RUHR
ECONOMIC PAPERS

Timo Mitze
Alfredo R. Paloyo
Björn Alecke

Quasi-experimental Methods in Empirical Regional Science and Policy Analysis – Is there a Scope for Application?

rwi

#367

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 #367

Responsible Editor: Thomas K. Bauer

ISSN 1864-4872 (online) – ISBN 978-3-86788-422-8

Ruhr Economic Papers #367

Timo Mitze, Alfredo R. Paloyo, and Björn Alecke

Quasi-experimental Methods in Empirical Regional Science and Policy Analysis – Is there a Scope for Application?

Bibliografische Informationen der Deutschen Nationalbibliothek

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

http://dx.doi.org/10.4419/86788422
ISSN 1864-4872 (online)
ISBN 978-3-86788-422-8

Timo Mitze, Alfredo R. Paloyo, and Björn Alecke[1]

Quasi-experimental Methods in Empirical Regional Science and Policy Analysis – Is there a Scope for Application?

Abstract

Applied econometrics has recently emphasized the identification of causal parameters for policy analysis. This revolution has yet to fully propagate to the field of regional science. We examine the scope for application of the matching approach – part of the modern applied econometrics toolkit – in regional science and highlight special features of regional data that make such an application difficult. In particular, our analysis of the effect of regional subsidies on labor-productivity growth in Germany indicates that such policies are effective, but only up to a certain maximum treatment intensity. Although the matching approach is very appealing due to its methodological rigor and didactical clarity, we faced difficulties in balancing the set of covariates for our regional data given that the regions differ strongly with respect to the underlying structural characteristics. Thus, results have to be interpreted with some caution. The matching approach nevertheless can be of great value for regional policy analysis and should be the subject of future research efforts in the field of empirical regional science.

JEL Classification: C21, R11, R58

Keywords: Generalized propensity score; nearest neighbor matching; labor productivity growth; regional policy

September 2012

---

1 Timo Mitze, University of Southern Denmark; Alfredo R. Paloyo, RWI; Björn Alecke, Gesellschaft für Finanz- und Regionalanalysen (GEFRA). *– Earlier versions of this paper have been presented at the 50th ERSA Congress in Jönköping and the 51st ERSA Congress in Barcelona. The authors thank conference participants for helpful suggestions.* *– All correspondence to Timo Mitze, Department of Border Region Studies, University of Southern Denmark, Alsion 2, 6400 Sønderborg, Denmark, E-Mail: tmitze@sam.sdu.dk.*

1 INTRODUCTION

Recent advances in applied econometrics have revolutionized the way economists and allied social scientists have addressed the issue of causality when confronted with observational data (i.e., data that do not come from a randomized controlled trial). Angrist and Pischke (2010) call this the "credibility revolution" in empirical economics.[1] While much of the previous work concerned with econometric evaluation involved simultaneous-equations models derived from an explicit structural framework, current applied work is dominated by the "experimentalist" school of thought, where emphasis is on credibly estimating a particular causal parameter of interest, sometimes even in the absence of an explicit structural model.[2] When quasi-experimental evaluation methods[3] can be properly applied, such an approach has the potential to give us reliable estimates of "treatment effects" (Wooldridge 2010) without an appeal to the sometimes very strong assumptions that underlie much of structural work.[4]

Given that quasi-experimental evaluation methods are increasingly applied in regional sciences, we give a brief overview in this paper of recent innovations in the field and apply one of the tools of the experimentalist school to study the effects of regional policy measures on economic growth. More explicitly, we estimate the causal effect of private-sector investment subsidies and business-related infrastructure projects in Germany under the umbrella of the *Gemeinschaftsaufgabe "Verbesserung der regionalen Wirtschaftsstruktur"* (GRW) on labor productivity growth. The GRW is arguably the most powerful instrument of German regional policy. Since the reunification of West and East Germany, more than € 60 billion has been spent to foster regional growth in structurally weak regions. The GRW also includes EU regional policy grants by means of the European Regional Development Fund (ERDF).

In the present paper, the estimation of the general effectiveness of German regional policy is done in two complementary approaches. First, we use a binary treatment indicator and apply propensity-score-based matching to compare the growth performance of GRW-funded and non-funded regions. This empirical identification strategy allows us to answer the question whether treatment boosts growth at all. Second, we calculate—for the subgroup of funded regions—the limit at which regional support is able to induce

[1] This revolution is not really a great leap forward in terms of methodology (e.g., the instrumental-variables estimator, developed decades earlier by P.G. Wright (Stock and Trebbi 2003), plays a prominent role) inasmuch as it is a radical change in the focus of analysis. There is much in the experimental literature about credible "identification strategies" in the sense of isolating an exogenous variation in the data to recover unbiased or consistent estimates of the parameter of interest. For a recent survey, see Imbens and Wooldridge (2009).

[2] While this approach has been welcome in human resource economics and other "micro" fields, it has made much less progress in areas such as industrial organization and macroeconomics. We discuss the reasons for this below.

[3] The World Bank has a taxonomy of evaluation designs available in the following link: http://go.worldbank.org/7M4NUSKE10.

[4] Keane (2010) makes it clear, however, that the non- and quasi-experimental estimation methods are only seemingly without strong assumptions. Implicitly, these techniques rely on rather strong assumptions as well, some of which may not be entirely justified in the industrial organization or regional studies context.

higher growth using a generalized propensity-score approach (Hirano and Imbens 2004) for a continuous treatment indicator (i.e., the amount of subsidy). This latter identification strategy is able to give an estimate of up to what extent a higher treatment intensity is associated with a better growth performance.

Our result suggests that GRW-funded regions indeed experienced higher labor-productivity growth compared to non-funded regions, indicating that the GRW policy has the potential to balance standards of living in Germany. We find that up to a funding intensity which corresponds to the 67th percentile of the regional distribution of GRW payments—or roughly 105 thousand Euro (T€) of GRW payments per labor force—higher treatment intensities ensure higher productivity growth. Thereafter, more funding does not necessarily induce increased growth. Thus, in line with earlier work on EU regional funding by Becker et al. (2012), we provide empirical evidence for a maximum desired treatment level after which a further policy stimulus does not have any positive effect on the regional economic performance.

Although the experimentalist approach is very appealing from a methodological and didactic perspective, applications to regional data have to be interpreted with some caution. This is due to the fact that regional data is associated with special features that are likely to complicate empirical applications, particularly in terms of satisfying the so-called “balancing property”. In other words, for a fixed set of regional observations, it is very hard to find perfect statistical twins in order to make sensible comparisons of mean outcomes. Moreover, the crucial assumption of “no general-equilibrium effects” (i.e., the stable unit-treatment-value assumption or SUTVA) is difficult to justify in a regional setting. Bearing the caveats in mind, we conclude with some thoughts on how the research agenda should move forward in regional science.

2 METHODOLOGICAL APPROACHES IN REGIONAL SCIENCE

We take the current topology of the literature as given—that is, whether economists ought to be doing structural or experimental work is not within the scope of this article.[5] More modestly, our aim is to demonstrate that some of the non- and quasi-experimental methods primarily developed by labor economists have applicability in regional studies as well, but that the nature of regional data presents some difficulties that are not typically encountered in individual-level studies common in applied labor microeconometrics. Macroeconomics and industrial organization, both of which heavily influence the methods used in regional science, have not followed the experimentalist revolution in labor economics. By implication, regional studies has not fully benefited from the so-called credibility revolution, particularly its emphasis on identification of causal parameters of interest.

The question of whether more experimental studies should be done in regional science is something else. Holmes (2010) has delineated the existing approaches in re-

[5] For that, the reader is referred to the following articles and the references therein: Angrist and Pischke (2010), Heckman (2010), Heckman and Urzúa (2010), Deaton (2010), and Keane (2010).

gional science into three types: descriptive, experimentalist, and structural. As can be reasonably expected from a regional scientist, he seems to be more sympathetic to the structural approach, noting that analyses of this type have been successfully applied in industrial organization, which in turn has a leading influence on regional science. Nevertheless, he is not entirely dismissive of the experimentalist methodology. He correctly notes that this approach has encouraged researchers to think about causation more carefully. Descriptive studies which are prevalent in the regional studies literature now suffer from diminished credibility as a result of the emphasis of the experimentalist school of thought on the sanctity of the identification strategy.[6]

Though much of the applied microeconometric techniques now in use for causal inference were developed within the field of applied labor microeconomics, regional science has also developed unique methods of its own. In particular, there exists a tremendous emphasis on spatial econometrics (by necessity, of course—no serious regional scientist would discount the importance of space). This has spilled over (pun intended) into other fields which have thus far neglected the spatial dimension of the data.

For instance, labor economists working on minimum wages have now understood the importance of spatial dependence and heterogeneity. In the analysis of the effect of minimum wages on employment, Dube et al. (2010) show that accounting for spatial heterogeneity makes a difference in whether minimum wages are shown to have either a negative or a positive (indeed, possibly even zero) effect. This illustrates that regional science does not always lag behind macroeconomics, industrial organization, or applied microeconometrics; indeed, it is at the forefront of developing econometric techniques with wide applicability.

Certain regional scientists have, however, started to accommodate the insights of recent applied work or at least have been influenced by the experimentalist school's emphasis on credible estimation of the causal parameter of interest. Dell (2010) uses discontinuities in Peru's regions to examine the effect of historical institutions on economic development.[7] Similarly, Becker et al. (2010) apply a regression-discontinuity design to examine the effectiveness of EU regional policy. The authors make use of the institutional design of the EU Objective 1 subsidy scheme, which qualifies regions for structural funds payments if they have a per capita GDP level below 75 percent of the EU average. Using this threshold level, the authors exploit the discrete jump in the probability of EU transfer receipt for their empirical identification strategy and find that Objective 1 payments have a positive impact on GDP growth.

Similar to the method applied in the present article, Becker et al. (2012) and Mohl and Hagen (2008) examine the impact of EU subsidies to lagging regions by means of matching methods. The fundamental idea underlying the matching approach is to con-

[6] Descriptive studies are not entirely useless, of course. Such studies allow us to uncover stylized facts in the data, which could subsequently lead to theories that attempt to explain the observed distributions of the data. Naturally, these theories ought to be tested, and this is where both structural and experimental approaches come into play.
[7] The discontinuities were due to the *mita*, the forced mining labor system during the era of colonized Peru.

struct a counterfactual situation which is able to answer the question, "What would have happened to the regional growth paths of two regions if everything else is equal in these regions except that one region did not receive funding?". The latter situation calls for a binary treatment indicator that splits regional entities into treated and non-treated (comparison) regions. In the case of a continuous treatment variable one could ask, "What would have happened to the regional growth paths if everything else is equal except that one region received a higher or lower level of funding compared to the other?". Thus, the single most important task in matching estimation, as implied by its name, is to find statistical twins which only differ by treatment status and no other structural characteristics that might impact on the observed economic growth performance.

3 A FRAMEWORK FOR EVALUATING REGIONAL SUBSIDIES

In this section, we briefly describe the underlying formal framework of much of the experimental or quasi-experimental techniques used for causal inference. This is based on the so-called potential-outcome model developed in the statistics literature by Neyman as early as 1923.[8]

Suppose we have a sample of N individual observations (say, regions) denoted by i and there are only two time periods (pre- and post-treatment). Our response variable is y_i (say, labor-productivity growth) and a treatment indicator, d_i, equals 1 if i received the treatment (i.e., a regional subsidy) and 0 otherwise. Before treatment is administered, two potential outcomes exist, $y_i(0)$ and $y_i(1)$, which represent outcomes if i did not or did receive the treatment, respectively.

After the receipt of the subsidy, we only observe $y_i = y_i(d_i) = y_i(0)(1 - d_i) + y_i(1)(d_i)$. For a particular region, we cannot observe the gross grain or loss, $y_i(1) - y_i(0)$, because both outcomes cannot be observed simultaneously. This represents the "fundamental problem of causal inference" (Holland 1986). In terms of average effects, this implies the following:

$$\tau_{\text{ATE}} = \mathrm{E}[y_i(1) - y_i(0)]$$

$$\tau_{\text{ATET}} = \mathrm{E}[y_i(1) - y_i(0) | d_i = 1] \,,$$

where ATE and ATET refer to the average treatment effect and average treatment effect on the treated, respectively. The goal is to recover these average effects by estimating the counterfactual situation.

One has to note that the individual observations are characteristically different from each other in important dimensions that affect both the probability of receiving a

[8] In economic applications, variants of the potential-outcome model can be traced to Roy (1951) and Quandt (1972). Heckman (2008) has a comprehensive discussion. Interestingly, he notes that the econometric concept of causality is much broader than that espoused in the statistics literature. In the literature, this model is also referred to as the "Rubin causal model", named after Donald B. Rubin, who formalized the current framework in a series of papers (e.g., Rubin (1986)).

certain amount of treatment and the response variable. Without taking this into account, a simple comparison of mean outcomes at different treatment intensities (i.e., levels of funding) is unable to provide us with a consistent estimate of the treatment effect (i.e., the effect of a particular level of funding on economic growth) because of the selection-into-treatment bias:

$$\mathrm{E}[y_i|d_i = 1] - \mathrm{E}[y_i|d_i = 0] = \{\mathrm{E}[y_i(1)|d_i = 1] - \mathrm{E}[y_i(0)|d_i = 1]\} + \{\mathrm{E}[y_i(0)|d_i = 1] - \mathrm{E}[y_i(0)|d_i = 0]\}.$$

The equation above shows that the observed difference can be decomposed into two parts. One is the ATET since $\mathrm{E}[y_i(1)|d_i = 1] - \mathrm{E}[y_i(0)|d_i = 1]$ is equal to $\mathrm{E}[y_i(1) - y_i(0)|d_i = 1]$. This is equal to the observed difference only when the second component, $\mathrm{E}[y_i(0)|d_i = 1] - \mathrm{E}[y_i(0)|d_i = 0]$, which represents the selection bias, is equal to zero.

The selection bias might manifest itself in this particular case of regional-policy evaluation by virtue of the fact that underperforming regions are precisely the ones that are given the subsidy. In other words, the recipients of the treatment are characteristically different from the non-recipients, and these characteristics are most likely correlated with the response variable of interest. In this case, ordinary regression estimates that do not take these differences into account are likely biased and inconsistent, and, therefore, are of very little use for evaluating the effectiveness of the policy.

4 The matching approach

One way to address this evaluation problem is to employ a matching approach based on the generalized propensity score (GPS) to eliminate biases generated by the inherent differences between regions as captured by the covariates (Hirano and Imbens 2004). This approach is a generalized version of conventional propensity-score matching (Rosenbaum and Rubin 1983) in that matching on the GPS allows for the continuous (as opposed to binary) nature of the treatment variable.[9] In this paper, we address the problem of regional policy evaluation in two steps: first, we use a binary treatment indicator; second, we take into account the intensity of treatment.

Before we describe matching on the GPS, we begin with the simpler case of a binary treatment to illustrate the mechanics of matching methods. The basic idea underlying the matching approach is to obtain a statistical twin of a treated region but which comes from the untreated group. Under fairly mild assumptions, the mean of the differences in outcomes between treated and untreated regions represents an estimate of the policy effect (Rosenbaum and Rubin 1983). For a specific matched pair, the outcome for the untreated labor-market region is therefore construed as the counterfactual situation for the

[9] We give a succinct overview of matching in this paper and refer the interested reader to Caliendo and Kopeinig (2008) for more details.

treated region—that is, it represents economic growth in a region which received funding had that particular region not, in fact, receive funding.[10]

What is essential for matching methods to generate consistent estimates is for the assumption of conditional independence to hold:

$$(y_i(1), y_i(0)) \perp d_i | \mathbf{x}_i,$$

where $\perp$ denotes statistical independence and $\mathbf{x}_i$ is a vector of covariates; i indexes units of observation. This implies that, although regions may differ in their observed characteristics $\mathbf{x}_i$, adjusting for these observable characteristics eliminates the biases associated with them. Thus, the conditional-independence assumption (CIA) is equivalently formulated as

$$\Pr(d_i = 1 | y_i(1), y_i(0), \mathbf{x}_i) = \Pr(d_i | \mathbf{x}_i),$$

where $\Pr(d_i = 1 | \mathbf{x}_i)$ is called the propensity score, i.e., the probability of treatment receipt. Notably, this equation precludes any "selection on unobservables", i.e., the covariates $\mathbf{x}_i$ capture all factors that determine the treatment probability.

Rosenbaum and Rubin (1983) show that when the CIA holds, then it is also true that $(y_i(1), y_i(0)) \perp d_i | \Pr(d_i = 1 | \mathbf{x}_i)$.[11] This important result states that if conditioning on the covariates $\mathbf{x}_i$ eliminates the selection bias, then conditioning on the propensity score can just as well do the same.

The level of potential subsidy is a continuous variable: $\gamma \in [\gamma_0, \gamma_1]$. In this case, it is more appropriate to use the GPS so we can calculate the dose–response function, i.e., the treatment effect at every level of the subsidy. Following Hirano and Imbens (2004) and Bia and Mattei (2008), define the potential outcomes as $y_i(\gamma)$ and maintain that CIA holds: $y_i(\gamma) \perp g_i | \mathbf{x}_i$, where g_i is the actual subsidy received. We assume that this condition—called "weak unconfoundedness"—holds for all $\gamma \in G = [\gamma_0, \gamma_1]$. The fundamental problem of causal inference remains since we are only able to observe the triple $\{\mathbf{x}_i, g_i, y_i(\gamma = g_i)\}$ once the subsidy is supplied; the entire set of potential outcomes $y_i(\gamma)$ remains unknown.

The GPS is defined as $s_i \equiv \omega(g_i, \mathbf{x}_i)$, where $\omega(\gamma, \mathbf{x})$ is the conditional density of the treatment variable given the covariates. An important result derived by Hirano and Imbens (2004) is that weak unconfoundedness given the GPS is equivalent to weak unconfoundedness given the covariates:

$$y_i(\gamma) \perp g_i | s_i \; \forall \; g_i \in \gamma.$$

[10] Standard practice based on more recent developments in matching techniques involves taking weighted averages of the untreated observations to build matches and to calculate the program effect (Heckman, Ichimura, and Todd 1997, 1998; Dehejia and Wahba 1999).

[11] Actually, for the matching approach to work, one must also assume that units with the same $\mathbf{x}_i$ have an equal probability of being a recipient of the treatment and being part of the control group. In other words, there must be some overlap between the treatment and control groups. This "common support" assumption, together with the CIA, is the case of "strong ignorability" (Rosenbaum and Rubin 1983).

Therefore, the GPS has the same bias-elimination property in the continuous treatment case as that demonstrated by the propensity score in the case of binary treatments.

To assess the quality of the matching procedure, researchers typically test whether the treatment and comparison groups are balanced. Caliendo and Kopeinig (2008) list a few methods to evaluate covariate balance: the use of the standardized bias, a *t*-test, a test using the pseudo-R^2, a test for joint significance (Sianesi 2004), and a stratification test based on Dehejia and Wahba (1999, 2002). The basic idea behind these approaches is to check whether systematic differences between treatment and control groups remain even after conditioning on the propensity score.

Analyzing the sensitivity of the estimation results is another important feature in applied work (Caliendo and Kopeinig 2008). A focal point here is to test for the potential role of hidden biases stemming from unobserved variables that influence the probability of receiving treatment. A prominent test to quantify this source of bias is to calculate Rosenbaum (2002) bounds. As DiPrete and Gangl (2004) point out, the Rosenbaum bounding approach can be interpreted as a worst-case scenario to test for the stability of the estimated outcome differences between treated and non-treated individuals given the existence of unobserved influencing factors. Rosenbaum bounds then quantify what the necessary strength of an unmeasured influence has to be in order to significantly impact the estimated ATET operating through selection effects.

5 EMPIRICAL APPLICATION

The *Gemeinschaftsaufgabe Verbesserung der regionalen Wirtschaftsstruktur* (henceforth, GRW) is the most important regional policy instrument in Germany and operates as a coordinated policy between the German federal government, the state-level governments and the EU's European Regional Development Funds (ERDF). The goal of the GRW is to provide subsidies for investments of the private business sector in economically underdeveloped regions as well as the provision of business-related public infrastructure.[12] Since the German reunification, roughly € 61 billion has been spent to foster the equalization of living standards in the different regions of Germany, with a large part of the subsidy allocated to the East German recovery. About two-thirds of the overall funding volume was assigned to private-sector investment subsidies (€ 39 billion).

We use annual data for the period 1993–2008 allocated to the 413 NUTS-3 districts in Germany in order to assess the effectiveness of the GRW. Descriptive statistics of the variables used throughout the empirical exercise are given in the Appendix. The response variable is the growth rate in labor productivity defined as the annual growth rate of GDP per worker. In the first step, our binary treatment indicator takes the value of 1 if a region received GRW payments for at least one year in the period 1993–2008 and is 0 otherwise.

[12] For a detailed description of the institutional setup, see, for instance, Alecke et al. (2012).

To estimate the propensity score (i.e., the probability of receiving treatment) for each region, we use a probit specification that models the receipt of GRW as a function of the following control variables: (1) the initial income gap in 1992 relative to the maximum income level observed in the sample period (as a proxy for steady-state income), (2) the average firm size defined as the number of workers per firm in each region, (3) the regional share of manufacturing-sector employment in total regional employment, (4) the region's human-capital endowment, (5) the population density defined as the population per area, as well as two dummy variables indicating (8) whether the region is an independent urban municipality (*kreisfreie Stadt*) with more than 100,000 inhabitants or belongs to a greater administrative district otherwise (*Landkreis*) and finally (9) an ordinal variable based on a classification of the regional settlement structure, which takes values from 1 (center of an agglomerated area) to 9 (rural area in periphery).[13]

The control variables were basically selected based on theoretical reasons and institutional facts. For instance, the inclusion of the initial labor-productivity gap in 1992 aims at capturing the institutional features of the GRW scheme, which assigns regions as eligible for funding if they are classified as "structurally weak" by means of a composite indicator using different socioeconomic criteria (including historical and projected information on unemployment rates, income levels, infrastructure equipment, etc.).[14] Though the GRW thus does not have a strict linear relationship with relative productivity levels as compared to the institutional setting of the EU structural funds, relative income gaps may be seen as a key indicator which is highly correlated with other socioeconomic criteria such as unemployment rates.

Likewise, the average firm size and the regional employment share of manufacturing sectors serve as empirical proxies for the underlying regional business structure, which are likely to influence the probability of receiving GRW funding as well. Finally, human-capital endowment, population density, and the included indicator variables mark further transmission channels that are theoretically expected to affect the receipt of GRW grants by regions. Thus, our approach does not aim at replicating the classification scheme of the GRW, but rather makes use of a portfolio of regional characteristics in order to find proper comparison regions for our treatment group that justifies the CIA.

We estimate the probit model of GRW receipt both in cross-sectional settings averaged over different time spans (1994–1998, 1998–2002, and 2000–2004), as well as for a pooled specification, which makes use of three-year averages in the entire interval 1993–2008.[15] The motivation for the design of different subsamples is twofold. First, we want to

---

[13] Data were kindly provided by the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR 2011).
[14] Details with regard to the selection of indicators for the classification of GRW-funded regions can be found at http://www.bbsr.bund.de/cln_032/nn_495082/BBSR/DE/Raumentwicklung/-StrukturAusgleichspolitik/RegionaleStrukturpolitik/Projekte/FoerdergebieteMittelaufteilung/-FoerdergebieteMittelaufteilung.html.
[15] All estimations are performed using Stata Version 12. Binary matching was performed using "psmatch2" (Leuven and Sianesi 2003); generalized propensity-score matching was performed using "doseresponse" (Bia and Mattei 2008). Sensitivity checks have been performed based on "mhbounds" (Becker and Caliendo 2007).

quantify the effectiveness of GRW for different time periods. Second, we synthetically define pre-treatment periods and control for pre-treatment difference among the regions' initial position in order to exclude feedback effects throughout the matching approach. This procedure is important for the success of the matching approach in terms of excluding any simultaneity bias stemming from feedback effects of the output variable on the treatment indicator and the vector of conditioning factors $\mathbf{x}_i$.

Estimation results for the alternative sample periods are shown in Table 1. Consistent with our theoretical expectations, the initial (logged) productivity gap is statistically significant and positively correlated with the probability of GRW receipt. The same holds for the average firm size. In contrast, the share of manufacturing-sector employment in total regional employment and the regional human-capital endowment show negative coefficient signs. The negative correlation of the latter variables can be explained with regard to the specific situation of supported regions in East German. On one hand, these regions are still characterized by a large fraction of employees with a high level of formal education. On the other hand, these regions have also faced severe structural breaks in terms of transforming and deindustrializing their local economies in the aftermath of German reunification. As a result of this "unification shock", East German regions experienced as strong decline in manufacturing-sector activity and still show, on average, a low level of industrial concentration compared to the West German average. At the same time, they receive large amounts of GRW support, which drives the observed negative correlation between GRW funding and the share of manufacturing-sector employment in total regional employment.

Table 1: Propensity-score (PS) estimation for GRW receipt (probit specification)

Treatment (0/1): Receipt of GRW funding	**Cross-section 1994–1998**	**Cross-section 1998–2002**	**Cross-section 2000–2004**	**Pooled 1993–2008**
Log(initial income gap)	1.452***	1.823***	2.425***	2.085***
(S.E.)	(0.2428)	(0.2518)	(0.3024)	(0.1235)
Log(average firm size)	1.325***	1.244***	0.559*	0.733***
(S.E.)	(0.2897)	(0.2981)	(0.3110)	(0.1335)
Log(share manufacturing sector)	−1.684***	−1.482***	−1.093***	−1.245***
(S.E.)	(0.2758)	(0.2747)	(0.2809)	(0.1212)
Log(human capital)	−0.892***	−0.761***	−0.398	−0.491***
(S.E.)	(0.2585)	(0.2636)	(0.2812)	(0.1206)
Log(population density)	−0.192	−0.118	−0.166	−0.088*
(S.E.)	(0.1222)	(0.1166)	(0.1286)	(0.0472)
Urban municipality indicator	−0.232	−0.280	−0.569*	−0.286**
(S.E.)	(0.2874)	(0.2762)	(0.3005)	(0.1246)
Settlement structure	0.034	0.041	0.000	0.021
(S.E.)	(0.0391)	(0.0394)	(0.0415)	(0.0175)
Observations	398	408	408	2020
Pseudo R^2	0.29	0.30	0.35	0.32
Time fixed effects	No	No	No	Yes

Notes: Pooled data using three-year intervals. S.E. is standard error. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The remaining variables (population density, urban municipality indicator, and settlement structure) turn out to be statistically insignificant in most specifications. Only for the pooled specification do we get empirical evidence for a negative correlation of population density and the urban-municipality indicator with the receipt of GRW funding, indicating that GRW funds—controlling for the city status—were mainly directed to agglomerated regions.

Having estimated the propensity score as the prerequisite for the selection of an appropriate comparison group, we can proceed with the actual matching approach. We chose the k nearest-neighbor algorithm, where each treated region is matched by its five ($k = 5$) nearest neighbors measured in terms of the estimated propensity score according to Table 1. We further apply a common-support restriction to our 5-NN matching routine in order to minimize the risk of bad matches and to avoid introducing bias. The results for the different sample designs are shown in Table 2. The table reports both the mean value of labor-productivity growth for the treated as well as the non-treated comparison group.

As outlined above, for the different cross-sectional sample designs, we use an evaluation interval of five years which is not allowed to overlap with the sample period for the propensity-score estimation in order to eliminate direct feedback effects. To illustrate this point, we estimate the outcome difference between treated and non-treated throughout the year 1999–2004 if the propensity score was calculated for the period 1994–1998 and so forth. For the pooled-data case, we use a three-year lead in the matching approach compared to the calculation of the associated propensity score. As the table shows, the estimated ATET parameter (τ_{ATET}) turns out to be positive and statistically significant for most sample periods except for the first evaluation period 1999–2004. While the latter result may be motivated by a general business cycle downturn for that period, which also led to a significant reduction in the growth rate differential among German regions, the general impression from Table 2 is that growth in labor productivity is higher for GRW-funded regions compared to non-funded comparison units. The additional growth impulse is around 0.5 percentage points, which is about 20 percent of the total growth rate of treated regions.

To evaluate the sensitivity of the obtained results with regard to the “balancing properties” of the propensity-score estimation, we use the pseudo-R^2 test proposed by Sianesi (2004). The approach involves a re-estimation of the propensity score model only for the matched sample and then a comparison of the resulting pseudo-R^2 to the one obtained before matching. Since matching should balance the two groups, the pseudo-R^2 based on the matched sample should be low. As shown in Table 2 (compared to Table 1), the ex-post pseudo-R^2 indeed drops by almost two-thirds of its initial “fit” (8–12 percent compared to 29–35 percent in the first-stage estimation).

However, if we additionally compute a likelihood-ratio test of whether the ex-post pseudo-R^2 is statistically different from zero, the null hypothesis of zero explanatory power of the covariates in the matched sample is still rejected. This result raises some critical reflections on the reliability of the estimation results given that a complete balancing of covariates is not possible for the sample at hand. The implication of the likelihood-ratio test is that the regional variation captured by the set of covariates may not be sufficient in order to isolate the causal effect of GRW on productivity growth. Stated differently, the assumption of conditional independence is less plausible in the present situation. Of course, the result is not surprising given the rather small set of regional entities at hand (N = 413), where only few covariates are at our disposal while the regional units itself form aggregated observations stemming from complex structural interdependencies at the sub-regional level.

Table 2: Nearest-neighbor ($k = 5$) matching and Rosenbaum bounds

Outcome: Average growth rate of labor productivity	**Cross-section 1999–2004**	**Cross-section 2003–2007**	**Cross-section 2005–2008**	**Pooled 3-Year Lead**
Mean of Treated	0.019	0.023	0.024	0.023
Number (com. sup.)	214 (191)	210 (189)	183 (132)	1027 (169)
Mean of Control	0.018	0.019	0.019	0.018
Number (com. sup.)	184 (184)	198 (198)	225 (225)	993 (993)
τ_{ATET}	0.001	0.004**	0.006**	0.005***
(S.E.)	(0.0021)	(0.0019)	(0.0024)	(0.0016)
Ex-post pseudo-R^2	0.087	0.119	0.086	0.120
LR test	46.1***	62.2***	31.3***	283.2***
Γ_1	0.001	0.004***	0.005***	0.004***
(p-value)	(0.15)	(0.00)	(0.00)	(0.00)
$\Gamma_{1.5}$	−0.001	0.002**	0.003**	0.000
(p-value)	(0.92)	(0.05)	(0.03)	(0.31)
Γ_2	−0.003	0.000	0.001	−0.002
(p-value)	(0.99)	(0.49)	(0.28)	(0.99)

Notes: S.E. is standard error. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

As a second sensitivity test, we apply the Rosenbaum bounding approach to quantify the probability that, for two regions with identical observed covariates, their chances of receiving GRW treatment actually differ due to unobservable characteristics. If the latter probability is not zero, both regions will differ in their odds of receiving treatment by a factor that involves a parameter Γ. The computation of different values for Γ in Table 2 reveals that an unobserved factor needs to cause the odds-ratio to differ by at least a factor of 1 to 1.5 in order to result in statistically insignificant outcome differences as a worst-case scenario. To illustrate the magnitude of a hidden bias that would force us to revise our statistical findings, we can equate the magnitude of this bias in terms of equivalent effects for observed covariates for which we can actually calculate it. For instance, a critical level of $\Gamma = 1.5$ is attained at a difference in human-capital endowments of more than 3.5 percent (with the sample mean being equal to 7 percent). Thus, the unobserved effect needs to be rather substantial compared to the distribution of the variable in order to have a statistically significant impact on the obtained result.

Keeping the potential pitfalls in mind, we may carefully argue that we have established a positive effect of GRW receipt on regional productivity growth. For the pooled specification, we thus obtain—on average—an additional annual growth effect for labor productivity of roughly 0.5 percentage points for GRW-funded regions. On top of this result, we finally want to take a closer look at the relationship between the actual funding volume and the regional productivity-growth performance. This allows us to identify a maximum level of funding with positive marginal growth effects, that is, the level beyond no further growth effects can be observed. This second step involves the use of a generalized propensity score (GPS) to compute dose–response functions.

In Table 3, we report the OLS estimates for the GPS estimates, where the dependent variable is the GRW intensity defined as GRW volume per unit of labor force (in 1000 €) for German NUTS-3 regions.[16] In order to have a sufficiently high number of observations, we focus on two pooled specifications here:[17] (1) a sample design with three-year averages for the period 1993–2008 in analogy to the binary matching approach outlined above and (2) an annual model. The set of regressors comprises lagged levels and growth rates of labor productivity and employment, as well as the investment intensity, the average firm size, foreign turnover, the share of manufacturing sector in total employment, human capital, population density, and the two indicator variables for the municipality status and settlement structure as introduced above.

Since the GPS approach requires Normally distributed residuals, we chose a Box–Cox transformation for our dependent variable in Table 3. The latter transformation is the only operationalization that ensures Normally distributed errors as indicated by the results of a Kolmogorov–Smirnov test conducted for the variable in levels, logarithmic, as well as Box–Cox transformation. Based on the estimated GPS as well as the treatment variable GRW, we can then compute the dose–response function by first regressing[18]

$$\mathrm{E}[y_i|GRW_i, GPS_i] = \alpha_0 + \alpha_1 GRW_i + \alpha_2 GRW_i^2 + \alpha_3 GPS_i + \alpha_4 (GRW_i \times GPS_i)$$

and using the obtained parameters to estimate the average potential outcome at each treatment level τ:

$$\mathrm{E}[y_i|\widehat{GRW_i}, GPS_i] = \frac{1}{N}\sum_{i=1}^{N}\left[\widehat{\alpha_0} + \widehat{\alpha_1}\tau + \widehat{\alpha_2}\tau^2 + \widehat{\alpha_3}GPS(\tau, X_i) \quad + \widehat{\alpha_4}(\tau \times GPS(\tau, X_i)\tau\right].$$

Additionally, the first derivate of the dose-response function with respect to the GRW transfer intensity can be computed as the so-called treatment-effect function. As Becker et

16 The variable labor force (LF) for each NUTS-3 region has been calculated according to the formula

$$LF = \frac{1}{N}[(Pop_t - Pop_t^{<18} - Pop_t^{>65}) \ (Pop_{t-1} - Pop_{t-1}^{<18} - Pop_{t-1}^{>65})],$$

where *Pop* denotes Population and the superscripts denote subgroups of population with less than 18 years and more than 65 years, respectively. The subscript t and $t - 1$ define the time periods.

17 An extension of the GPS method to longitudinal data is given by Moodie and Stephens (2012).

18 As suggested by Hirano and Imbens (2004), data was organized in groups of the treatment intensity. We have chosen to discretize the treatment intensity into five groups according to the quintiles of the distribution.

al. (2012) point out, the latter can be used to infer the maximum desirable treatment intensity of regional policy. In order to reduce the sensitivity of the estimates with respect to large outliers, we restrict the calculation of the dose–response function up to the 90th percentile of the distribution of GRW funding.

Table 3: Generalized propensity-score (GPS) estimation

Treatment: GRW intensity	**Three-year interval 1993–2008**	**Annual 1993–2008**
Log(lagged labor productivity level)	−0.159***	−0.127***
(S.E.)	(0.0197)	(0.0146)
Log(lagged labor productivity growth)	9.158***	9.425***
(S.E.)	(2.687)	(1.562)
Log(lagged employment level)	−0.078	0.049
(S.E.)	(0.1705)	(0.1261)
Log(lagged employment growth)	0.754	3.451
(S.E.)	(0.1891)	(3.2264)
Log(investment intensity)	0.755***	0.774***
(S.E.)	(0.1891)	(0.1255)
Log(average firm size)	−0.502	−0.811**
(S.E.)	(0.3463)	(0.2563)
Log(foreign turnover)	−0.135	−0.358***
(S.E.)	(0.1457)	(0.1036)
Log(share manufacturing sector)	0.219	0.514**
(S.E.)	(0.2679)	(0.2035)
Log(human capital)	3.133***	3.456***
(S.E.)	(0.3191)	(0.2281)
Log(population density)	−0.121	−0.095
(S.E.)	(0.1172)	(0.0641)
Net migration indicator	−1.056***	−0.786***
(S.E.)	(0.2178)	(0.1325)
Urban Municipality indicator	0.209***	1.102***
(S.E.)	(0.0429)	(0.2231)
Settlement structure	0.067***	0.239***
(S.E.)	(0.0177)	(0.0319)
Observations	820	1832
Time fixed effects	Yes	Yes
Kolmogorov–Smirnov test	0.032	0.014
(p-value)	(0.36)	(0.84)

Notes: S.E. is standard error. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Estimated dose–response function

Outcome: Growth rate of labor productivity	Three-year interval 1993–2008	Annual 1993–2008
GRW	−0.003***	−0.002
(S.E.)	(0.0007)	(0.0096)
GRW^2	0.0002***	0.0001
(S.E.)	(0.00006)	(0.0007)
GPS	−0.053*	−0.071*
(S.E.)	(0.0302)	(0.0383)
GRW × GPS	0.015***	0.016***
(S.E.)	(0.0037)	(0.0048)
Observations	820	1832
Balancing Property (*F*-test)	Reject	Reject
(*p*-value)	(0.00)	(0.00)

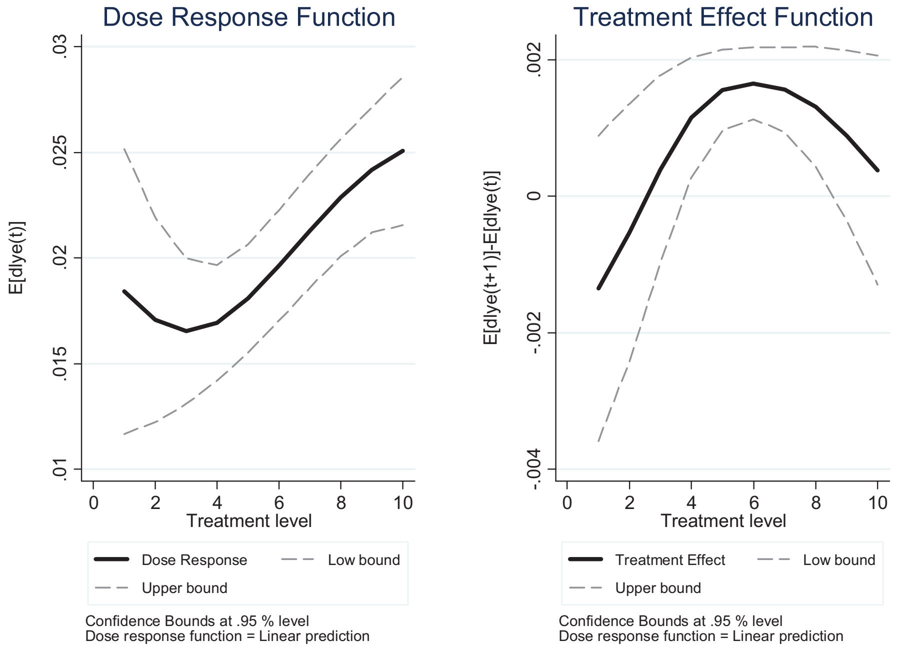
Notes: GRW as GRW per employee based on a Box–Cox transformation. The constant is not reported. S.E. is standard error. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The dose–response function shows how labor-productivity growth responds to changes in the GRW intensity. In order to interpret the results of the estimated dose–response function as shown in Table 4, we plot the dose–response and treatment-effect functions in Figure 1. The treatment-effect function is the first derivative of the dose–response function. Of particular interest is the graph of the treatment-effect function on the right-hand side of Figure 1, since it allows us to identify the treatment level which is associated with zero marginal increase in regional productivity growth. As the figure shows, this is the case for a treatment level of approximately 8 (in its Box–Cox transformation), which corresponds to a GRW intensity of roughly T€ 105 per unit of labor force and is about two-thirds of the maximum observed funding intensity (67th percentile of the distribution of GRW intensity).[19]

For higher funding intensities, the GRW support is shown to be ineffective since it fails to induce an additional growth stimulus. Theoretically, a maximum desired treatment level can be explained by the existence of diminishing returns to investment, that is, increasing funding intensities are associated with lower returns on investment. Additionally we can also observe that a minimum treatment intensity is necessary to induce a positive growth stimulus (at roughly 20 percent of the distribution of GRW intensity, which corresponds to T€ 10 per unit of labor force). Together with the maximum treatment intensity, this results in an inverted U-shape of the treatment-effect function as shown in Figure 1.

[19] Here, the lower bound of the 95-percent confidence level intersects the zero line of no additional productivity growth effect. In all, the distribution of GRW per labor force can take values from T€ 0.5 to T€ 420.

Figure 1: Dose–response and treatment-effect function for GRW intensity



Notes: Dose–response and treatment-effect functions are based on the estimates for the annual panel 1993–2008. Confidence intervals have been calculated using bootstrap methods.

Our empirical results for Germany thus lie between the findings of Becker et al. (2012), who find a maximum treatment intensity equal to the 18th percentile of the regional distribution of EU transfers as share of GDP, and Hagen and Mohl (2008), who find positive but statistically insignificant effects for EU regional transfers. However, for the binary case, our results have to be interpreted with some caution, since the balancing property of the covariates in the matched sample is not satisfied (using an *F*-test as indicated in Table 4). This supports our expectations from above that, for regional data, where only a fixed (and small) set of covariates are available, it is rather hard to find perfect statistical twins.

Does this mean one should not apply the matching approach in regional science and policy analysis at all? Clearly not, since this problem is not uniquely restricted to the matching approach. To make this point clearer, one can simply bear in mind that the regression approach can be seen as a particular form of matching (for details, see Angrist and Pischke (2010)). This close relationship between matching and regression may also be seen when one does a weighted regression, with the weights equal to the inverse probability of being selected into treatment. Ordinary least squares may be viewed as matching with equal weights. The point is that the difficulty of comparing apples to apples and oranges to oranges in the matching context carries over to the regression framework (just that regression analysis typically does not address this problem properly). One potential solution to circumvent this problem would be to rely much more on individual- and firm-level datasets in the conduct of regional-policy analysis. However, the disadvantage of such an approach is that it is typically not possible to identify regional net effects if the level of the analysis is the individual firm.

6 Conclusion

In this paper, we have applied methodological advances stemming from the so-called experimentalist school of applied econometrics to the analysis of regional-policy evaluation in Germany. Starting with a general overview of this relatively new field in the context of empirical regional science, we have discussed its merits and likely pitfalls. The experimentalist approach is very appealing since it puts a strong emphasis on the politically relevant notion of causality and the identification of causal effects of policy actions.

Taking the evaluation of the German *Gemeinschaftsaufgabe "Verbesserung der regionalen Wirtschaftsstruktur"* (GRW) as an empirical case study, our results for a binary propensity-score-matching approach show that GRW-funded regions indeed experienced a higher labor-productivity growth compared to non-funded regions throughout the sample period 1993–2008. This indicates that the GRW policy is successful in fostering convergence and equalizing standards of living in Germany. Using a generalized propensity-score model with a continuous treatment variable, we also find that, up to a funding intensity of roughly two-thirds of the regional distribution of GRW payments, higher treatment intensities ensure higher productivity growth. Thus, in line with earlier work on German and EU regional funding, we obtain empirical evidence that regional policy is effective but only up to a certain treatment level.

As our empirical application has also shown, there are some caveats though. The most severe problem of the application of the experimentalist approach in regional science is that regional data exhibit special features that are likely to complicate empirical applications, particularly in terms of satisfying the so-called "balancing property". As both the estimation results for the binary and for the generalized propensity-score-based matching approaches have shown, it is very hard to find proper statistical twins for a fixed set of regional observations. One potential solution to circumvent this problem would be to rely much more on individual- and firm-level datasets in the conduct of regional policy analysis. However, the disadvantage of such an approach is that it is typically not possible to identify regional net effects if the level of the analysis is the individual firm.

Moreover, the crucial assumption of "no general-equilibrium effects" (SUTVA) is difficult to justify in a regional setting, for instance, in the presence of spatial spillovers. Nevertheless, this does not mean that applications of the experimentalist school are a dead end in regional science and policy analysis. First, standard regression approaches have the same problems while lacking the transparency and rigor to isolate causal effects. Second, recent applications such as in Chagas et al. (2012) seek to find ways to include spatial effects in the analysis of matching models. An alternative approach would be to apply spatial filtering techniques in order to augment the (generalized) propensity-score-matching approach. Thus, the tools applied in this paper appear to be an interesting addition to the standard toolkit in empirical regional science and policy analysis and mark a fruitful research agenda for the future.

References

Alecke, Björn, Timo Mitze, and Gerhard Untiedt (2012). "Growth effects of regional policy in Germany: Results from a spatially augmented multiplicative interaction model", *Annals of Regional Science*, *online first.*

Angrist, Joshua D. and Jörn-Steffen Pischke (2010). "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics", *The Journal of Economic Perspectives* **24**(2):3–30.

BBSR (2011). "Laufende Raumbeobachtungen des Bundesinstitutes für Bau-, Stadt- und Raumforschung", various issues. Bundesamt für Bauwesen und Raumordnung: Bonn.

Becker, Sascha and Marco Caliendo (2007). "mhbounds – Sensitivity Analysis for Average Treatment Effects", IZA Discussion Paper No. 2542.

Bia, Michela and Alessandra Mattei (2008). "A Stata package for the estimation of the dose–response function through adjustment for the generalized propensity score", *The Stata Journal* **8**(3):354–373.

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

Chagas, André Luis Squarize, Rudinei Toneto, and Carlos Roberto Azzoni (2012). "A spatial propensity score matching evaluation of the social impacts of sugarcane growing on municipalities in Brazil", *International Regional Science Review* **35**(1):48–69.

Deaton, Angus (2010). "Instruments, randomization, and learning about development", *Journal of Economic Literature* **48**(2):424–455.

Dehejia, Rajeev H. and Sadek Wahba (1999). "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs", *Journal of the American Statistical Association* **94**(448):1053–1062.

Dehejia, Rajeev H. and Sadek Wahba (2002). "Propensity score-matching methods nonexperimental causal studies", *The Review of Economics and Statistics* **84**(1):151–161.

Dell, Melissa (2010). "The persistent effects of Peru's mining *mita*", *Econometrica* **78**(6):1863–1903.

DiPrete, Thomas and Markus Gangl (2004). "Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments", *Sociological Methodology* **34**(1):271–310.

Dube, Arindrajit, T. William Lester, and Michael Reich (2010). "Minimum wage effects across state borders: Estimates using contiguous counties", *The Review of Economics and Statistics* **92**(4):945–964.

Heckman, James J. (2008). "Econometric causality", *International Statistical Review* **76**(1):1–27.

Heckman, James J. (2010). "Building bridges between structural and program evaluation approaches to evaluating policy", *Journal of Economic Literature* **48**(2):356–398.

Heckman, James J., Hidehiko Ichimura, and Petra Todd (1997). "Matching as an econometric evaluation estimator: Evidence from evaluating a job training program", *Review of Economic Studies* **64**(4):605–654.

Heckman, James J., Hidehiko Ichimura, and Petra Todd (1998). "Matching as an econometric evaluation estimator", *Review of Economic Studies* **65**(2):261–294.

Heckman, James J. and Sergio Urzúa (2010). "Comparing IV with structural models: What simple IV can and cannot identify", *Journal of Econometrics* **156**(1):27–37.

Hirano, Keisuke and Guido W. Imbens (2004). "The propensity score with continuous treatments". In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, Andrew Gelman and Xiao-Li Meng, eds. West Sussex, England, UK: John Wiley & Sons Ltd.

Holland, Paul W. (1986). "Statistics and causal inference", *Journal of the American Statistical Association* **81**(396):945–960.

Holmes, Thomas J. (2010). "Structural, experimentalist, and descriptive approaches to empirical work in regional economics", *Journal of Regional Science* **50**(1):5–22.

Imbens, Guido W. and Jeffrey M. Wooldridge (2009). "Recent developments in the econometrics of program evaluation", *Journal of Economic Literature* **47**(1):5–86.

Keane, Michael P. (2010). "Structural vs. atheoretical approaches to econometrics", *Journal of Econometrics* **156**(1):3–20.

Leuven, Edwin and Barbara Sianesi (2003). "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing", Statistical Software Components S432001, Boston College Department of Economics.

Mohl, Philipp and Tobias Hagen (2008). "Which is the right dose of EU cohesion policy for economic growth?", ZEW Discussion Paper No. 08-104.

Moodie, E.; Stephens, D. (2007): "Estimation of dose-response functions for longitudinal data using the generalized propensity score", *Statistical Methods in Medical Research*, **21**(2): 149-166.

Neyman, Jerzy (1923(1990)). "On the application of probability theory to agricultural experiments. Essays on Principles. Section 9", *Statistical Science* **5**(4):465–472.

Quandt, Richard E. (1972). "A new approach to estimating switching regressions", *Journal of the American Statistical Association* **67**(338):306–310.

Rosenbaum, Paul R. (2002). *Observational Studies*. 2nd ed. New York: Springer.

Rosenbaum, Paul R. and Donald B. Rubin (1983). "The central role of the propensity score in observational studies for causal effects", *Biometrika* **70**(1):41–55.

Roy, A.D. (1951). "Some thoughts on the distribution of earnings", *Oxford Economic Papers* **3**(2):135–146.

Rubin, Donald B. (1986). "Comment: What ifs have causal answers?", *Journal of the American Statistical Association* **81**(396):961–962.

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

Stock, James H. and Francesco Trebbi (2003). "Who invented instrumental variable regression?", *The Journal of Economic Perspectives* **17**(3):177–194.

Wooldridge, Jeffrey M. (2010). *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: The MIT Press.

Appendix: Descriptive statistics of variables in the empirical analysis

Variable	Description	Mean	S.D.	Min.	Max.
Labor productivity	Regional GDP per Employee (in 1000 Euro)	50.71	9.44	22.70	119.52
Employment	Employment level (in 1000)	93.41	121.17	18.38	1638.02
Investment intensity	Gross Fixed capital formation (GFCF) in manufacturing sector as share of total turnover in manufacturing (in %)	4.92	4.02	0.46	76.50
Average firm size	Number of employees per firm	134.92	111.50	35.65	1857.33
Foreign turnover	Share of foreign turnover in total turnover for manufacturing sector (in %)	29.25	14.46	0.00	420.31
Share manufacturing sector	Percentage share of employment in manufacturing sector relative to total employment	24.48	11.07	1.94	696.00
Human capital	Percentage share of school graduates with university qualification (in %)	7.14	3.36	1.93	25.27
Population density	Total population per square kilometer	14.33	94.93	0.38	2637.01
Net migration indicator	Binary indicator whether region has received a net surplus in migrants (internal and external), 0 otherwise	0.63	0.48	0	1
Urban municipality indicator	Binary indicator whether region belongs to a greater administrative district, 0 otherwise	0.73	0.44	0	1
Settlement structure	Indicators variable for different classes of settlement structure (classified according to an ordinal scale with 1 = highly agglomerated to 9 = highly peripheral)	5.39	2.52	1	9
GRW	Binary variable for receipt of GRW subsidies, 0 otherwise	0.45	0.50	0	1
GRW intensity	Volume of GRW subsidies per labor force (in 1000 Euro, T€), where labor force is defined as $LF = \frac{1}{N}[(Pop_t - Pop_t^{<18} - Pop_t^{>65})\ (Pop_{t-1} - Pop_{t-1}^{<18} - Pop_{t-1}^{>65})]$, where Pop is population and the superscripts denote subgroups of population with less than 18 years and more than 65 years, respectively. The subscript t and $t-1$ define the time periods.	68.71	180.39	0.00	3071.29

Notes: Descriptive statistics are given for the whole sample range of 1993–2008. Specific subsample information as used throughout the empirical applications can be obtained from the authors upon request.