European Economic Review 56 (2012) 648–668

Contents lists available at SciVerse ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

Too much of a good thing? On the growth effects of the EU's regional policy

Sascha O. Becker [a,b,c,d,e], Peter H. Egger [f,g,b,c,d,*], Maximilian von Ehrlich [f]

[a] University of Warwick, United Kingdom
[b] Ifo, Germany
[c] CEPR, United Kingdom
[d] CESifo, Germany
[e] IZA, Germany
[f] ETH Zürich, KOF, Weinbergstrasse 35, 8092 Zürich, Switzerland
[g] WIFO, Austria

ARTICLE INFO

Article history:
Received 21 September 2010
Accepted 7 March 2012
Available online 27 March 2012

JEL classification:
C21
O40
H50
R11

Keywords:
EU regional policy
Regional growth
Generalized propensity score estimation
Quasi-randomized experiment

ABSTRACT

The European Union (EU) provides grants to disadvantaged regions of member states from two pools, the Structural Funds and the Cohesion Fund. The main goal of the associated transfers is to facilitate convergence of poor regions (in terms of per-capita income) to the EU average. We use data at the NUTS3 level from the last two EU budgetary periods (1994–1999 and 2000–2006) and generalized propensity score estimation to analyze to which extent the goal of fostering growth in the target regions was achieved with the funds provided and whether or not more transfers generated stronger growth effects. We find that, overall, EU transfers enable faster growth in the recipient regions as intended, but we estimate that in 36% of the recipient regions the transfer intensity exceeds the aggregate efficiency maximizing level and in 18% of the regions a reduction of transfers would not even reduce their growth. We conclude that some reallocation of the funds across target regions would lead to higher aggregate growth in the EU and could generate even faster convergence than the current scheme does.

1. Introduction

As the budget of the European Union (EU) becomes tighter and major recipients of European regional transfers struggle with debt crises, questions about the proper utilization and effectiveness of transfers from the central EU budget to Europe's poorest regions are hotly debated. Since 1975, when the European Regional Development Fund (ERDF) was founded, a significant budget has been devoted to the reduction of regional imbalances, especially, in terms of per-capita income.[1] The Treaty of Lisbon which entered into force in 2009 acknowledges *regional cohesion* as one of the key goals of the European Union.[2]

* Corresponding author at: ETH Zürich, KOF, Weinbergstrasse 35, 8092 Zürich, Switzerland. Tel.: +41 44 632 41 08.
E-mail addresses: s.o.becker@warwick.ac.uk (S.O. Becker), egger@kof.ethz.ch (P.H. Egger), vonEhrlich@kof.ethz.ch (M. von Ehrlich).

[1] The European Social Fund (ESF) and the European Agricultural Guidance and Guarantee Fund (EAGGF) were already founded in 1958 and 1962, respectively, but were focused on specific duties and were limited in scope. The Cohesion Fund was founded as late as 1992.

[2] Article 174 of the Treaty on the Functioning of the European Union states: "[⋯] the Union shall aim at reducing disparities between the levels of development of the various regions and the backwardness of the least favored regions" (see Official Journal C 115/127 09/05/2008).

0014-2921/$ - see front matter
http://dx.doi.org/10.1016/j.euroecorev.2012.03.001

The Union's regional policy goals are rooted in the perception that a common market requires a certain degree of homogeneity in economic development which is not necessarily an automatic outcome of the integration process but, eventually, has to be assisted by active policy interventions. Accordingly, with the EU enlargements to the south[3] and, more recently, to the east,[4] the disparities among the member countries of the Union increased sharply, and so did the scope of regional transfers. During the years 1975–1988, the ERDF budget represented on average 6.8% of the total Community budget, while during the current 2007–2013 programming period expenses aimed at cohesion make up 35.7% of the total Community budget, or 347.41 billion Euros at current prices (see European Commission, 1989, 2008). These expenses on cohesion policy stem from different funds: the ERDF contributes about 58%, the European Social Fund (ESF) about 22%, and the *Cohesion Fund* about 20%. The ERDF and the ESF are commonly referred to as the *Structural Funds* where the former focuses on infrastructure investments and the latter on employment measures.[5] The Cohesion Fund was established in the treaty of Maastricht and is intended to support the Structural Funds in strengthening the economic and social cohesion in the Union. The Cohesion Fund mainly finances environmental projects and trans-European transport infrastructure networks. In contrast to the Structural Funds, the Cohesion Fund operates on the national rather than the regional level.[6]

The heterogeneity of regional transfer intensity – defined as the amount of EU transfers in percent of a target region's beginning-of-period GDP – across recipient regions and programming periods is remarkable. Whereas some NUTS3 regions[7] received only negligible amounts of EU transfers in the order of less than a thousandth of a percent of their GDP, others faced a transfer intensity of 29% of their beginning-of period GDP. We will discuss this heterogeneity in more detail below.

It is sometimes argued that some regions use EU transfers increasingly inefficiently as they receive more transfers. Due to a lack of administrative capacity, part of the funds is not spent as intended but used for consumption purposes or subject to corruption.[8] If there are diminishing returns to EU regional transfers, knowing that they foster growth *on average*, as in Becker et al. (2010), is not enough.[9] In fact, it is important to understand how a varying treatment *intensity* (different amounts of EU transfers relative to GDP) affects regional growth. This will allow us to see up to which level transfers serve the intended goal of fostering regional growth and beyond which a further allocation of funds becomes inefficient. Estimation of that threshold for the EU's regional policy programmes calls for an identification strategy that goes beyond a binary transfer indicator and exploits variation in transfer intensity.

An argument for a declining treatment effect – and, eventually, existence of a *maximum desirable level of regional transfers* – arises naturally from neoclassical production theory and the assumption of diminishing returns to investment and investment-stimulating transfers (see Hirshleifer, 1958). Suppose that investment projects are financed and undertaken in the order of expected returns on investment. Then, a bigger number of investment projects carried out would be associated with a lower return to investments (or transfers). If diminishing returns to transfers were relevant, we could identify a maximum desirable level of the treatment intensity. Above that level, no additional (or even lower) per-capita income growth effects would be generated than at or below that threshold.

There is a similar argument for a *minimum necessary level of regional transfers* which is based on the big-push or poverty-trap theory of development, which states that transfers (or aid) have to exceed a certain threshold in order to become effective. For instance, this would be the case if the marginal product of capital were extremely low at too small levels of infrastructure or human capital (see Sachs et al., 2004). Alternatively, this could be the case if regions lagging behind were isolated from other developed regions (see Murphy et al., 1989, for arguments along those lines). When applying the big-push or poverty-trap theory to the least-developed NUTS3 regions in the EU, one would expect to find a

---

[3] Greece joined the EU in 1981, and Spain and Portugal in 1986.

[4] Cyprus, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, the Slovak Republic, and Slovenia joined in 2004, and Bulgaria and Romania in 2007.

[5] Until 2006, the Structural Funds included the European Agricultural Guidance and Guarantee Fund (EAGGF) and the Financial Instrument for Fisheries Guidance (FIFG) which have been replaced by the European Agricultural Fund for Rural Development (EAFRD) and the European Fisheries Fund (EFF), respectively. Both funds are no longer directly involved in cohesion policy.

[6] Member states qualify for transfers from the Cohesion Fund if their GDP per-capita falls below 90% of the community average. The most significant amount of Structural Funds is transferred to regions with a per-capita GDP below 75% of the community average (so-called Objective 1 regions).

[7] Eurostat defines NUTS3 regions as entities of between 150 and 800 thousand inhabitants. An exception is large cities with population of more than 800 thousand which are still usually NUTS3 regions in their entirety. The counterpart to a NUTS3 region in the United States would be a county. In France, they represent *Départements*, in Germany, they are equivalent to *Landkreise*, in Spain, they correspond to *Comunidades Autónomas*, and in the United Kingdom, they are associated with the *Unitary Authorities*.

[8] See euobserver.com from October 20, 2009, ‘‘EU funds still vulnerable to fraud in Bulgaria’’, Handelsblatt from March 2, 2010 ‘‘Korrumpierter Staatsapparat: EU duldet Griechenlands Betrug seit Jahren’’, the New York Times from August 23, 2008, ‘‘EU cuts back funding to Bulgaria’’, or euractive.com from December 8, 2008, ‘‘Time to redesign the Structural Funds system’’.

[9] Becker et al. (2010) provide an overview of the literature on the effects of the EU's regional transfers and conduct an evaluation of Objective 1 transfers, which make up two thirds of the EU's Structural Funds Programmes. More specifically, Becker et al. (2010) use a binary treatment indicator in a regression discontinuity design to study the causal effects of Objective 1 funds on GDP per-capita growth in recipient versus non-recipient regions. The discontinuity arises from the rule that EU regions whose GDP per-capita falls below 75% of the EU average are eligible for Objective 1 funds whereas regions above the 75% threshold are ineligible. Their results suggest that, in a best-case scenario, Objective 1 transfers generate a multiplier of approximately 1.2 so that every Euro of transfers generates 20 extra cents of GDP. However, that multiplier effect relates to Objective 1 treatment only, since other parts of the Structural and Cohesion Funds do not follow a clearly defined rule (75% threshold) and do not lend themselves to a regression discontinuity design for identification.

minimum desirable level of regional transfer intensity only above but not below which transfers generate positive growth effects. Then, it would be reasonable to allocate more transfers to a few very poor regions in order to ensure that they induce noticeable effects.

With regions above a maximum desirable treatment intensity or below a minimum necessary treatment intensity, the overall EU budget could be reduced without any negative growth effects and, hence, there would be scope for unambiguous efficiency gains. In this analysis we also ask what the empirically *optimal transfer intensity* is. This will be the transfer level above which an additional Euro transferred yields less than a Euro of additional GDP. Hence, what we dub optimal transfer intensity here is associated with a transfer multiplier of unity. Accordingly, a reallocation of transfers from regions above the optimal transfer intensity to ones below it would enhance aggregate growth (although it might hurt growth in the regions from which transfers are taken away).

We aim at identifying the functional form of the relationship between EU-NUTS3 regional transfer *intensity* and per-capita income growth by way of dose–response function estimation.[10] Unlike the study of Becker et al. (2010) and other studies using a binary indicator for EU regional transfer treatment, the dose–response function allows us to ask to which extent the European Commission in conjunction with regional authorities at the national and sub-national levels provide and use transfers in an efficient – here to be interpreted as *per-capita-income growth maximizing* – way.[11] We identify the GDP per-capita growth-maximizing transfer intensity, which allows us to determine how many and which regions receive too much funding and how many and which regions receive too little funding out of the Structural and Cohesion Funds Programme.

The results for the two programming periods 1994–1699 and 2000–2006 point to a non-linear relationship between the treatment intensity of EU regional transfers and per-capita GDP growth. More specifically, we find evidence of a *maximum desirable treatment intensity*. At a transfer intensity beyond this level, the null hypothesis of zero (or even negative) growth effects induced by additional transfers can no longer be rejected. Contrary to the big push hypothesis, within the EU there is no evidence for the existence of a *minimum necessary level of regional transfers* to induce positive per-capita income growth effects.

The estimates suggest that, up to a *maximum desirable treatment intensity* of about 1.3% of a region's GDP, EU transfer receipts from Structural Funds or the Cohesion Fund lead to positive marginal income growth effects. However, beyond a treatment intensity of 1.3%, per-capita income growth can on average not be increased any further through *additional* EU transfers. About 18% of NUTS3 recipient regions received transfers above that threshold. According to our results, a reallocation of the transfers away from those regions would not harm them, but might benefit other regions. When applying the stricter criterion of an *optimal treatment intensity*, we find that transfers should not exceed a treatment intensity of about 0.4%. According to our estimates, the transfer-multiplier fell short of unity for about 36% of the NUTS3 recipient regions across the two periods considered. This leads to the conclusion that there is significant scope for greater efficiency at the level of Structural and Cohesion Funds transfers regarding their growth-maximizing allocation for the Union as a whole as well as its poorest regions.

The remainder of the paper is structured as follows. The next section presents details on the sources and the construction of data at the level of NUTS3 regions for the two programming periods 1994–1999 and 2000–2006. Also, that section summarizes descriptive statistics. Section 3 discusses the econometric methodology applied for the identification of causal effects of the EU's regional transfers on growth. Section 4 presents the results and interprets the findings against the background of efficiency. The last section concludes with a summary of the most important findings.

2. Data and descriptive statistics

Our data stem from several sources. Information on EU transfers to NUTS3 regions has been kindly provided by ESPON (European Spatial Planning Observation Network) for the years 1994–1999 and the European Commission's Directorate General for Regional Policy for the year 2000–2006. The latter information originates from a study on the expenditure of EU regional policy carried out by SWECO. We link those data to various regional characteristics from Cambridge Econometrics' Regional Database and a measure of countries' voting power in the EU Council (measured by the Shapley and Shubik, 1954 index) which is taken from Felsenthal and Machover (1998) for the first programming period and from Widgren (2009) for the second programming period.

In total, our data-set consists of 2280 region-programming-period observations out of which 2078 received transfers through one of the two programmes considered here (Structural Funds or Cohesion Fund). Of the 2078 treated units, 702 classify as Objective 1 regions which received the lion's share of total EU transfers considered (74% on average across the two programming periods). A total of 363 of the 2078 treated units received transfers from the Cohesion Fund. Table 1 provides details on the number and characteristics of NUTS3 regions during the two programming periods 1994–1999 and 2000–2006. We pool the two programming periods for the sake of greater precision of the estimated relationship between

[10] Earlier studies by Becker et al. (2008) and Hagen and Mohl (2008) used variation in the extent of transfers but did not have access to data at the disaggregated NUTS3 level as we do now, so that robust identification of the functional relationship between EU regional transfer intensity and per-capita income growth effects was not possible there.

[11] Note that we take the revenue side of the EU budget as given because each country contributes a fixed percentage of GDP and VAT to the EU budget so taxation is non-progressive. Moreover, data on sub-national contributions to the EU budget are not available. Taking the revenue side as given implies that we disregard the (hardly quantifiable) efficiency costs of raising the necessary tax revenue for transfers.

Table 1
EU regional transfers and GDP per-capita growth in NUTS3 regions.

	Mean	Std. dev.	Min	Max	Treated obs.
Annual transfers per treated region					
Sample: all regions receiving EU transfers from either Structural Funds or Cohesion Funds budget					
Total EU transfers (mn. Euros)	23.141	49.744	0.005	778.531	2078
Total EU transfers/GDP (%)	0.759	1.512	0.00009	29.057	2078
Sample: regions receiving EU transfers from the Structural Funds budget under the Objective 1 heading					
Objective 1 transfers (mn. Euros)	52.131	68.869	0.603	778.531	702
Objective 1 transfers/GDP (%)	1.991	2.103	0.076	29.057	702
Sample: regions receiving EU transfers from the Cohesion Funds budget					
Cohesion Fund transfers (mn. Euros)	21.479	36.090	0.018	334.935	363
Cohesion Fund transfers/GDP (%)	0.659	0.950	0.002	6.338	363
Annual GDP per-capita growth	0.042	0.017	−0.039	0.138	2078

Notes: Our pooled sample consists of 1091 EU15 NUTS3 regions in the 1994–1999 programming period and 1213 EU25 NUTS3 regions in the 2000–2006 programming period. We miss information on the four French overseas-départements and the two autonomous Portuguese regions Madeira and Azores for both periods. In the second period we loose 12 regions that cannot be assigned to the 1994–1999 data due to a territorial reform in Saxony-Anhalt. Hence, in total we have 2280 treated and untreated observations. In order to obtain annual transfers per GDP we divide the annual transfers by the GDP prior to the start of the respective programming period. This is 1993 for the EU12 in the first period but 1994 for the countries joining in 1995 (Austria, Finland, and Sweden), and 1999 for the EU15 in the second period but 2003 for the accession countries of 2004. Moreover, we adjust for the number of years the respective countries actually received funds. This is 6 years for the EU12 in the first period and 5 years for the countries joining in 1995, and 7 years for the EU15 but 3 years for the new accession countries of 2004 in the second period.

treatment intensity and per-capita income growth. By design, NUTS3 regions of EU member countries as of 1999 are observed twice in the data while EU entrants during 2000–2006 are observed only once. Accordingly, we adjust standard errors of parameters and confidence bounds of treatment effects to account for such repeated observations. Pooling more than two budgetary periods for NUTS3 regions is infeasible since detailed information on treatment intensity for programming periods prior to 1994–1999 is not available at the required disaggregated level.

Table 1 displays the average annual transfers per *treated* NUTS3 region adjusting for the number of years the respective regions actually received transfers. The reason for this adjustment is that Austria, Finland, and Sweden joined the EU only in 1995 and did not receive transfers for the whole programming period 1994–1999. The same is true for the accession countries in 2004 that did not receive transfers for the whole 2000–2006 programming period. As mentioned before, 2078 of the 2280 covered EU-NUTS3 region-programming-period observations received transfers from either the Structural Funds or Cohesion Fund budgets. Table 1 shows that the transfer intensity in these 2078 units varied dramatically. Whereas the Greek region of Grevena displayed a transfer intensity of 29.057% in the 1994–1999 programming period, the Swedish region of Hallands län received EU transfers of only 0.00009% – or 5345 Euros – in the programming period 2000–2006. This variation in NUTS3 regional transfer intensity has three roots: first, the variation in GDP (and per-capita GDP) across NUTS3 regions; second, the variation in transfers to countries, NUTS2 regions,[12] and NUTS3 regions as provided by the European Commission[13]; third, the discretion at the national level or the level of NUTS2 regions about the allocation of funds to NUTS3 entities which fall into their jurisdiction.

The EU spent about 21,934 mn. Euros on regional transfers per annum (out of the Structural Funds and the Cohesion Fund programmes) across the two periods under consideration of which 2952 mn. Euros were spent through the Cohesion Fund and 18,982 mn. Euros were transferred through the Structural Funds programmes. Objective 1 regions received about 16,301 mn Euros from the central EU budget per annum across the two periods.[14]

When using the respective relevant GDP of the year prior to the start of the programming period in the denominator, the average annual regional transfer intensity amounted to 0.759% for all regional transfers, to 1.991% for Objective 1 transfers only, and to 0.659% for Cohesion Fund transfers. While most of the NUTS3 regions received *some* transfers from the central budget, there is considerable variation in the transfer intensity as indicated above. Fig. 1 displays the geographical distribution of total EU transfer per GDP for both programming periods under consideration.

---

[12] NUTS2 regions are somewhat larger clusters of NUTS3 regions with 0.8–3 million inhabitants.

[13] For some types of transfers, such as those falling under the auspices of Objective 1 in the Structural Funds Programmes, eligibility for transfers is determined at the level of NUTS2 regions (with a few exceptions which determine transfers to NUTS3 regions; see Becker et al., 2010, for a detailed description of the rules for Objective 1 treatment). Other types of transfers are determined at the NUTS3 level or the national level.

[14] Note that these figures refer to the transfers the *EU spent annually* and cannot be directly compared to the figures in Table 1 which refer to the average annual funds *recipient* regions received. The latter adjusts for the number of years regions actually received funds. For that, we calculate transfers in 1994–1999 divided by 6 or 5 for the EU12 and for the 3 new EU15 members, respectively, and transfers in 2000–2006 divided by 7 or 3 for the EU15 and for the 10 new EU25 members, respectively, before we pool the data and take the average across all observations. In total, there are 12,477 region-year observations with positive transfers. Multiplying the average annual transfers in Fig. 1 with 12,477 and dividing by 13 (the number of years covered by the two programming periods) yields approximately the annual funds the EU spent but does not take into account that regions which received transfers in less than 13 years may have received higher or lower annual funds than other regions in those years they were eligible for funding.

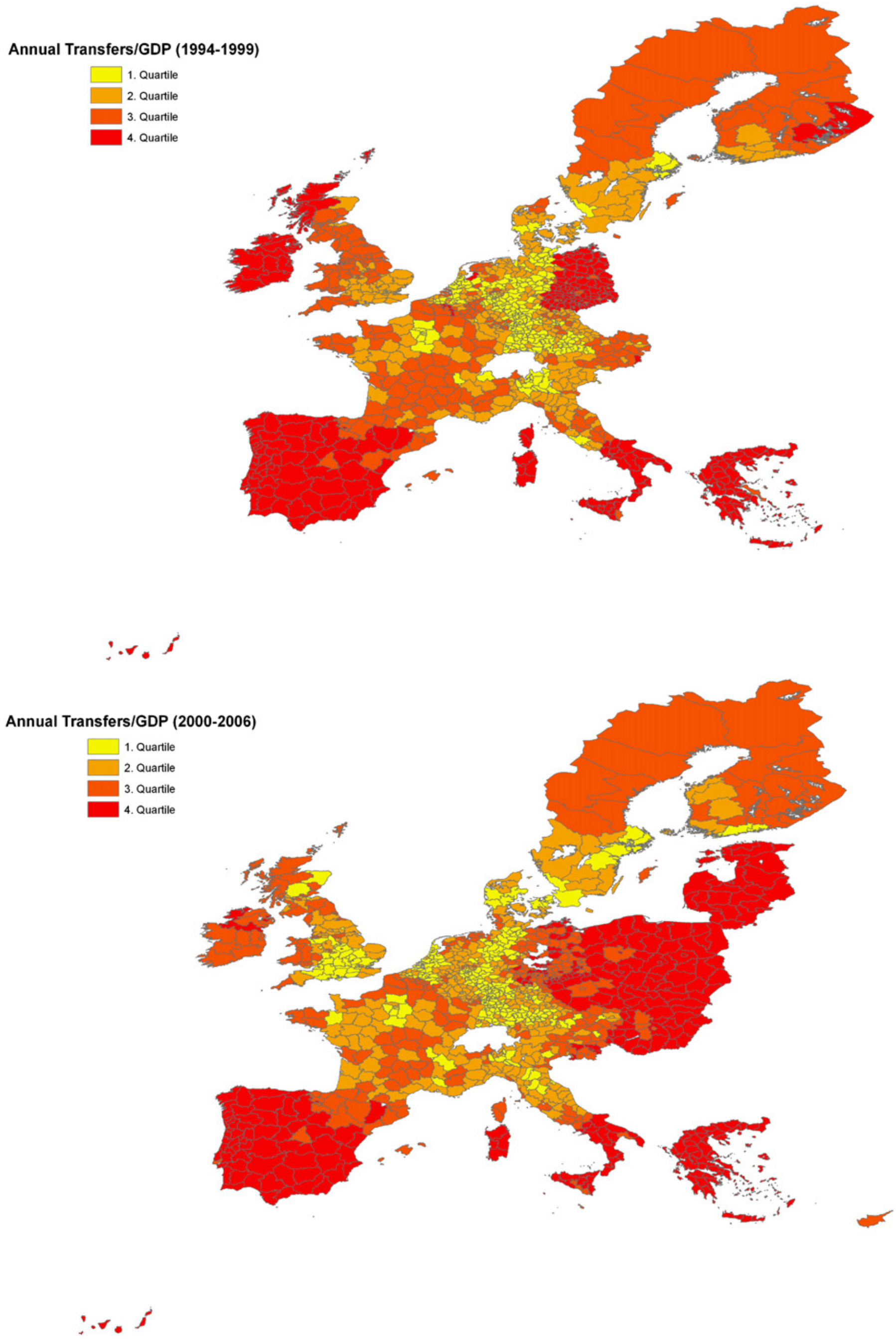

Fig. 1. Regional distribution of EU transfers. *Note*: The maps indicate the annual transfer intensity (total EU transfers per GPD) for the 1994–1999 and 2000–2006 programming periods.

In the subsequent analysis, we focus on those 2078 observations that received regional transfers through either the Structural Funds or the Cohesion Fund programmes. As can be seen from the final row of Table 1, those regions' per-capita income measured at Purchasing Power Parity grew by about 4.2% per annum during the two considered programming periods.[15] However, there is a fair amount of variation in the data. Table 1 suggests that the minimum growth rate across NUTS3 regions reflects a decline by almost 4% per annum while the maximum growth rate was almost 14% per annum within the sample period.

[15] Per-capita income growth is expressed as an average change in log-transformed per-capita income.

Table 2
Descriptive statistics.

Covariates and statistics	Mean (1)	Std. dev. (2)	Min (3)	Max (4)
GDP per-capita	9.583	0.367	8.068	11.038
(GDP per-capita)^2	91.971	7.024	65.098	121.835
(GDP per-capita)^3	883.945	101.057	525.232	1344.806
Shapley–Shubik index	0.090	0.041	0	0.134
Budgetary period dummy	0.478	0.500	0	1
Border region dummy	0.249	0.433	0	1
Employment	4.567	0.919	0.331	7.712
Industrial employment	3.286	1.009	−2.765	6.603
Service employment	4.075	0.976	0.320	7.427
Population density	0.448	0.957	0.002	20.381
Observations	2078			

Notes: Annual GDP per-capita growth is measured at PPP where we use logarithmic growth rates between 1993 and 1999 for the first period, and logarithmic growth rates between 1999 and 2006 for the second period. Time-varying covariates as per-capita GDP (PPP), employment measures, etc. refer to initial values, i.e., 1993 for the first period and 1999 for the second period. Total employment, industrial employment, service employment, and per-capita GDP are measured in logarithmic terms. We miss information on the four French overseas-départements and the two autonomous Portuguese regions Madeira and Azores for both periods. In the second period we loose 12 regions that cannot be assigned to the 1994–1999 data due to a territorial reform in Saxony-Anhalt.

In our empirical analysis we employ various covariates: the GDP per-capita level (at purchasing power parity, PPP) prior to the respective programming period, total regional employment, sectoral employment, population density, a measure of countries' voting power in the EU, a period dummy and a variable that indicates whether a region is located at the EU border. Table 2 provides descriptive statistics for the data used, where per-capita GDP, and the employment information are measured in logarithmic terms.

3. Generalized propensity scores

3.1. Methodology

To estimate the causal effect of transfer intensity on per-capita income growth, we resort to generalized propensity score (GPS) estimation, a non-parametric method to estimate treatment effects conditional on observable determinants of treatment *intensity*. Propensity score matching represents a well-suited econometric technique for policy evaluation as it is able to correct for selection bias into different levels of treatment intensity by comparing units that are similar in terms of their observable characteristics. Following the seminal paper by Rosenbaum and Rubin (1983) propensity score matching became very popular in the case of binary treatment (see, e.g., Heckman et al., 1997; Dehejia and Wahba, 1999). The binary case was extended to categorial multivalued treatment by Imbens (2000) and, more recently, to continuous treatments (see Hirano and Imbens, 2004; Imai and Van Dijk, 2004).[16] In the following, we outline the method developed by Hirano and Imbens (2004) and apply it to our research question.

Index the regions by $i=1,\ldots,N$ and consider the *unit-level dose–response function* of outcomes $Y_i(\tau)$ (annual per-capita income growth) as a function of treatments $\tau \in \mathcal{T}$ (transfer intensity). We focus on $\tau_0 > 0$, i.e., regions with positive transfers. In the binary case, the treatment would be restricted to $\mathcal{T}=\{0,1\}$. However, our objective is not to analyze whether or not receiving transfers at all boosts growth, but to what extent a higher treatment intensity yields stronger or weaker effects than a lower treatment intensity. Furthermore, we want to derive the optimal treatment intensity. Employing the generalized propensity score methodology, we aim at estimating the average dose–response function across all regions i, $\mu(\tau)=E[Y_i(\tau)]$.

The key challenge is to compare regions with sufficiently similar characteristics but different treatment intensity in order to construct a quasi-experimental setting. For each observation i we observe the vector of covariates X_i, the treatment intensity T_i, and the outcome corresponding to the level of treatment received, $Y_i = Y_i(T_i)$. Let us drop index i for simplicity and assume that $Y(\tau)_{\tau\in\mathcal{T}}$,$T$,$X$ is defined on a common probability space, τ is continuously distributed with respect to a Lebesgue measure on $\mathcal{T}$, and $Y=Y(T)$ is a well defined random variable.

For such a setting, the concept of *unconfoundedness* for binary treatments was generalized by Hirano and Imbens (2004) to one of *weak unconfoundedness* for continuous treatments

$$Y(\tau) \perp T \,|\, X \quad \text{for all } \tau \in \mathcal{T}. \tag{1}$$

Regions differ in their characteristics X such that some are more or less likely to receive a high treatment intensity than others. Weak unconfoundedness means that, after controlling for observable characteristics X, any remaining difference

[16] See Becker and Muendler (2008) and Kluve et al. (forthcoming) for recent applications of GPS estimation in different contexts.

in treatment intensity T across regions is independent of the potential outcomes $Y(\tau)$. Eq. (1) is referred to as weak unconfoundedness because it does not require *joint independence* of all potential outcomes, $Y(\tau)_{\tau\in[\tau_0,\tau_1]},T,X$. Instead, it requires *conditional independence* to hold at given treatment levels.

The generalized propensity score is defined as

$$R = r(T,X), \tag{2}$$

where $r(\tau,x) = f_{T|X}(\tau|x)$ is the conditional density of the treatment given the covariates. Similar to the conventional propensity score with binary treatments, the generalized propensity score is assumed to have a *balancing property* which requires that, within strata of $r(\tau,X)$, the probability that $T=\tau$ does not depend on the value of X. In other words, when looking at two observations with the same ex ante probability (conditional on observable characteristics X) of being exposed to a particular treatment intensity, their actual treatment intensity is independent of X. That is, the generalized propensity score summarizes all information in the multi-dimensional vector X so that

$$X \perp 1\{T=\tau\}\,|\,r(\tau,X). \tag{3}$$

This is a mechanical property of the generalized propensity score, and does not require unconfoundedness. In combination with weak unconfoundedness, the balancing property implies that assignment to treatment is *weakly unconfounded given the generalized propensity score*: if assignment to treatment is weakly unconfounded given pre-treatment characteristics X, then

$$f_T(\tau|r(\tau,X),y(T)) = f_T(\tau|r(\tau,X)) \tag{4}$$

for every τ (see Hirano and Imbens, 2004, for a proof). Hence, we can evaluate the generalized propensity score at a given treatment level by considering the conditional density of the respective treatment level τ. In that sense, we use as many propensity scores as there are treatment levels, but never more than a single score at one treatment level.

We eliminate biases associated with differences in the covariates in two steps (for a proof that the procedure removes bias, see Hirano and Imbens, 2004):

1. Estimate the conditional expectation of per-capita income growth as a function of two scalar variables, the treatment level T and the generalized propensity score R, $\beta(\tau,r) = E[y|T=\tau,R=r]$.
2. Estimate the dose–response function at a particular level of the treatment intensity by averaging this conditional expectation over the generalized propensity score at that particular level of treatment intensity, $\mu(\tau) = E[\beta(\tau,r(\tau,X))]$.

For the latter, one does not average over the generalized propensity score $R=r(T,X)$, but over the score evaluated at the treatment level of interest, $r(\tau,X)$. In other words, one fixes τ and averages over X_i and $r(\tau,X_i)\forall i$.

3.2. Estimating the generalized propensity score and the balancing of covariates

In the following, we apply the methodology outlined above to our data-set of 2078 NUTS3-programming-period observations receiving different levels of transfers from the European central budget. The treatment intensity of interest, T_i, is the average annual amount of EU transfers relative to the NUTS3 level GDP prior to the beginning of the respective programming period (see Table 1 for a summary of treatment intensities). Following Hirano and Imbens (2004), we assume a normal distribution for the treatment intensity given the covariates:

$$T_i|X_i \sim N(\beta_0+X_i\beta_1,\sigma^2), \tag{5}$$

where X_i is a row vector and β_1 a column vector. Since the empirical distribution of EU regional transfers per GDP is positively skewed, we chose a logarithmic transformation. According to the Kolmogorov–Smirnov test (and other conventional test statistics), the log-transformed treatment intensity variable does not violate the assumption of normality. As determinants of treatment intensity, we employ the following observables in X_i. First of all, we use log GDP per-capita (at purchasing power parity) measured prior to the respective programming period. This variable should be included, since it matters for the treatment assignment rule for some types of EU transfers.[17] To allow for a non-linear relationship between treatment intensity and log per-capita income, we include a quadratic and a cubic of log GDP per-capita along with the main effect. Moreover, we include the Shapley and Shubik (1954) index of a country's voting power prior to a budgetary period to account for effects of power-play and lobbying at the country level. Finally, we include several variables characterizing the economic structure of a region such as log employment, log industrial employment, log service employment, an EU border dummy, and population density (measured as inhabitants per square kilometer) prior to a budgetary period.[18] The economic structure of a region is considered to be a key determinant of regional

[17] For instance, NUTS2 regions qualify for Objective 1 transfers if their per-capita GDP falls short of 75% of the EU average. Moreover, the level of GDP per-capita prior to a programming period is a key determinant of subsequent per-capita income growth which we should condition on in order to isolate the impact of transfer intensity on growth.

[18] In a sensitivity analysis, we used employment shares instead of log employment levels. All results are qualitatively and quantitatively insensitive to this choice. However, we use the log-level specification in the paper as it fits the data on treatment intensity better in terms of R^2 than a model using employment shares.

Table 3
Estimation of the generalized propensity score (GPS).

Covariates and statistics	Coef.	Std. err.
GDP per-capita	403.226	55.040***
(GDP per-capita)^2	−42.016	5.787***
(GDP per-capita)^3	1.443	0.203***
Shapley–Shubik index	−4.903	0.809***
Budgetary period dummy	0.672	0.063***
Border region dummy	−0 .054	0.067
Employment	1.964	0.278***
Industrial employment	−0.957	0.107***
Service employment	−0.867	0.197***
Population density	0.055	0.029*
Constant	−1284.350	174.267***
Observations	2078	
R^2	0.561	

Notes: ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

transfers. Table 2 summarizes moments (such as mean, standard deviation, minimum, and maximum) of the distribution of these variables. In Appendix A we perform a sensitivity analysis which takes into account regional infrastructure endowments as well as data on regional environmental issues as determinants of EU regional transfer intensity. The estimated relationships turn out qualitatively and quantitatively similar to the benchmark results which exploit a larger sample.[19]

We estimate Eq. (5) by ordinary least squares as reported in Table 3. Using the observable variables in Table 2 plus a constant, we can explain regional transfer intensity fairly well. According to Table 3, the included covariates explain about 56% of the variation in treatment intensity. All of the covariates except one, namely an indicator variable identifying regions at the EU border, exert a significant impact on treatment intensity at least at 10% (using two-tailed test statistics and robust standard errors).

Building on this estimation, the GPS is calculated as

$$\hat{R}_i = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{1}{2\hat{\sigma}^2}(T_i - \hat{\beta}_0 - X_i\hat{\beta}_1)^2\right). \tag{6}$$

As stated above, the GPS allows us to remove any bias in the estimate of the dose–response function, $E[Y_i(\tau)]$, if the covariates are sufficiently balanced. That is, Eq. (3) has to be satisfied. Furthermore, focusing on the common-support region between treated and control units in the sample is helpful. This avoids perfect predictability of the treatment intensity given a specific value of the GPS. Within the common-support region, units with a certain treatment intensity and respective propensity scores have counterparts with similar GPS but different treatment intensity. In the following, we illustrate that focusing on the common-support region and controlling for the GPS improves comparability of observations with different treatment intensity tremendously in the data at hand.

To assess the performance of the GPS, Hirano and Imbens (2004) suggest to organize the data in groups of treatment intensity. We chose to discretize the treatment intensity according to the quartiles of the distribution which leaves us with four treatment groups. The first and the third group consist of 520 observations, respectively, while the second and fourth group consist of 519 observations, respectively. As is illustrated in Table 4, these groups differ starkly in the observed covariates. The four columns report t-statistics on whether the mean of each covariate in the respective group is significantly different from the mean of the covariates in the three other groups. According to Table 4, only 8 of the 40 t-values are lower than 1.96. Overall, 80% of the observables display a significant difference between treated units in a given group and control units with a treatment intensity belonging to another group when using two-tailed test statistics and a 5% significance level. The median t-value across all tests is 3.46 and the average mean t-value is 7.76. Accordingly, *ex ante*, the risk of biased causal inference with continuous treatments is particularly large due to such stark differences in observables determining treatment intensity.

Choosing a coarser or finer classification by assigning the observations to fewer or more treatment groups does not affect our results in a decisive way. In Appendix B we report the results for sensitivity checks with three and five treatment groups instead of four groups as used in the main text.

For each treatment group $j \in \{1,2,3,4\}$ we calculate the median treatment intensity T_{Mj} and evaluate the GPS for the whole sample at median treatment intensities. Hence, we calculate $\hat{R}_i(T_M^j, X_i)$ for each group j and each observation

[19] Data on regional infrastructure endowments and, especially, data on environmental characteristics are missing for a number of regions. Hence, the corresponding augmented treatment intensity models in the Appendix exploit variation from a smaller data sample. This limits the scope of conclusions for economic policy in an unnecessary way, since the functional form of the dose–response function is quite similar to the one based on the more parsimonious benchmark specification of regional transfer treatment intensity.

Table 4
Treatment groups and covariates.

Covariates and statistics	Group 1	Group 2	Group 3	Group 4
GDP per-capita	**−22.942**	**−9.585**	**2.803**	**33.683**
(GDP per-capita)^2	**−23.220**	**−9.339**	**3.107**	**33.041**
(GDP per-capita)^3	**−23.462**	**−9.074**	**3.402**	**32.348**
Shapley–Shubik index	**−6.211**	**−3.286**	**2.932**	**6.575**
Budgetary period dummy	**2.384**	0.103	−1.674	−0.810
Border region dummy	**3.123**	1.333	**−1.097**	**−3.361**
Employment	**−5.649**	**−2.473**	**3.316**	**4.796**
Industrial employment	**−7.915**	**−3.464**	**3.463**	**7.919**
Service employment	**−6.053**	**−3.353**	1.757	**7.699**
Population density	**−6.906**	−0.548	0.606	**6.850**
Observations	520	519	520	519
Median t-value	**3.46**			
Mean t-value	**7.79**			

Notes: The groups are generated according to the quartiles of total EU transfers per GDP. t-Values reported in boldface indicate significance at the 5% level. The median and mean t-values are calculated on the basis of the t-statistics across all groups and covariates as reported in the table.

$i=1,\ldots,N$ using the estimates $\hat{\beta}_0,\hat{\beta}_1,\hat{\sigma}^2$ reported in Table 3. We test the common-support condition by plotting the GPS values $\hat{R}_k(T_M^j,X_k)$ where $k\in j$ for observations k being part of group j, against the GPS values $\hat{R}_l(T_M^j,X_l)$ where $l\notin j$ of observations l not belonging to group j. Both the GPS of observations k and the GPS of observations l are evaluated at the median treatment intensity of group j (T_M^j). Only observations $l\notin j$ and observations $k\in j$ featuring GPS values with common support are used for estimation of the dose–response function. Hence, we only use observations l for which

$$\mathrm{Min}\{\hat{R}_k(T_M^j,X_k)\}\leq \hat{R}_l(T_M^j,X_l)\leq \mathrm{Max}\{\hat{R}_k(T_M^j,X_k)\}\ \forall j\in\{1,2,3,4\}$$

holds true, where $k\in j$ and $l\notin j$. Put differently, we require compared observations to display a sufficient degree of similarity in the observable characteristics determining treatment intensity.

The histogram of GPS values evaluated at median treatment intensities of each group are illustrated in Fig. 2, where the yellow bars represent observations of group j and the black bars represent all other observations not belonging to j. We display separate histograms for each group $j\in\{1,2,3,4\}$. As can be seen in Fig. 2, in groups 2 and 3 there are black bars outside the range of the yellow bars, i.e. there are control observations outside the common support. Similarly, in groups 1 and 4, there are control observations outside the common support. This cannot be seen in the figure because the lack of common support occurs in the left half of the lowest bar. In the following analysis, we restrict our sample to observations that satisfy the common-support condition.

Geographically, these observations often turn out to be NUTS3 regions in the new member countries of the EU. Fig. 3 indicates which NUTS3 regions are inside (in white) or outside of the common-support region in a given programming period (in red). Fig. 3 suggests that the regions outside the common support are typically peripheral ones in any of the two budgetary periods. When using the GPS to construct comparable units of observation, we find that there are 1693 of the 2078 region-programming-period units with common support.

After imposing the common-support condition, we check whether the generalized propensity scores achieve a sufficient balancing of covariates and thereby eliminate the selection bias potentially affecting the dose–response function estimates. As explained above, four groups are determined on the basis of the variation in the continuous regional transfer intensity. In addition, we determine 10 blocks *within* each group based on the estimated GPS. We define the blocks for each group $j\in\{1,2,3,4\}$ by the deciles of the GPS evaluated at the median of the group $\hat{R}_k(T_M^j)$ where $k\in j$. Then, we assign each observation $i\in N$ to the respective block according to its GPS evaluated at T_M^j. Note that the blocks are determined for each group separately, and only "treated" observations that are part of the respective group are relevant for the calculation of the deciles. By design, the sum of observations over blocks in a group yields the total number of observations in that group.

Table 5 illustrates the group-and-block structure generated from this algorithm. For instance, the first of the 10 blocks has in total 678 observations of which 40 are located in group 1 and 638 in all other groups together. Taking the sum over all blocks and adding the respective group and control observations yields the total number of 1693 observations in the common-support region. An organization of the data in this way helps identifying comparable observations with the same *predicted* treatment intensity (blocks) but different actual treatment intensity (groups). Following Hirano and Imbens (2004) to test the *balancing property*, we compare observed characteristics of units *within* a specific block of predicted transfer intensity *across* groups of actual treatment intensity. For instance, we compare the 40 observations in cell *group* 1/*block* 1 to the 678 observations in cell *control* 1/*block* 1 and test for equality of covariates. Accordingly, we conduct 10 two-tailed t-tests for *each* group across all covariates. Table 6 reports the mean t-statistics for each group across all covariates, where we weight the t-statistics by the number of observations in the respective block in order to calculate the mean t-statistic.

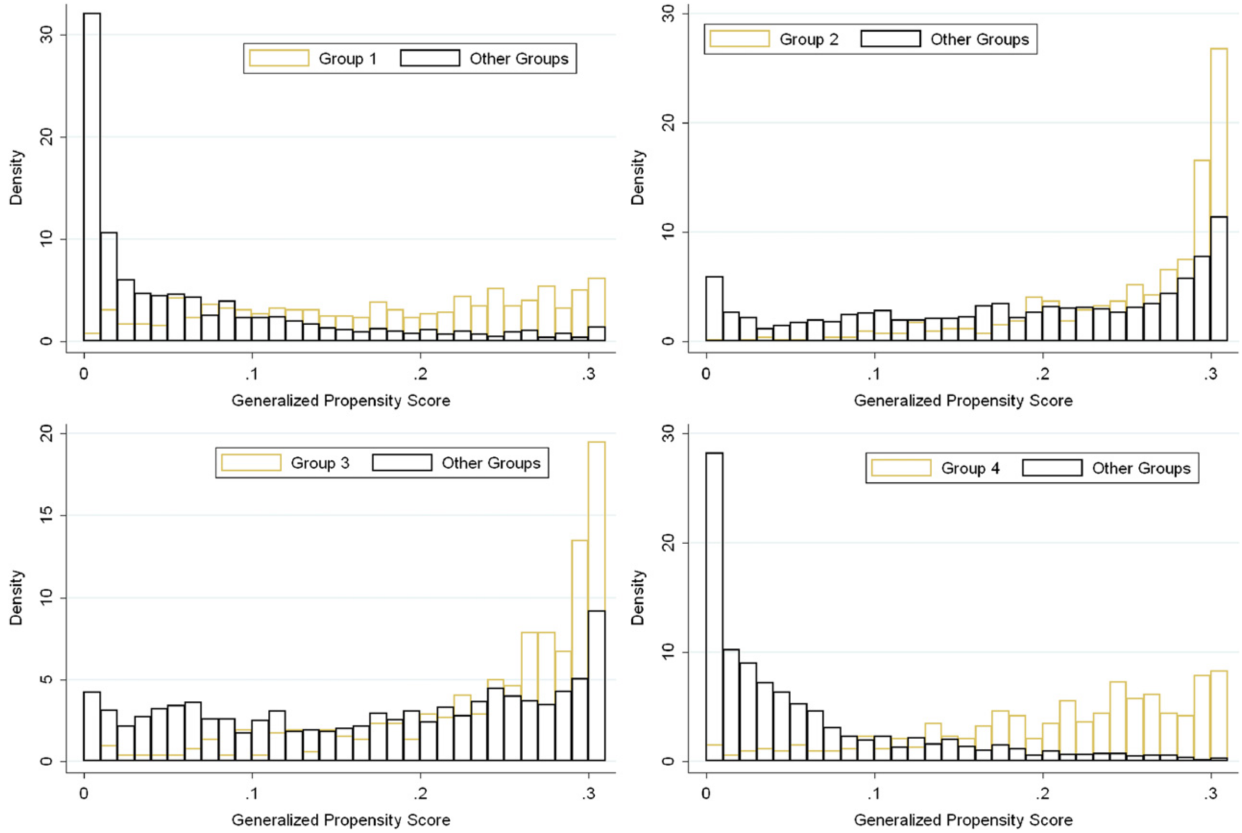

Fig. 2. Common support of the generalized propensity score. *Note*: The groups are generated according to the quartiles of total EU transfers per GDP. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The degree of bias reduction through matching on the GPS is considerable. This can be seen from a comparison of the *t*-values in Table 6 which contrasts units within the support region *after matching* based on the GPS with the respective ones in Table 4 before matching. While the median and average absolute *t*-values were 3.46 and 7.79, respectively, in Table 4, the corresponding values in Table 6 are 0.53 and 0.63, respectively. Before matching, almost all *t*-values were statistically significant while only 2 out of 40 *t*-values remain marginally significant after controlling for the GPS.[20] Accordingly, we argue that the estimated generalized propensity scores perform well in reducing potential treatment-intensity selection bias.

3.3. Estimating the dose–response function

After having largely removed selection bias into different treatment intensities, we can proceed to estimating and visualizing the relationship between regional transfer intensity and regional GDP growth. To do so, the following "second-stage" regression model specifies the conditional expectation of Y_i given T_i and R_i:

$$E[Y_i|T_i,R_i] = \alpha_0+\alpha_1 T_i+\alpha_2 T_i^2+\alpha_3 T_i^3+\alpha_4 \hat{R}_i+\alpha_5 \hat{R}_i^2+\alpha_6 \hat{R}_i^3+\alpha_7 \hat{R}_i T_i \tag{7}$$

using the GPS values estimated in the first stage ($\hat{R}_i$) and the observed treatment intensities (T_i). The parameters are estimated by ordinary least squares, where we implement a block-bootstrap procedure (with 1000 replications) which takes into account that the GPS is not observed but estimated and that some NUTS3 regions are repeatedly observed across programming periods. The GPS terms in the regression are the ones "controlling" for selection into treatment intensities. If selectivity indeed matters, we expect those terms to be jointly statistically significant. In Table 7, we show coefficient estimates from Eq. (7) and find that all GPS-based polynomial terms matter both individually as well as jointly. Hence, GPS

[20] It might be possible to improve the balancing property even further by either using more than 10 blocks or eliminating extreme per-capita income growth rates from the distribution. However, using too many blocks may lead to a small-sample bias of the estimates. We have experimented with dropping units with extreme values of per-capita income growth, but this does not have a visible impact on the estimated non-parametric dose–response function. Hence, to avoid a small-sample bias and an ad-hoc judgment about sample trimming, we decided to use 10 blocks and not drop further observations from the data when assessing the balancing property.

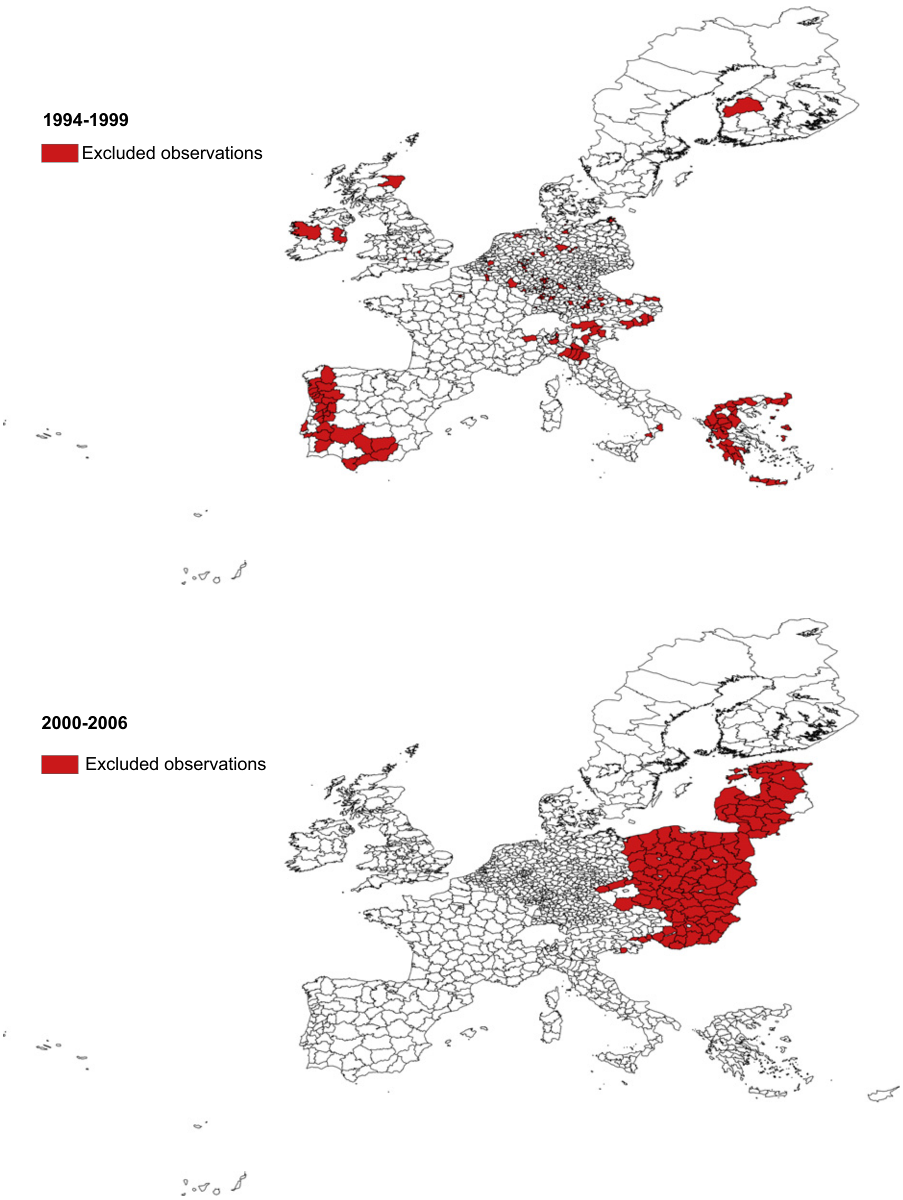

Fig. 3. Observations failing common support restriction. *Note*: The maps indicate the region-period observations that are dropped due to their GPS values lying out of the common support region. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Cell size for comparison of treated and control units in the matrix of 10 blocks and 4 groups.

Block	Group 1	Control 1	Group 2	Control 2	Group 3	Control 3	Group 4	Control 4
1	40	638	49	398	49	336	31	848
2	42	183	47	155	49	183	31	256
3	41	114	48	113	49	121	32	102
4	40	91	50	105	49	147	32	56
5	41	73	48	99	49	73	30	30
6	40	57	47	98	49	63	32	29
7	41	53	50	49	49	89	32	20
8	41	30	48	74	49	52	31	15
9	41	27	48	67	49	67	31	11
10	40	20	48	52	49	72	31	13

Notes: The groups are generated according to the quartiles of EU transfers per GDP whereas the blocks are generated according to the deciles of the GPS evaluated at the median treatment intensity of each group.

Table 6
Balance of covariates accounting for the GPS.

Covariates and statistics	Group 1	Group 2	Group 3	Group 4
GDP per-capita	−0.474	**−1.989**	−1.039	1.006
(GDP per-capita)^2	−0.462	**−1.961**	−1.001	0.981
(GDP per-capita)^3	−0.451	−1.932	−0.962	0.956
Shapley–Shubik index	−0.017	−0.182	0.728	−0.522
Budgetary period dummy	0.547	−0.530	−0.828	0.120
Border region dummy	0.606	−0.104	−0.283	1.253
Employment	0.448	−0.096	0.383	0.486
Industrial employment	0.219	−0.099	0.774	0.687
Service employment	0.405	−0.171	0.059	0.680
Population density	0.124	−0.317	−0 .835	0.558
Observations	407	483	490	313
Median *t*-value	0.53			
Mean *t*-value	0.63			

Notes: The groups are generated according to the quartiles of total EU transfers per GDP. Observations which do not satisfy the common support condition are excluded from the respective groups. In order to control for the GPS values we discretize them into deciles. *t*-Values reported in boldface indicate significance at the 5% level. The median and mean *t*-values are calculated on the basis of the *t*-statistics across all groups and covariates as reported in the table. See main text for details.

estimation is indeed relevant and significantly reduces the bias of the estimated response of per-capita income growth to changes in regional transfer intensity.

With the parameters estimated in the second stage, we can now estimate the average potential outcome at treatment level τ, the so-called *dose–response function*:

$$\widehat{E[Y_\tau]} = \frac{1}{N}\sum_{i=1}^{N}[\hat{\alpha}_0+\hat{\alpha}_1\tau+\hat{\alpha}_2\tau^2+\hat{\alpha}_3\tau^3+\hat{\alpha}_4\hat{R}(\tau,X_i)+\hat{\alpha}_5\hat{R}^2(\tau,X_i)+\hat{\alpha}_6\hat{R}^3(\tau,X_i)+\hat{\alpha}_7\hat{R}(\tau,X_i)\tau]. \tag{8}$$

In addition to the *dose–response function* itself we display its derivative with respect to the regional transfer intensity—which is commonly referred to as the *treatment effect function*. The latter allows us to infer the aforementioned *minimum necessary*, the *optimal*, and the *maximum desirable* treatment intensities of EU regional transfers.

4. Results

4.1. Estimates for total EU regional transfers

The dose–response function based on the GPS is a non-parametric estimate of the functional relationship between per-capita income growth and regional transfer intensity, and so is the treatment effect function. Fig. 4 displays each of those two non-parametric functions in the center as well as the corresponding block-bootstrapped 90% confidence interval. Fig. 4 is obtained for all EU regional transfers at the NUTS3 level under the auspices of the Structural Funds and the Cohesion Fund as in Tables 3–7.

According to the dose–response function in the left panel of Fig. 4, the response of regional per-capita income growth increases monotonically with regional transfer intensity. However, a marginal increase of transfer intensity at a given transfer level does not *necessarily* lead to statistically significantly higher per-capita income growth. This can be seen from the derivative of the dose–response function with respect to transfer intensity in the right panel of Fig. 4. Since the dose–response function is concave, the treatment effect function declines monotonically. The 90% confidence band of the treatment effect function includes zero per-capita income growth at a treatment intensity of about or more than 1.3%. The latter level is indicated by a dotted black bar in the treatment effect plot of Fig. 4. Below this regional transfer intensity level, an increase in regional transfer intensity leads to an unambiguous increase in the per-capita income growth response. NUTS3 regions with a regional transfer intensity of more than 1.3% do no longer unambiguously gain from *additional* EU transfers. In other words, for regions above the 1.3% threshold, a reduction of EU transfers to 1.3% of their GDP would not necessarily harm their growth prospects.

The estimated dose–response function also confirms the results of our previous study where we concluded that transfers under the Objective 1 scheme raised annual growth in the recipient regions on average by about 1.6 percentage points (see Becker et al., 2010). In our data-set, Objective 1 regions received on average EU transfers in the amount of 1.9% of their GDP per annum. At such a transfer intensity, the dose–response function in Fig. 4 predicts an annual growth response of about 5.1% for the average Objective 1 region. The non-Objective 1 regions in our data-set had an average annual growth rate of about 3.6% which yields an average treatment effect of Objective 1 treatment of about 1.5 percentage points. Accordingly, the magnitudes of the average treatment effects as derived from the regression discontinuity design and from the generalized propensity score approach are quite similar. Yet, as argued above, the dose–response function provides insights beyond those of our earlier study which aimed at estimating a homogeneous local average treatment effect.

Table 7
Estimation of the dose–response function.

Covariates and statistics	Coef.	Std. err.
ln(Total EU transfers/GDP)	0.012	0.0007***
ln(Total EU transfers/GDP)^2	0.001	0.0001***
ln(Total EU transfers/GDP)^3	0.00004	4.61e−06***
ln(GPS)	0.001	0.0002***
ln(GPS)^2	0.0005	0.00006***
ln(GPS)^3	0.00003	3.89e−06***
ln(GPS)*ln(Fund/GDP)	9.00e−06	0.00003
Constant	0.084	0.002***
Observations	1693	
R^2	0.11	

Notes: ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. We estimate the dose–response function by blockwise bootstrapping (i.e., drawing from the regional level and then merging respective periods) with 1000 iterations that take into account first-stage estimations.

The results in Fig. 4 also point to the existence of a *maximum desirable* level of regional transfers in terms of target region GDP beyond which the per-capita income growth stimulus becomes unimportant so that additional transfers, on average, are wasted. Of all the 2078 observations receiving transfers in the two considered programming periods (this number includes units within and outside of the common-support region applied in Fig. 4), 1698 display a transfer intensity below the maximum desirable level of 1.3%, and 380 units are treated in excess of 1.3%. The sum of regional transfers to those 380 observations amounted to 148,450.38 mn. Euros. Suppose the European Commission had limited the transfers to those 380 observations to exactly 1.3% of their initial GDP. This would have entailed a reduction of transfers by 32,237.091 mn. Euros in the first programming period and by 31,716.078 mn. Euros in the second programming period. Suppose that the European Commission had used those saved funds in a financially neutral way and spent it in other regions so as to promote aggregate growth in the Union. Ignoring region size, the Commission would then have allocated the saved funds to the regions with a low regional transfer intensity. Suppose the Commission had allocated the funds to the 25% regions with the lowest transfer intensity in each programming period. In 1994–1999 these were 272 regions featuring an average treatment intensity of about 0.014% and in 2000–2006 these were 248 regions featuring an average treatment intensity of about 0.026%. Moreover, assume that the reallocation had been administered so as to provide each of these regions with the same annual transfer intensity after redistribution.[21] Then, the average treatment intensity could have been increased by 0.246 and 0.164 percentage points in the first and in the second programming period, respectively, without any additional funds required.[22] According to our point estimates in Tables 3 and 7 this would have raised annual growth in the average region benefiting from this kind of redistribution by about 1.12 percentage points in the first programming period and by about 0.76 percentage points in the second programming period. Since the reduction of transfers to recipient regions above a transfer intensity of 1.3% should not affect their growth rates in a significant way, this kind of redistribution would have entailed unambiguous efficiency gains.

Another important concept is what we dubbed the *optimal transfer intensity* which was defined as the threshold where an additional Euro transferred yields exactly one Euro of additional GDP in the average recipient region. Accordingly, the optimal transfer intensity has to satisfy the condition

$$\frac{\partial \widehat{E[Y_\tau]}}{\partial \tau}\frac{\partial \tau}{\partial \Im} = \ln(GDP+1) - \ln(GDP) \Leftrightarrow \frac{\partial \widehat{E[Y_\tau]}}{\partial \tau} \approx 0.01, \tag{9}$$

where $\Im$ is the absolute level of transfers, $\tau = (\Im/GDP) \times 100$, and $\partial \widehat{E[Y_\tau]}/\partial \tau$ is the treatment effect function as displayed in the right panel of Fig. 4.[23] If the treatment effect function exceeds 0.01, an additional Euro of transfers boosts GDP in the recipient region by more than one Euro such that a higher level of regional redistribution would benefit the Union's total GDP. On the contrary, if the treatment effect function falls short of 0.01, an additional Euro transferred yields less than a Euro in a recipient region such that the volume of transfers is inefficiently high.

The optimal transfer intensity is indicated by a dotted black bar in right panel of Fig. 4. Across the two periods under consideration the optimal transfer intensity in Fig. 4 amounts to about 0.4% of regional GDP. Note that a transfer intensity

[21] This kind of reallocation generates the biggest possible effect for those regions given that leapfrogging is to be avoided.

[22] In the first programming period, the targeted 272 regions would have received transfers for 6 years. These regions featured an average GDP of 8,230.146 mn. Euros and received average annual transfers of 1.505 mn. Euros. By the mentioned redistribution scheme, the treatment intensity in those regions could have been raised to about $(272 \times 6 \times 1.505\ \text{mn.€} + 32{,}237.091\ \text{mn.€})/(272 \times 6 \times 8{,}230.146\ \text{mn.€}) \times 100 = 0.26\%$ in the first programming period. The 248 targeted regions in the second programming period would have received transfers for 7 years. These regions featured an average GDP of 10,995.18 mn. Euros and received average annual transfers of 2.308 mn. Euros. In those regions, the transfer intensity could have been raised by the mentioned reallocation scheme to about $(248 \times 7 \times 2.308\ \text{mn.€} + 31{,}716.078\ \text{mn.€})/(248 \times 7 \times 10{,}995.18\ \text{mn.€}) \times 100 = 0.19\%$ in the second programming period.

[23] Other things equal, an additional Euro boosts the growth rate by $1/GDP$ and the percentage transfer intensity by $100/GDP$ such that the optimal transfer intensity is reached where the estimated treatment effect function $\partial \widehat{E[Y_\tau]}/\partial \tau$ equals 0.01.

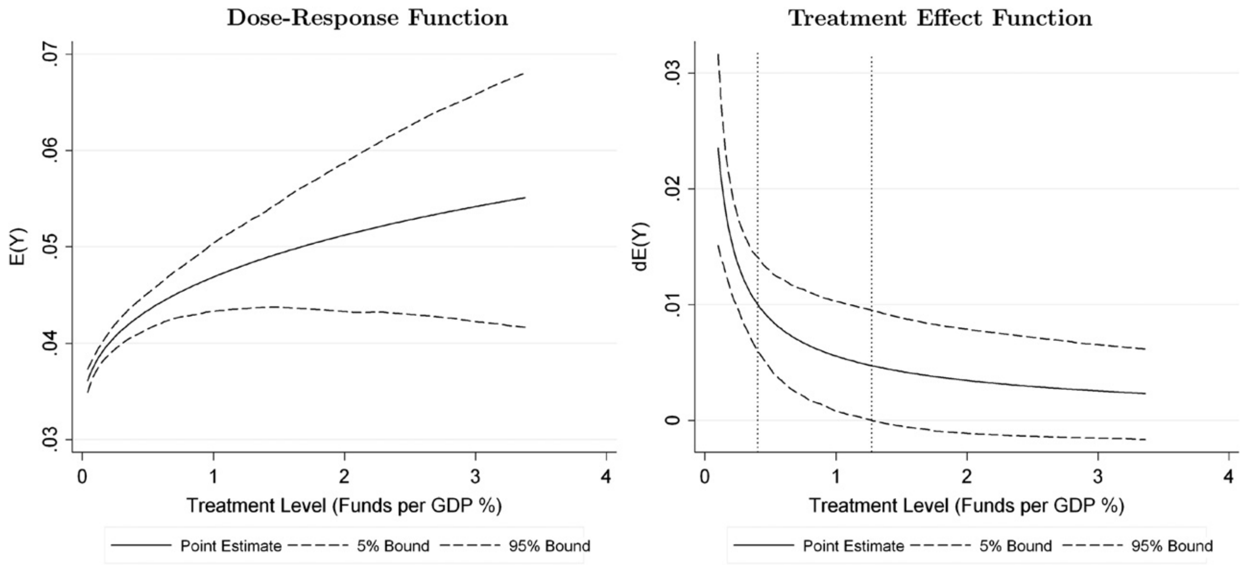

Fig. 4. Effects of total EU transfers. *Note*: Observations with treatment levels in the highest and lowest 5% are trimmed. The dotted bars in the treatment effect function indicate the optimal treatment intensity and the maximum desirable treatment intensity, respectively. The functions in the center of each graph are surrounded by their block-bootstrapped 90% confidence bands.

above the optimum desirable level may still be below the maximum desirable transfer intensity so that a given recipient region with a transfer intensity in that range may still significantly benefit from additional EU transfers.

While the maximum desirable transfer intensity requires only a significant impact on recipient regions, the optimal transfer intensity requires a transfer multiplier above one. Hence, the latter concept is closely linked to aggregate efficiency. Suppose the European Union's single objective had been aggregate growth in the two programming periods under consideration. Then, the Union should have cut transfers to regions with a transfer intensity in excess of 0.4% (344 and 397 regions in the 1994–1999 and 2000–2006 programming periods, respectively) and have raised transfers to regions below the optimal transfer intensity (741 and 596 regions in the 1994–1999 and 2000–2006 programming periods, respectively). Yet, such a policy would have been in conflict with the political goal of regional cohesion, since it would have implied a reallocation of transfers from less developed regions with a high transfer intensity to rather prosperous regions with a low transfer intensity. Such a trade-off between regional cohesion and aggregate efficiency would have been pertinent for 162 regions in the 1994–1999 period and 199 in the 2000–2006 period featuring a transfer intensity above the optimal level but below the maximum desirable level.[24] In any case, according to the reported estimates, cutting transfers to regions beyond the maximum desirable transfer intensity enhances efficiency without harming regional cohesion.

Using NUTS3-level data on the gap between a region's per-capita income level to the (unweighted) average and the transfer intensity together with the estimated treatment function in Fig. 4, we can classify regions along two dimensions. First, regions with a transfer multiplier smaller than unity (i.e., regions to the right of the red dotted line in Fig. 4) and regions with a transfer multiplier of unity or greater than that (i.e., regions at or to the left in Fig. 4). Second, regions with a non-positive per-capita income gap (i.e., ones with a real per-capita income of or below the EU average) and ones with a positive per-capita income gap (i.e., regions with a per-capita income above the unweighted EU average across NUTS3 units). This classification of regions leads to four possible regimes in a given programming period.[25] Applying the estimated treatment effect function to all regions including the ones outside of the common support of the generalized propensity score, we may determine which group a NUTS3 region belongs to.

Fig. 5 illustrates the outcome for the two programming periods NUTS3-level transfer data are available for. In essence, the results from this study suggest that – in pursuit of the two goals of an effective use of funds and the closure of the per-capita income gap within the Union – the European Commission and national governments together might have reduced transfers to regions which are colored light-red and, even more so, dark-red, and have reallocated those transfers to the dark-blue regions. The figures suggest that the two aforementioned goals could have been followed in a better way if – with a few exceptions – transfers had been reallocated from the geographical periphery of the EU towards its core in both

[24] Note also that the value of the treatment intensity where the upper bound of the displayed confidence interval is equal to 0.01 (i.e. the value where the multiplier is unity; see (9)) is below the maximum desirable treatment intensity. The fact that the upper bound of the optimal treatment intensity lies below the maximum desirable treatment intensity indicates that there is, for some regions, a trade-off between aggregate efficiency and regional cohesion (see Martin, 1999 and Boldrin and Canova, 2001 for a theoretical elaboration on this trade-off).

[25] There is a fifth group of NUTS3 regions, namely the ones which did not get any transfers as considered in this study in a given programming period.

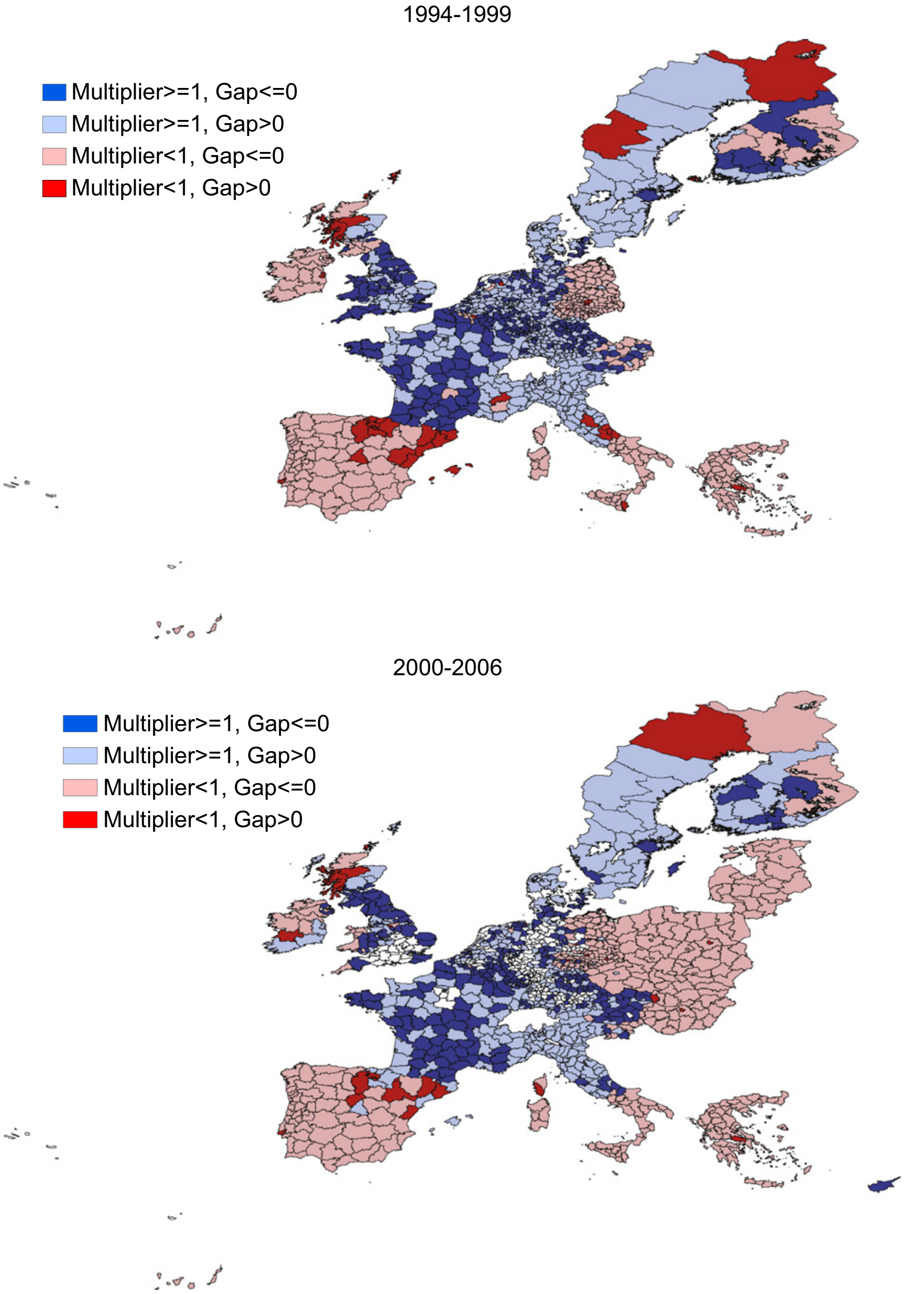

Fig. 5. Transfer multiplier and per-capita GDP gap. *Note*: The regions are assigned to the four groups according to the predicted transfer multiplier and their per-capita GDP gap prior to the respective programming period, i.e. 1993 and 1999. The transfer multiplier is derived using the treatment effect function from Fig. 4 and the respective treatment intensity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the 1994–1999 and the 2000–2006 programming periods. The reason for this outcome is that many of the poorest regions in the Union display a much weaker response to transfers than ones that are closer to but still below the Union's average per-capita income level in the geographical core.

4.2. *Estimates for specific treatments*

We can produce similar estimates for different sub-components of the EU transfer budget. Since Structural Funds transfers account for the lion's share (about 87% on average) of all of the EU's regional transfers, the results for all transfers and Structural Funds transfers alone are very similar. However, we can consider somewhat smaller budgets such as

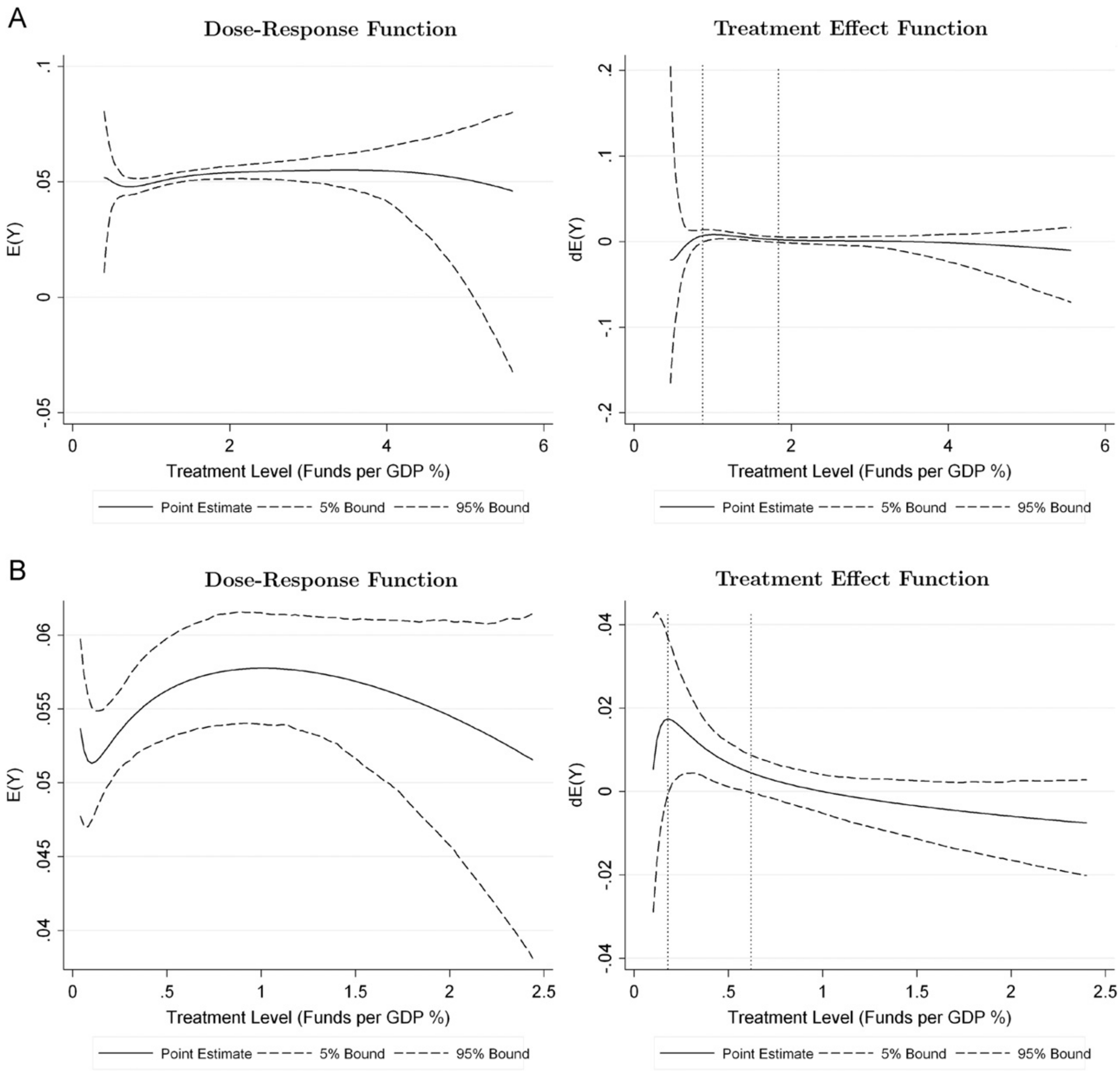

Fig. 6. Effects of (A) Objective 1 and (B) cohesion fund transfers. *Note*: Observations with treatment levels in the highest and lowest 5% are trimmed. The dotted bars in the treatment effect functions indicate the minimum necessary and the maximum desirable treatment intensities. The functions in the center of each graph are surrounded by their block-bootstrapped 90% confidence bands.

transfers to Objective 1 regions (which account for about 74% of all EU-administered regional transfers) and, alternatively, for Cohesion Fund transfers (about 13% of total transfers). Again, we can estimate the dose–response function and the treatment effect function.[26]

Panels A and B in Fig. 6 summarize the results for transfers to Objective 1 regions and Cohesion fund transfers, respectively, akin to Fig. 4 for all transfers. Either one of the two figures displays a similar pattern. First of all, neither the dose–response function nor the treatment effect function is monotonic but hump-shaped. In particular, the confidence bands of the treatment effect function cross the abscissa twice. Hence, the figures suggest that there is a minimum necessary level of transfer treatment in the two sub-categories and a maximum desirable level. However, one reason for the existence of the former is that the number of observations with a very low treatment intensity is relatively small and the estimated variance in response is relatively large for those units. Hence, the statistical evidence of existence of a

[26] Obviously, the validity of GPS estimation again depends on balancing of the covariates as with all regional transfers. For the sake of brevity, we suppress the documentation of balancing here, but results are available from the authors upon request. It turns out that the balancing property tests are as successfully met as in the case of all EU regional transfers combined.

maximum desirable treatment level is stronger than the one of a minimum necessary one. According to Fig. 6, the maximum desirable treatment threshold is at about 1.8% for Objective 1 regional transfers and at about 0.61% for Cohesion Fund transfers.

5. Conclusions

This paper focuses on the estimation of the response of average annual GDP per-capita growth to changes in the intensity of regional transfers provided by the European Commission under the auspices of the Structural Funds and Cohesion Fund programmes. We use NUTS3 data, the most disaggregated regional data available, covering the two budgetary periods 1994–1999 and 2000–2006. Non-parametric generalized propensity score analysis allows us to estimate the causal effect of different levels of EU transfers on regional per-capita income growth.

Our results point to an optimal transfer intensity of 0.4% of target region GDP and a maximum desirable intensity of 1.3%. Additional transfers to regions below a transfer intensity of 0.4% enhance aggregate efficiency as they exhibit a multiplier above one. Regions with an EU transfer intensity below 1.3% of their beginning-of-period GDP could grow faster in response to additional EU transfers. Regions with a transfer intensity of more than 1.3% of GDP could give up EU transfers without experiencing a significant drop in their average annual per-capita income growth rate. For a certain range of transfer intensities, we detect a trade-off between aggregate efficiency and regional cohesion. Reducing the transfers to regions below the maximum desirable transfer intensity significantly harms their growth prospects but may enhance aggregate efficiency, if the transfer intensity is above the optimal level. A reallocation of EU transfers from the 18% of regions that received more than 1.3% of their initial GDP as EU transfers to regions below that threshold would have been efficient and could have boosted regional convergence even further in the two considered programming periods.

Acknowledgments

We gratefully acknowledge numerous useful comments on an earlier version of the manuscript by two anonymous referees and the editor in charge (Theo Eicher). Furthermore, we thank Serguei Kaniovski and Rafal Raciborski for providing us with data. We are indebted to seminar participants at various seminars and conferences. S. Becker gratefully acknowledges financial support from the Scottish Institute for Research in Economics (SIRE). M. von Ehrlich gratefully acknowledges financial support from the German Science Foundation (DFG).

Appendix A. Infrastructure endowments and environmental hazards as additional determinants of treatment intensity

There are good reasons to expect the transfer allocation within the EU to depend also on regional infrastructure endowments and on local environmental issues as the Structural Funds and the Cohesion Fund finance many infrastructure and environmental projects. We aim at capturing these effects by the regional road density and by an index that reflects environmental hazards in an augmented empirical analysis. The data on road endowments were provided by the Office for Regional Science, Planning and Geographical Information (RRG). We include the cumulative length of roads in the respective NUTS3 regions (in kilometers) prior to each programming period, i.e., 1993 and 1999, weighted by the total area (in square kilometers) in addition to the observable covariates in X_i. The environmental hazards index stems from the ESPON regional database where we chose the index capturing all weighted environmental hazard values. The hazard dimensions include snow avalanches, droughts, earthquakes, extreme temperatures, floods, forest fire, landslides, the occurrence of storm surges, the tsunami potential, the risk of volcanic eruptions, winter storms and tropical storms, the air traffics hazard potential, the risk from chemical plants, the risk of radioactive contamination, and oil spills.

The infrastructure variable as well as the environmental hazard index turn out to vary significantly across treatment groups which suggests that they indeed affect the transfer treatment intensity. Including these variables also raises the explanatory power of the first-stage regression from an R^2 of 56% to one of 58%. After imposing the common support condition and matching on the GPS we are still able to reach a sufficient balancing of covariates and thereby eliminate the selection bias. While 39 out of 48 t-tests on mean differences of covariates indicate significant differences across the four treatment groups before matching on the GPS, none of the t-tests remains significant after controlling for the GPS. The average t-statistic drops from 5 to 0.4 after controlling for the GPS.[27] Again, we follow the procedure as outlined in Section 3.3 for estimating the dose–response function and the treatment effect function for the augmented model. Due to missing data on the infrastructure and environmental variables the augmented second-stage regression is based on only 1332 observations. Yet, despite the considerably smaller sample with common support, the point estimates as well as the confidence intervals of the dose–response as well as the treatment effect functions do not change much compared to our benchmark estimation (see Fig. A1). Provided that the results are qualitatively and quantitatively similar to the benchmark results in the main text, we consider the main results preferable since they are based on a substantially larger sample of observations.

[27] The corresponding tables are available upon request.

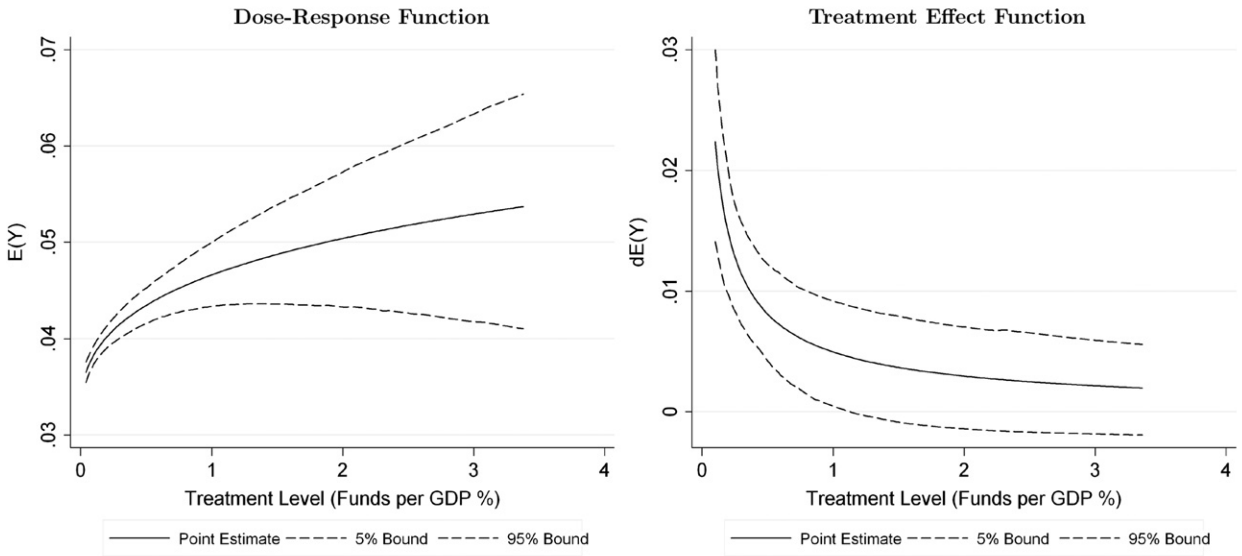

Fig. A1. Effects of total EU transfers accounting for infrastructure endowments and environmental hazards. *Note*: Observations with treatment levels in the highest and lowest 5% are trimmed. The functions in the center of each graph are surrounded by their block-bootstrapped 90% confidence bands.

Table B1
Treatment groups and covariates (3 groups).

Covariates and statistics	Group 1	Group 2	Group 3
GDP per-capita	−**26.660**	−**3.986**	**33.481**
(GDP per-capita)^2	−**26.852**	−**3.643**	**33.070**
(GDP per-capita)^3	−**26.993**	−**3.296**	**32.595**
Shapley–Shubik index	−**5.079**	−**2.166**	**7.303**
Budgetary period dummy	1.226	**2.115**	−**3.346**
Border region dummy	**3.315**	−0.054	−**3.260**
Employment	−**7.557**	**2.930**	**4.555**
Industrial employment	−**9.352**	1.913	**7.344**
Service employment	−**8.468**	1.476	**6.929**
Population density	−**6.590**	0.254	**6.328**
Observations	693	692	693
Median *t*-value	**4.82**		
Mean *t*-value	**9.40**		

Notes: The groups are generated according to the terciles of total EU transfers per GDP. *t*-Values reported in boldface indicate significance at the 5% level. The median and mean *t*-values are calculated on the basis of the *t*-statistics across all groups and covariates as reported in the table.

Appendix B. Sensitivity with respect to the chosen number of treatment groups

In Section 3.2, we assigned the observations to four treatment groups that were generated according to the quartiles of the distribution of EU transfer intensity in terms of regional GDP (see Table B3). In general, there is a trade-off between the coarseness of the classification and the violation of the balancing property. The coarser the classification, the more likely will the balancing property be violated (i.e., the less comparable are treated and control units), but the more observations will have common support. On the contrary, the finer the classification, the less likely will the balancing property be violated, but the less observations will have common support. Our results appear quite robust to the chosen number of treatment groups.

Choosing a coarser classification than for the benchmark results with three treatment groups according to the terciles of the treatment distribution amplifies the differences in covariates across groups compared to the classification with four treatment groups. The average *t*-statistic in the balancing score tests increases from 7.79 to 9.4 (see Tables 4 and B1). However, conditioning on the GPS still renders the differences across treatment groups insignificant as is obvious from Table B2. Table B3 provides the distribution of observations across the cells underlying Table B2.

Choosing a finer classification than for the benchmark results with five treatment groups of approximately the same size reduces the differences in covariates across treatment groups as can be seen from comparing Tables 4 and B4 and Tables 6 and B5 (see Table B6 for the distribution of observations across the cells underlying Tables B4 and B5).

Table B2
Balance of covariates accounting for the GPS (three groups).

Covariates and statistics	Group 1	Group 2	Group 3
GDP per-capita	−1.390	−0.982	0.534
(GDP per-capita)^2	−1.359	−0.960	0.527
(GDP per-capita)^3	−1.329	−0.938	0.520
Shapley–Shubik index	1.229	−0.143	−0.871
Budgetary period dummy	−0.470	0.319	−0.070
Border region dummy	0.394	−0 .597	1.195
Employment	0.224	0.794	−0.713
Industrial employment	0.264	0.786	−0.401
Service employment	0.046	0.711	−0.688
Population density	00.286	−0 0.439	−0.291
Observations	578	663	471
Median t-value	0.57		
Mean t-value	0.65		

Notes: The groups are generated according to the terciles of total EU transfers per GDP. Observations which do not satisfy the common support condition are excluded from the respective groups. In order to control for the GPS values we discretize them into deciles. t-Values reported in boldface indicate significance at the 5% level. The median and mean t-values are calculated on the basis of the t-statistics across all groups and covariates as reported in the table. See main text for details.

Table B3
Cell size for comparison of treated and control units in the matrix of 10 blocks and 3 groups.

Block	Group 1	Control 1	Group 2	Control 2	Group 3	Control 3
1	57	566	67	262	47	670
2	59	166	66	178	48	263
3	58	108	66	114	47	119
4	58	68	67	109	47	71
5	57	57	66	96	47	46
6	58	66	66	66	47	21
7	58	33	66	53	47	19
8	58	35	67	72	47	13
9	58	19	66	50	47	11
10	57	16	66	49	47	8

Notes: The groups are generated according to the terciles of EU transfers per GDP whereas the blocks are generated according to the deciles of the GPS evaluated at the median treatment intensity of each group.

Table B4
Treatment groups and covariates (five groups).

Covariates and statistics	Group 1	Group 2	Group 3	Group 4	Group 5
GDP per-capita	**−21.184**	**−11.046**	**−3.072**	**7.816**	**30.424**
(GDP per-capita)^2	**−21.497**	**−10.874**	**−2.818**	**8.010**	**29.801**
(GDP per-capita)^3	**−21.780**	**−10.682**	**−2.562**	**8.185**	**29.139**
Shapley-Shubik index	**−8.055**	0.591	**−4.645**	**3.388**	**8.731**
Budgetary period dummy	**2.564**	−1.396	**3.337**	**−3.268**	−1.230
Border region dummy	**3.618**	0.373	1.072	−0.831	**−4.236**
Employment	**−3.922**	**−5.476**	0.793	**3.781**	**4.815**
Industrial employment	**−6.442**	**−5.373**	−0.343	**3.935**	**8.272**
Service employment	**−4.079**	**−6.656**	−0.026	**2.878**	**7.925**
Population density	**−6.761**	−1.425	−0.638	**2.287**	**6.528**
Observations	415	417	415	415	416
Median t-value	**4.16**				
Mean t-value	**6.97**				

Notes: The groups are generated according to the quintiles of total EU transfers per GDP. t-Values reported in boldface indicate significance at the 5% level. The median and mean t-values are calculated on the basis of the t-statistics across all groups and covariates as reported in the table.

The variation of the common support sample with respect to the number of treatment groups is only minor: while the benchmark specification's common support sample contains 1693 observations the specifications with three and five treatment groups yield common support samples of 1712 and 1599 observations, respectively. Accordingly, the estimates

Table B5
Balance of covariates accounting for the GPS (five groups).

Covariates and statistics	Group 1	Group 2	Group 3	Group 4	Group 5
GDP per-capita	0.533	−1.888	−1.598	−0.948	0.853
(GDP per-capita)^2	0.532	−1.840	−1.594	−0.935	0.851
(GDP per-capita)^3	0.530	−1.791	−1.590	−0.923	0.850
Shapley–Shubik index	−1.343	1.768	−1.295	0.396	−0.160
Budgetary period dummy	0.584	−1.197	0.834	−1.094	0.402
Border region dummy	0.732	−0.033	−0.165	−0.003	1.040
Employment	1.290	−0.508	−0.082	0.148	0.756
Industrial employment	0.882	−0.105	−0.331	0.653	0.955
Service employment	1.310	−0.827	−0.049	−0.149	1.001
Population density	0.334	−0.441	−1.234	−0.402	1.168
Observations	260	349	390	378	221
Median *t*-value	0.84				
Mean *t*-value	0.82				

Notes: The groups are generated according to the quintiles of total EU transfers per GDP. Observations which do not satisfy the common support condition are excluded from the respective groups. In order to control for the GPS values we discretize them into deciles. *t*-Values reported in boldface indicate significance at the 5% level. The median and mean *t*-values are calculated on the basis of the *t*-statistics across all groups and covariates as reported in the table. See main text for details.

Table B6
Cell size for comparison of treated and control units in the matrix of 10 blocks and 5 groups.

Block	Group 1	Control 1	Group 2	Control 2	Group 3	Control 3	Group 4	Control 4	Group 5	Control 5
1	26	670	35	407	39	377	37	408	23	849
2	27	180	35	162	39	170	39	234	22	199
3	26	117	35	135	39	128	38	167	22	137
4	26	106	35	106	39	100	37	97	21	50
5	26	66	35	80	39	109	38	42	22	49
6	26	79	35	79	39	74	38	61	23	14
7	25	36	35	66	39	83	38	49	22	33
8	27	44	35	73	39	62	38	46	21	17
9	25	20	35	55	39	48	38	73	23	13
10	26	20	34	87	39	58	37	44	22	17

Notes: The groups are generated according to the quintiles of EU transfers per GDP whereas the blocks are generated according to the deciles of the GPS evaluated at the median treatment intensity of each group. See main text for details.

of the second-stage model which underlies the dose–response function and the treatment effect function estimates remain almost unaffected.[28]

References

Becker, S.O., Muendler, M.-A., 2008. The effect of FDI on job security. B.E. Journal of Economic Analysis & Policy (Advances) 8.
Becker, S.O., Egger, P.H., von Ehrlich, M., 2010. Going NUTS: the effect of EU structural funds on regional performance. Journal of Public Economics 94, 578–590.
Becker, S.O., Egger, P.H., von Ehrlich, M., Fenge, R., 2008. Going NUTS: The Effect of EU Structural Funds on Regional Performance. CESifo Working Paper No. 2495, November.
Boldrin, M., Canova, F., 2001. Europe's regions—income disparities and regional policies. Economic Policy 32, 207–253.
Dehejia, R., Wahba, S., 1999. Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. Journal of the American Statistical Association 94, 1053–1062.
European Commission, 1989. ERDF in Figures 1988. Commission of the European Communities, Brussels.
European Commission, 2008. Regional Policy—Inforegio. 〈http://ec.europa.eu/regional_policy/policy/fonds/index_en.htm〉.
Felsenthal, D., Machover, M., 1998. The weighted voting rule in the EU's council of ministers, 1958–95: intentions and outcomes. Electoral Studies 16, 33–47.
Hagen, T., Mohl, P., 2008. Which is the Right Dose of EU Cohesion Policy for Economic Growth? ZEW Discussion Paper No. 08-104, December.
Heckman, J.J., Ichimura, H., Todd, P., 1997. Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. Review of Economic Studies 64, 605–654.
Hirano, K., Imbens, G.W., 2004. The propensity score with continuous treatments. In: Andrew, G., Meng, X.-L. (Eds.), Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives, Wiley. (Chapter 7).
Hirshleifer, J., 1958. On the theory of optimal investment decision. Journal of Political Economy 66, 329–352.

[28] The corresponding figures are available upon request.

Imai, K., Van Dijk, D.A., 2004. Causal inference with general treatment regimes: generalizing the propensity score. Journal of the American Statistical Association 99, 854–866.
Imbens, G.W., 2000. The role of the propensity score in estimating dose–response functions. Biometrika 87, 706–710.
Kluve, J., Schneider, H., Uhlendorff, A., Zhao, Z. Evaluating continuous training programs using the generalized propensity score. Journal of the Royal Statistical Society, Series A., forthcoming.
Martin, P., 1999. Public policies, regional inequalities and growth. Journal of Public Economics 73, 85–105.
Murphy, K.M., Shleifer, A., Vishny, R.M., 1989. Industrialization and the big push. Quarterly Journal of Economics 106, 503–530.
Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.
Sachs, J.D., McArthur, J.W., Schmidt-Traub, G., Kruk, M., Bahadur, C., Faye, M., McCord, G., 2004. Ending Africa's poverty trap. Brookings Papers on Economic Activity 2004, 117–216.
Shapley, L.S., Shubik, M., 1954. A method for evaluating the distribution of power in a committee system. American Political Science Review 48, 787–792.
Widgren, M., 2009. The impact of council voting rules on EU decision-making. CESifo Economic Studies 55, 30–56.