0.9000	-5.8	88.9	-0.69(0.49)
high-skilled	0.5034	0.5021	0.4	95.6	0.04(0.96)
foreign ownership	0.0529	0.0529	0.0	100.0	0.00(1.00)

$^{+++}$ Notes: p-values in parenthesis; matching method: NN-matching; number of neighbors: one; caliper: no; treatment variable: investment abroad in the business years 2004 and/or 2005.

Table 6: Hotelling's T-squared Test by Propensity Score 3-Quantile - FDI

Quantile	Frequency treatments	Frequency matched controls	T-squared statistics	F-Test statistics	p-value
First	52	48	38.825	0.7924	0.7654
Second	52	48	26.216	0.7511	0.7908
Third	66	33	21.143	0.7530	0.7700

$^{+++}$ Notes: Hotelling's T-squared Test for 3 Quantile for all covariates jointly; matching method: NN-matching; number of neighbors: one; no caliper; treatment variable: investment abroad in the business years 2004 and/or 2005.

Table 7: Heckman and Hotz Pre-Test - FDI

Time	OLS for FDI	ATT for FDI
t-1	0.029**	0.013
	(0.012)	(0.019)

+++ Notes: standard errors in parenthesis; *** 99% significance level, ** 95% significace level, * 90% significance level; OLS DiD: Difference-in-Difference estimator with robust standard errors; matching method: kernel matching; weighting: epanechnikov; bandwidth: 0.01; standard errors are generated via bootstrapping with 500 replications; treatment variable: investment abroad in the business years 2004 and/or 2005.

Table 8: Balancing Tests from Nearest-Neighbor-Matching - Relocation

Covariate	Mean treatment group	Mean matched control group	Percent bias	Percent bias reduction	Mean difference test
log total employment	4.4883	4.4352	2.7	96.0	0.11(0.92)
exports	30.721	32.349	6.1	92.6	-0.22(0.82)
affiliate	0.3256	0.3721	-11.0	73.0	-0.45(0.66)
works council	0.4419	0.4651	-4.9	86.8	-0.21(0.83)
high technology	0.6047	0.6744	-14.5	8.0	-0.67(0.51)

+++ Notes: p-values in parenthesis; matching method: NN-matching; number of neighbors: one; caliper: no; treatment variable: displacement of an in-house activity to a foreign country in period 01.07.2006 to 30.06.2007.

Table 9: Hotelling's T-squared Test by Propensity Score 3-Quantile - Relocation

Quantile	Frequency treatments	Frequency matched controls	T-squared statistics	F-Test statistics	p-value
First	12	23	26.368	0.7990	0.6756
Second	15	15	60.285	0.4485	0.9157
Third	16	16	66.911	0.6505	0.7975

+++ Notes: Hotelling's T-squared Test for 3 quantiles for all covariates jointly; matching method: NN-matching; number of neighbors: one; no caliper; treatment variable: displacement of an in-house activity to a foreign country in period 01.07.2006 to 30.06.2007.

Table 10: Heckman and Hotz Pre-Test - Relocation

Time	OLS relocation	ATT Relocation
t-1	-0.042 (0.027)	-0.038 (0.058)

+++ Notes: standard errors in parenthesis; *** 99% significance level, ** 95% significace level, * 90% significance level; OLS DiD: Difference-in-Difference estimator with robust standard errors; matching method: kernel matching; weighting: epanechnikov; bandwidth: 0.01; standard errors are generated via bootstrapping with 500 replications; treatment variable: displacement of an in-house activity to a foreign country in period 01.07.2006 to 30.06.2007.