Munich Personal RePEc Archive

A conditional directional distance function approach for measuring regional environmental efficiency: Evidence from the UK regions

George Halkos and Nickolaos Tzeremes

University of Thessaly, Department of Economics

April 2012

Online at https://mpra.ub.uni-muenchen.de/38147/
MPRA Paper No. 38147, posted 16. April 2012 18:54 UTC

A conditional directional distance function approach for measuring regional environmental efficiency: Evidence from the UK regions

By

George E. Halkos* and Nickolaos G. Tzeremes

University of Thessaly, Department of Economics,

Korai 43, 38333, Volos, Greece

Abstract

This paper, by using conditional directional distance functions as introduced by Simar and Vanhems [J. Econometrics 166 (2012) 342-354] modifies the model by Färe and Grosskopf [Eur. J. Operat. Res. 157 (2004) 242-245], examines the link between regional environmental efficiency and economic growth. The proposed model using conditional directional distance functions incorporates the effect of regional economic growth on regions' environmental efficiency levels. The results from the UK regional data reveal that economic growth has a negative effect on regions' environmental performance up to a certain GDP per capita level, where after that point the effect becomes positive. This indicates the existence of a Kuznets type relationship between the UK regions' environmental performance and economic growth.

JEL classification: C14; C6; Q5; R1

Keywords: Regional environmental performance; Directional distance function; Conditional measures; UK regions.

* **Address for Correspondence:** Professor George Halkos, Department of Economics, University of Thessaly, Korai 43, 38333, Volos, Greece. Email: halkos@econ.uth.gr, http://www.halkos.gr/ Tel.: 0030 24210 74920 FAX : 0030 24210 74701

1. Introduction

The measurement of environmental technology in Data Envelopment Analysis (DEA) literature has been an open challenge for researchers. The problem lies on the treatment of the pollutant[1] in a production function framework. The tradeoff between environmental quality and economic development has been firstly modeled by Färe et al. (1989) with the use of distance functions in a nonparametric setting. It was the first nonparametric model measuring environmental technology in a production function framework.

In addition the model introduced by Färe et al. (1989) has treated pollutant as output of the production process and by imposing strong and weak disposability developed environmental performance indicators (hereafter EPIs)[2]. Later, Tyteca (1997) introduced another EPI based on the same principles as Färe et al. (1989) but with different assumptions. Since then, the construction of EPIs has been introduced by several papers that incorporate them into their analysis.

Moreover, Chung et al. (1997) using the weak disposability assumption of outputs constructed a Malmquist–Luenberger index, creating for the first time environmental productivity indexes. The original work of Färe et al. (1989) assumed strong (for desirable outputs) and weak (for undesirable outputs) disposability treating environmental impacts as undesirable outputs in a hyperbolic efficiency measure. Generally the property of weak disposability of detrimental variables is well known and has been used in

[1] The pollutant is also referred to the literature of measuring environmental technology as 'bad' output.
[2] Other studies treat the pollutant as input in a DEA framework (Reinhard et al., 2000, Korhonen and Luptacik, 2004).

several formulations (Färe et al., 1996, 2004; Chung et al., 1997; Tyteca, 1996, 1997; Zofio and Prieto 2001; Zhou et al., 2006, 2007)[3].

Another well known treatment of bad outputs when measuring environmental performance in DEA setting is the one introduced by Seiford and Zhu (2002). They developed a radial DEA model, in order to improve efficiency via increasing desirable and decreasing undesirable outputs. They have introduced a linear monotone decreasing transformation and thus undesirable outputs can be treated as desirable.

However, Färe and Grosskopf (2004) commented on that transformation claiming that the transformation proposed provides different efficiency results due to the fact that it does not resort to ad hoc treatment of undesirable outputs as inputs (as a result of the imposition of strong disposability assumption for all outputs). Furthermore, Färe and Grosskopf suggested an alternative approach based on directional output distance function. Later, Seiford and Zhu (2005) replied to the critic made proposing that the model based on directional output distance function is very similar to the weighted additive model (Ali et al., 1995; Thrall, 1996; Seiford and Zhu, 1998) where the bad outputs are treated as controllable inputs.

Several scholars following the modeling principle by Färe et al. (1989), for country level studies have examined the relationship between economic growth and environmental performance (Zaim and Taskin, 2000a, 2000b, 2000c; Taskin and Zaim, 2001; Zofio and Prieto, 2001; Zaim, 2004; Managi, 2006; Yörük and Zaim, 2006; Picazo-Tadeo and García-Reche, 2007; Halkos

[3] This approach is widely accepted among the environmental economists, however, several remarks have been raised regarding the 'operationalization of weak disposability in empirical production analysis' (Hailu and Veeman, 2001; Färe and Grosskopf, 2003; Hailu, 2003; Kuosmanen, 2005; Färe and Grosskopf, 2009; Kuosmanen and Podinovski, 2009).

and Tzeremes, 2009). These studies are based on the works of Selden and Song (1994) and Grossman and Krueger (1995) which have found an inverted U-type (Environmental Kuznets Curve-EKC)[4] relationship between economic activity and environmental quality.

Additionally the pre-mentioned DEA studies relating the link between environmental performance and economic activity have used country level data. In their formulation they have used GDP per capita (as a proxy of countries' economic growth) as independent variable and EPI as dependent variable. In this way, the existence of a Kuznets type relationship was examined in a second stage panel data econometric analysis[5]. In general the two-stage analysis of DEA efficiency scores has been very popular among the researchers[6]. However, as has been critically stated by Simar and Wilson (2011) several assumptions regarding the data generating process (most of the times unsupported by economic data) are needed in order for the researchers to perform second-stage regressions involving DEA efficiency scores.

Simar and Wilson (2007) provided an alternative model which involves bootstrap algorithm alongside with a truncated regression, which provides a consistent estimation when analyzing DEA efficiency scores. Another approach for explaining the efficiency scores is the one introduced by Daraio and Simar (2005, 2006, 2007) which is based on conditional measures of a probabilistic approach of efficiency measures. In addition one of the main

---

[4] Kuznets (1955) showed that income disparities first rise and then begin to fall during economic development stages, many studies tried to link a similar type relationship between economic growth (in per capita terms) and environmental degradation/performance.

[5] Most of the studies have used fixed and random effect models missing dynamic effects which can be revealed with the application of the generalized method of moments (GMM) estimators (Managi, 2006; Managi and Jena, 2008; Halkos and Tzeremes, 2009)

[6] For critical discussion for two-stage DEA analysis see Banker and Morey (1986), Hoff (2007), Simar and Wilson (2007), Banker and Natarajan (2008), Park et al. (2008), McDonald (2009) and Simar and Wilson (2011).

advantages of this approach is that does not require a ristrictive 'separability' condition between the input–output space and the space of exogenous environmental factors[7].

Recently, Simar and Vanhems (2012) based on the probabilistic formulation of the production process introduced by Cazals et al. (2002) and Daraio and Simar (2005), defined for the first time conditional directional distance functions and their nonparametric estimators, where conditioning was on environmental factors that may influence the production process[8]. Based on the work of Simar and Vanhems (2012) our paper extents Färe and Grosskopf 's (2004) directional distance function model incorporating bad outputs in order to account for the effect of economic growth. More specifically, we propose a conditional distance function model with the treatment of bad outputs in productivity analysis, which is conditioned on the effect of economic growth. As a result we will be able to model the effect of economic growth on environmental performance avoiding all the 'unrealistic' assumptions involved in most of the two-stage DEA formulations (Simar and Wilson, 2007, 2011).

Finally, as an illustrative example we use NUTS 2 level data from the UK regions in order to examine the link of environmental performance-economic growth relationship.

---

[7] One of the most unrealistic assumptions of the two-stage DEA studies is the requirement of the separability condition between the input–output space and the space of the exogenous factors, assuming that these factors have no influence on the attainable set, affecting only the probability of being more or less efficient (Bădin et al. 2010, p.634).

[8] On their paper Simar and Vanhems (2012) show how directional distance functions can be expressed on radial and hyperbolic measures (therefore, negative values can be included in the formulation) and they were also defined conditional and unconditional directional distance functions (also for α-quantile or order-m partial frontiers) both for radial and hyperbolic measures.

2. Data and variables

To our knowledge few studies have examined regions' environmental efficiency levels. Most of them concentrated in the regions of China (Watanabe and Tanaka, 2007; Bian and Yang, 2010; Guo et al., 2011; Shi et al., 2010; Wang et al., 2012). In addition Mandal and Madheswaran (2010) measured regional environmental efficiency for 20 Indian states in terms of cement production, whereas Macpherson et al. (2010) used a directional distance function approach in order to measure regional environmental-economic assessments for the case of Mid-Atlantic region of the USA. Finally, for an EU region, the first study that developed regional environmental performance indicators is the one by Halkos and Tzeremes (2012). They have measured German regions' environmental efficiency by using Kuosmanen (2005) technology in a directional distance function approach for modeling municipality wastes.

In our analysis we are using regional data collected from two different regional databases (EUROSTAT[9] and OECD[10]) for the year 2007. Most of the studies measuring regional environmental efficiencies analyze administrative regions (in NUTS 2 level) in order to grasp the effect of regional regulatory environmental style within the countries (Knill and Lenschow, 1998, Halkos and Tzeremes 2011, 2012). Similarly, our analysis is referring to NUTS 2 level for 37 U.K. regions[11].

Based on several other studies similar to ours (Färe et al., 1989, 1996, 2004; Färe and Grosskopf, 2003, 2004; Chung et al., 1997; Tyteca, 1996, 1997;

[9] http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/regional_statistics/data/main_tables.
[10] http://stats.oecd.org/Index.aspx?DataSetCode=REG_LAB_TL3.
[11] http://en.wikipedia.org/wiki/NUTS_of_the_United_Kingdom.

Taskin and Zaim, 2001; Zofio and Prieto, 2001; Zaim, 2004; Managi, 2006; Yörük and Zaim, 2006; Picazo-Tadeo and García-Reche, 2007; Halkos and Tzeremes, 2009) in order to model regional environmental efficiency we are using two inputs. These are the total regional labour force (employed people-all NACE activities in thousands) and regional capital stock (millions of euros). Regional capital stock for the year 2007 is not available; therefore we have calculated it following the perpetual inventory method (Feldstein and Foot, 1971; Epstein and Denny, 1980) as:

$$K_t = I_t + (1-\delta)K_{t-1} \tag{1}$$

where K_t is the regional gross capital stock in current year; K_{t-1} is the regional gross capital stock in the previous year; I_t is the regional gross fixed capital formation and δ represents the depreciation rate of capital stock. In our study, following Zhang et al. (2011), we set δ to 6%.

Furthermore our study uses regional gross domestic product (million PPS) as good output and three greenhouse gases (GHGs) as bad outputs (realised from all NACE activities). More analytically we use data from the European Environmental Agency[12] and are referring to the regional quantities of carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) measured in metric tones. Greenhouse gases (GHGs) include carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O) as well as high Global Warming Potential gases such as hydrofluorocarbons (HFCs), perfluorocarbons (PFCs) and sulphur hexafluoride (SF_6). CO_2 emissions from the burning of fossil fuels and the change in the use of human land are considered as the most important

[12] http://prtr.ec.europa.eu

anthropogenic sources. Methane and nitrous oxide are naturally present in the atmosphere. Methane is caused by emissions from landfills, livestock, rice farming and fertilizers. These three gases are among the most significant GHGs.

Then in our second stage analysis and in order to test the link between regional environmental efficiency and regional economic growth, we follow several other regional studies (He, 2008; Diao et al., 2009; Brajer et al., 2011) using regional GDP per capita (GDPPC) (measured in euro) as a proxy of regional economic growth. Table 1 presents the descriptive statistics of the variables used. As can be realized there are a lot of disparities among the thirty seven regions of our analysis.

Table 1: Descriptive statistics of the variables used

	Inputs		Good output	Exogenous variable
	Capital Stock	*Labour force*	*Current GDP*	*GDPPC*
Min	8607011.688	234.300	11142.000	21200.000
Max	10375626.651	2772.800	290091.000	96600.000
Mean	9443410.963	831.249	55482.351	31570.270
Std	527350.090	513.995	48346.944	12218.494
	Bad Outputs			
	CH_4	CO_2	N_2O	
Min	1440.000	121000.000	12.900	
Max	49168.000	173222000.000	6748.000	
Mean	14222.000	11177540.541	420.668	
Std	10761.241	28643694.624	1179.242	

3. Modeling regional environmental efficiency

3.1 Directional distance functions for measuring regional environmental efficiency

Following the model proposed by Färe and Grosskopf (2004) we let $P(x)$ to denote an input vector $x \in \Re_{+}^{N}$ which can produce a set of undesirable outputs $u \in \Re_{+}^{K}$ and desirable outputs $v \in \Re_{+}^{M}$. Then in order to determine the environmental technology several assumptions are needed to be taken following Shephard (1970), Färe and Primont (1995). We assume that the output sets are closed and bounded and that inputs are freely disposal. In addition $P(x)$ can be an environmental output set if:

1. $(v,u) \in P(x)$ and $0 \leq \theta \leq 1$ then $(\theta v, \theta u) \in P(x)$ (i.e. the outputs are weakly disposable) and

2. $(v,u) \in P(x)$, $u = 0$ implies that $v = 0$ (i.e. the null jointness assumption of good and bad outputs).

The weak disposability assumption implies that the reduction of bad outputs is costly and therefore the reduction of bad outputs can be obtained only by a simultaneously reduction of good outputs. In addition the assumption which indicates that the good outputs are null-joint with bad outputs implies that the bad outputs are byproducts of the production process when producing good outputs. In order to formalize the environmental technology we use the data envelopment analysis (DEA) framework.

Let $k = 1,\ldots,K$ be the observations and then the environmental output can be formalized as:

$$
\begin{aligned}
P(x) = \Big\{ (v,u) : & \sum_{k=1}^{K} \omega_k v_{km} \geq v_m, m = 1,\dots,M, \\
& \sum_{k=1}^{K} \omega_k u_{kj} = u_j, j = 1,\dots,J, \\
& \sum_{k=1}^{K} \omega_k x_{kn} \leq x_n, n = 1,\dots,N, \\
& \omega_k \geq 0, k = 1,\dots,K \Big\}
\end{aligned} \tag{1}
$$

$\omega_k, k = 1,\dots,K$ indicate the intensity variables which are not negative and imply constant return to scale[13]. The inequality on the good outputs and the equality on the bad outputs help us to impose the weak disposability assumption and only strong disposability of good outputs. However the null-jointness is imposed by the following restrictions on bad outputs:

$$
\begin{aligned}
& \sum\nolimits_{k=1}^{K} u_{kj} > 0, j = 1,\dots,J, \\
& \sum\nolimits_{j=1}^{J} u_{kj} > 0, k = 1,\dots,K.
\end{aligned} \tag{2}
$$

Furthermore, we apply the directional distance function approach as in Chung et al. (1997) and in order to be able to reduce bad and expand good outputs[14]. In order to be able to model that in the directional distance function setting we use a direction vector $g = (g_v, -g_u)$, where $g_v = 1$ and $-g_u = -1$. Then the efficiency score for a region k' can be obtained from:

$$
\begin{aligned}
& D\left(x^{k'}, v^{k'}, u^{k'}; g\right) = \max \beta \\
& s.t. \left(v^{k'} + \beta g_v, u^{k'} - \beta g_u\right) \in P(x)
\end{aligned} \tag{3}
$$

In this way, the linear programming problem can be calculated as:

---

[13] Following Zelenyuk and Zheka (2006, p.149) our regional environmental efficiency measurement follows the most common assumption made in economics which is the constant returns to scale (CRS) assumption. In addition the CRS assumption provides us with greater discriminative power among the examined regions. Finally, due to the fact that we have a small sample size (37 regions) it is therefore better for our analysis to use more robust scale assumptions. Still if a researcher wants to impose variables returns to scale in this model, it is suggested to read first the remarks raised by Kuosmanen (2005), Färe and Grosskopf (2009) and Kuosmanen and Podinovski (2009).

[14] This is the most common assumption made for directional distance functions when measuring environmental efficiency levels. However, different directions can be chosen in order for the researcher to test the efficiency under different environmental policy scenarios (Halkos and Tzeremes, 2012).

$$
\begin{aligned}
& D\left(x^{k'}, v^{k'}, u^{k'}; g\right) = \max \beta \\
& s.t. \quad \sum_{k=1}^{K} \omega_k v_{km} \geq v_{k'm} + \beta g_{vm}, m = 1, \ldots, M, \\
& \qquad \sum_{k=1}^{K} \omega_k u_{kj} = u_{k'j} - \beta g_{uj}, j = 1, \ldots, J, \\
& \qquad \sum_{k=1}^{K} \omega_k x_{kn} \leq x_{k'n} \\
& \qquad \omega_k \geq 0, k = 1, \ldots, K.
\end{aligned} \qquad (4).
$$

Efficiency is next indicated when $D\left(x^{k'}, v^{k'}, u^{k'}; g\right) = 0$ and inefficiency by $D\left(x^{k'}, v^{k'}, u^{k'}; g\right) > 0$. Due to the fact that we are using the efficiency scores obtained in a second stage analysis we present the efficiency scores obtained in terms of Shephard's output distance function. In fact according to Chung et al. (1997) Shephard's output distance function is a special case of the directional distance function and can be calculated as:

$$
D(x, v, u) = 1 / \left(1 + D\left(x^k, v^k, u^k; v^k, u^k\right)\right) \qquad (5).
$$

3.2 Conditional directional distance functions incorporating bad outputs

Following Daraio and Simar (2005) who extent the probabilistic formulation of the production process firstly introduced by Cazals et al. (2002)[15], let the joint probability measure of $\left(X, Y^{v,u}\right)$ and the joint probability function of $H_{XY^{v,u}}(.,.)$ can be defined as[16]:

$$
H_{XY^{v,u}}\left(x, y^{v,u}\right) = \text{Prob}\left(X \leq x, Y^{v,u} \geq y^{v,u}\right) \qquad (6).
$$

In addition the following decomposition can be obtained as:

[15] For the theoretical background and the asymptotic properties of nonparametric conditional efficiency measures see Jeong et al. (2010).

[16] For simplicity of presentation $Y^{v,u}$ symbolizes bad (u) and good (v) outputs.

$$H_{XY^{v,u}}\left(x, y^{v,u}\right) = \text{Prob}\left(Y^{v,u} \geq y^{v,u} \middle| X \leq x\right)\text{Prob}\left(X \leq x\right) = S_{Y^{v,u}|X}\left(y^{v,u} \middle| x\right)F_X\left(x\right) \quad (7),$$

where $F_X(x) = \text{Prob}(X \leq x)$ and $S_{Y^{v,u}|X}\left(y^{v,u} \middle| x\right) = \text{Prob}\left(Y^{v,u} \geq y^{v,u} \middle| X \leq x\right)$.

In addition let $Z \in R^r$ denote the exogenous factors to the production process (in our case is the GDPPC). Then equation (6) becomes:

$$H_{XY^{v,u}|Z}\left(x, y^{v,u} \middle| z\right) = \text{Prob}\left(X \leq x, Y^{v,u} \geq y^{v,u} \middle| Z = z\right) \quad (8),$$

which complete characterizes the production process. According to Daraio and Simar (2005, 2006, 2007) the following decomposition can be derived:

$$\begin{aligned} H_{XY^{v,u}|Z}\left(x, y^{v,u} \middle| z\right) &= \text{Prob}\left(Y^{v,u} \geq y^{v,u} \middle| X \leq x, Z = z\right)\text{Prob}\left(X \leq x \middle| z\right) \\ &= S_{Y^{v,u}|X,Z}\left(y^{v,u} \middle| x, z\right)F_{X|Z}\left(x \middle| z\right) \end{aligned} \quad (9).$$

The estimator of the conditional survival function introduced above can be obtained from:

$$\hat{S}_{Y^{v,u}|X,Z}\left(y^{v,u} \middle| x, z\right) = \frac{\sum_{i=1}^{n} I\left(Y_i^{v,u} \geq y^{v,u}, X_i \leq x\right)K_h\left(Z_i, z\right)}{\sum_{i=1}^{n} I\left(X_i \leq x\right)K_h\left(Z_i, z\right)} \quad (10),$$

where $K_h\left(Z_i, z\right) = h^{-1}K\left(\left(Z_i - z\right)/h\right)$ with $K(.)$ being a univariate kernel defined on a compact support (Epanechnikov in our case) and h is the appropriate bandwidth calculated following Bădin et al. (2010)[17].

Recently Simar and Vanhems (2012) developed the probabilistic characterization of directional distance function taking the general form of:

$$D\left(x, y; g_x, g_y\right) = \sup\left\{\beta > 0 \middle| H_{XY}\left(x - \beta g_x, y + \beta g_y\right) > 0\right\} \quad (11)$$

and the conditional directional distance function of (x, y) conditional on $Z = z$ can then be defined as:

[17] The calculation of bandwidth by Bădin et al. (2010) is based on the Least Squares Cross Validation (LSCV) criterion introduced by Hall et al. (2004) and Li and Racine (2007).

$$D\left(x, y; g_x, g_y \middle| z\right) = \sup\left\{\beta > 0 \middle| H_{XY|Z}\left(x - \beta g_x, y + \beta g_y \middle| Z = z\right) > 0\right\} \qquad (12).$$

Based on those developments the probabilistic form of Färe and Grosskopf 's (2004) model (presented previously) measuring environmental efficiency will take respectively the form of:

$$D\left(x^{k'}, v^{k'}, u^{k'}; g_v, g_u\right) = \sup\left\{\beta > 0 \middle| H_{XY^{v,u}}\left(x^{k'}, v^{k'} + \beta g_v, u^{k'} - \beta g_u\right) > 0\right\} \qquad (13),$$

In addition the conditional form of the model will take the form of

$$D\left(x^{k'}, v^{k'}, u^{k'}; g_v, g_u \middle| z\right) = \sup\left\{\beta > 0 \middle| H_{XY^{v,u}|Z}\left(x^{k'}, v^{k'} + \beta g_v, u^{k'} - \beta g_u \middle| Z = z\right) > 0\right\} (14).$$

Finally, the DEA program for the environmental efficiency score for a region k' when using the conditional output oriented directional distance function can be calculated as:

$$\begin{aligned} & D\left(x^{k'}, v^{k'}, u^{k'}; g_v, g_u \middle| z\right) = \max \beta \\ & s.t. \quad \sum_{\substack{k=1,\ldots,K| \\ |Z_k - z| \leq h}} \omega_k v_{km} \geq v_{k'm} + \beta g_{vm}, m = 1, \ldots, M, \\ & \qquad \sum_{\substack{k=1,\ldots,K| \\ |Z_k - z| \leq h}} \omega_k u_{kj} = u_{k'j} - \beta g_{uj}, j = 1, \ldots, J, \qquad (15) \\ & \qquad \sum_{\substack{k=1,\ldots,K| \\ |Z_k - z| \leq h}} \omega_k x_{kn} \leq x_{k'n} \\ & \qquad \omega_k \geq 0, k = 1, \ldots, K \text{ such that } \left|Z_k - z\right| \leq h. \end{aligned}$$

3.3 Determining the effect of the exogenous variables

In order to identify the effect of per capita regional economic growth on regional environmental efficiency (REE) levels without specifying in prior any functional relationship, our paper applies a nonparametric regression in the principles of Daraio and Simar (2005, 2006, 2007). Following, Li and

Racine (2007) and Racine (2008) let us have a random variable X (regional GDP per capita-GDPPC) with a probability density function (PDF) $f(x)$. Then the Gaussian kernel $K(x)$ can be defined as:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \tag{16}$$

and the PDF of $f(x)$ can be obtained from:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) \tag{17}$$

where h represents the bandwidth calculated by the least squares cross-validation data driven method as suggested by Hall et al. (2004). In addition let us have the variable Y to denote the ratio of $\frac{D\left(x^{k'}, v^{k'}, u^{k'}; g_v, g_u \middle| z\right)}{D\left(x^{k'}, v^{k'}, u^{k'}; g_v, g_u\right)} = Q$.

The joint PDF of (X, Y) can be defined as:

$$\hat{f}(x,y) = \frac{1}{nh_x h_y} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_x}\right) K\left(\frac{Y_i - y}{h_y}\right) \tag{18}$$

where (h_x, h_y) are representing the bandwidths calculated by the least squares cross-validation data driven method and $K(.)$ represents the Gaussian kernel defined previously.

The conditional PDF between the two variables accordingly can be obtained from:

$$\hat{g}(y|x) = \hat{f}(x,y) / \hat{f}(x) \qquad (19).$$

Then our nonparametric regression will have the general form of:

$$Y = g(X) + u \qquad (20),$$

but as we don’t know the functional form of $g(.)$ we will estimate it nonparametrically using kernel methods. In order to obtain the estimation we will need to interpret $g(x)$ as the conditional mean of Y given X. If we let Y and X be the dependent and independent variables accordingly $(Y = Q, X = GDPPC)$ following the proof from Li and Racine (2007, p. 59), $g(x) \equiv E(Y|X = x)$ then $E(Y|X)$ is the optimal predictor of Y given X. In this way we can estimate $g(x) \equiv E(Y|x)$ by:

$$\hat{g}(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{X_i - x}{h_x}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_x}\right)} \qquad (21).$$

Equation (21) represents the local constant estimator introduced from Nadaraya (1964) and Watson (1964).

In addition following the test proposed by Racine et al. (2006) and Racine (2008, p.67) we investigate the significance of regional GDPPC explaining the variations of regional REE. Specifically, if ξ denotes the explanatory variables that have be redundant from our model and X denotes the explanatory variable used (GDPPC in our case), then the null hypothesis can be written as $H_0 : E(y|x,\xi) = E(Y|\xi)$ almost everywhere. This can be equivalent to $H_0 : \frac{\partial E(y|x,\xi)}{\partial x} = \beta(x) = 0$ almost everywhere. Next the test statistic can be defined as:

$$I = E\left\{\beta(x)^2\right\} \qquad (22)$$

By forming a sample of average of I, we can replace the unknown derivatives with their nonparametric estimates (Racine, 1997). The test statistic can then be approximated as:

$$I_n = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}(X_i)^2 \tag{23}$$

where $\hat{\beta}(X_i)$ is the local constant partial derivative estimator presented above. Since I_n is a consistent estimator of I, $I_n \to 0$ under H_0 and $I_n \to I > 0$ under H_1. Finally, in order to obtain the distribution of the test statistic under the null hypothesis we apply bootstrap procedures as described in Racine (1997).

Based on the visualization effect proposed by Daraio and Simar (2005, 2006, 2007) of the exogenous variable (Z), if the regression line is increasing it indicates that Z is unfavourable to regions' environmental efficiency, whereas if it is decreasing then it is favourable. When Z is unfavourable then the per capita regional GDP acts like an extra undesired output to be produced demanding the use of more inputs in the environmental production activity. In the opposite case it plays a role of a substitutive input in the production process giving the opportunity to save inputs in the activity of production.

4. Empirical Results

The empirical results (table 2) present the REE scores of the U.K. regions both for unconditional $[D(x^{k'}, v^{k'}, u^{k'}; g_v, g_u)]$ and conditional to GDPPC $[D(x^{k'}, v^{k'}, u^{k'}; g_v, g_u | z)]$ measures. The unconditional REE values reveal that eight regions are environmentally efficient. These are Cornwall and Isles of Scilly, Devon, Greater Manchester, Herefordshire-Worcestershire and Warwickshire, Highlands and Islands, Inner London, Surrey-East and West Sussex and West Wales-The Valleys. In addition the eight regions with the lowest environmental performance are Hampshire-Isle of Wight, Tees Valley-Durham, Eastern Scotland, Lancashire, Essex, Derbyshire-Nottinghamshire, Shropshire-Staffordshire and East Anglia.

Table 2: The UK regions' environmental efficiency levels measured in Shephard's output distance functions

Regions (NUTS 2 level)	$D\left(x^{k'},v^{k'},u^{k'};g_v,g_u\right)$	$D\left(x^{k'},v^{k'},u^{k'};g_v,g_u \middle| z\right)$
Bedfordshire and Hertfordshire	0.5519	0.5677
Berkshire, Buckinghamshire and Oxfordshire	0.4276	0.4476
Cheshire	0.4288	0.3368
Cornwall and Isles of Scilly	1.0000	1.0000
Cumbria	0.6089	0.3486
Derbyshire and Nottinghamshire	0.3136	0.3916
Devon	1.0000	1.0000
Dorset and Somerset	0.8687	0.8484
East Anglia	0.2893	0.7606
East Wales	0.4181	0.3661
East Yorkshire and Northern Lincolnshire	0.4130	0.1808
Eastern Scotland	0.3725	0.5959
Essex	0.3350	0.4281
Gloucestershire, Wiltshire and Bristol/Bath area	0.4138	0.5855
Greater Manchester	1.0000	1.0000
Hampshire and Isle of Wight	0.3890	0.4822
Herefordshire, Worcestershire and Warwickshire	1.0000	0.9340
Highlands and Islands	1.0000	1.0000
Inner London	1.0000	0.5440
Kent	0.6538	0.5543
Lancashire	0.3635	0.3578
Leicestershire, Rutland and Northamptonshire	0.4178	0.5072
Lincolnshire	0.4588	0.2535
Merseyside	0.9145	0.8381
North Eastern Scotland	0.8216	0.3975
North Yorkshire	0.8131	0.4451
Northern Ireland (UK)	0.9993	1.0000
Northumberland and Tyne and Wear	0.4360	0.4534
Outer London	0.5957	1.0000
Shropshire and Staffordshire	0.2952	0.2929
South Western Scotland	0.6871	1.0000
South Yorkshire	0.9290	0.7283
Surrey, East and West Sussex	1.0000	1.0000
Tees Valley and Durham	0.3817	0.3486
West Midlands	0.9943	1.0000
West Wales and The Valleys	1.0000	1.0000
West Yorkshire	0.4459	0.4845
Mean	*0.6497*	*0.6346*
Max	*1.0000*	*1.0000*
Min	*0.2893*	*0.1808*
Std	*0.2737*	*0.2765*

Furthermore, when we account for the effect of GDPPC, ten regions are reported to be efficient. These are Cornwall-Isles of Scilly, Devon, Greater Manchester, Highlands-Islands, Surrey-East and West Sussex, West Wales-The Valleys, Northern Ireland (UK), West Midlands, South Western Scotland and Outer London. Similarly, the ten regions with the lowest environmental efficiency scores are North Eastern Scotland, Derbyshire-Nottinghamshire, East Wales, Lancashire, Cumbria, Tees Valley-Durham, Cheshire, Shropshire-Staffordshire, Lincolnshire and East Yorkshire-Northern Lincolnshire. The mean values suggest that the conditional REE values are slightly lower compare to the unconditional.

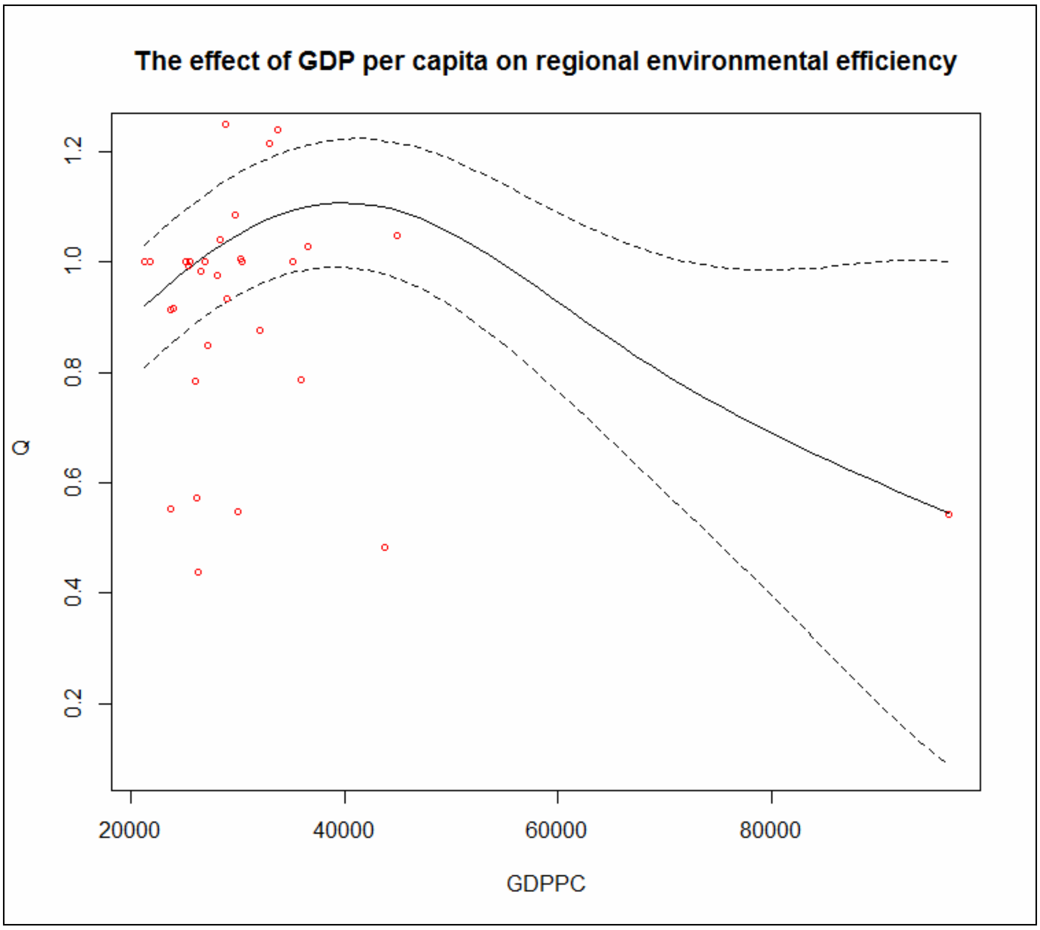
In the principles of Daraio and Simar (2005, 2006, 2007) figure 1 provides a graphical representation of the effect of regional GDPPC on the UK regions' environmental efficiency. For this task we use the 'Nadaraya-Watson' estimator, which is the most popular method for nonparametric kernel regression proposed by Nadaraya (1965) and Watson (1964). According to De Whitte and Marques (2007, p. 25) integrating conditional efficiency measures can help us to avoid main drawbacks of efficiency analysis and have some attractive features such as 1) the absence of separability condition, 2) avoiding the need of priory assumption on the functional form of the model and 3) allowing the exploration of the effect of environmental-exogenous variables. The significance of the effect of Z in the nonparametric regression setting was based on the procedure described previously (Racine, 1997; Racine et al., 2006; Li and Racine, 2007). We obtained a p-value of 0.025, which indicates significance at 5% level.

As such figure 1 illustrates the nonparametric estimate of the regression function using the conditional and unconditional regional environmental efficiency scores $Q = \frac{D\left(x^{k'}, v^{k'}, u^{k'}; g_v, g_u \middle| z\right)}{D\left(x^{k'}, v^{k'}, u^{k'}; g_v, g_u\right)}$. In addition it presents their variability bounds of pointwise error bars using asymptotic standard error formulas (Hayfield and Racine, 2008). As explained earlier when the regression is increasing, it indicates that the GDPPC factor is unfavourable to regions' environmental efficiency indicating a clear negative effect.

In our case figure 1 illustrates an increasing nonparametric regression line up to a point (40000 euros) indicating that GDPPC levels act as an extra bad output to the regional environmental production process. However, after that point the effect becomes positive (since the regression line is decreasing) and therefore the regional GDPPC levels acts as substitutive input in the regional environmental production process. Therefore, it provides regions with the opportunity to "save" inputs in the activity of environmental production.

Finally, figure 1 illustrates that there is a 'U' shape relationship between the UK regions' environmental efficiency levels and regional economic growth.

Figure 1: The effect of regional GDP per capita (GDPPC) levels on the UK regions' environmental efficiency levels (Q)



5. Conclusions

The contribution of our paper is twofold. First, it proposes an extension of the original model proposed by Färe and Grosskopf (2004) measuring environmental process of a decision making unit in order to incorporate the effect of an exogenous to the process variable. For that reason our paper applies the methodology illustrated on the work by Simar and Vanhems (2012) and develops conditional directional distance functions incorporating bad outputs. Moreover, in the principles of the studies of Daraio and Simar (2005, 2006, 2007) our paper illustrates the 'visualization' effect of the external-exogenous variable.

In addition the second contribution of our paper lies on its application of our proposed model. To our knowledge there are not any studies for EU regions investigating a Kuznets type relationship between regional environmental efficiency and regional economic growth. Our application investigates such a relationship for the 37 U.K. regions at NUTS 2 level. The results reveal the existence of a 'U' shape relationship between regional environmental efficiency and regional GDP per capita levels.

References

Ali, A.I., Lerme, C.S., Seiford, L.M., 1995. Components of efficiency evaluation in data envelopment analysis. European Journal of Operational Research 80, 462–473.

Bădin, L., Daraio, C., Simar, L., 2010. Optimal bandwidth selection for conditional efficiency measures: A Data-driven approach. European Journal of Operational Research 201, 633-640.

Banker, R.D., Morey, R.C., 1986. Efficiency analysis for exogenously fixed inputs and outputs. Operations Research 34, 513–521

Banker, R.D., Natarajan, R., 2008. Evaluating contextual variables affecting productivity using data envelopment analysis. Operations Research 56, 48–58.

Bian, Y., Yang, F., 2010. Resource and environment efficiency analysis of provinces in China: A DEA approach based on Shannon's entropy. Energy Policy 38, 1909-1917.

Brajer, V., Mead, R.W., Xiao, F., 2011. Searching for an Environmental Kuznets Curve in China's air pollution. China Economic Review 22, 383-397.

Cazals, C., Florens, J.P., Simar, L., 2002. Nonparametric frontier estimation: a robust approach. Journal of Econometrics 106, 1-25.

Chung, Y.H., Färe, R., Grosskopf, S., 1997. Productivity and undesirable outputs: A directional distance function approach. Journal of Environmental Management 51, 229– 240.

Daraio, C., Simar, L., 2005. Introducing environmental variables in nonparametric frontier models: a probabilistic approach. Journal of Productivity Analysis 24, 93–121

Daraio, C., Simar, L., 2006. A robust nonparametric approach to evaluate and explain the performance of mutual funds. European Journal of Operational Research 175, 516–542.

Daraio, C., Simar, L., 2007. Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. Journal of Productivity Analysis 28, 13–32

De White, K., Verschelde, M., 2010. Estimating and explaining efficiency in a multilevel setting: A robust two-stage approach. TIER working paper series, TIER WP 10/04, Top Institute for Evidence Based Education Research, University of Amsterdam, Maastricht University, University of Groningen.

Diao, X.D., Zeng, S.X., Tam, C.M., Tam, V.W.Y., 2009. EKC analysis for studying economic growth and environmental quality: A case study in China. Journal of Cleaner Production 17, 541–548.

Epstein, L., Denny, M., 1980. Endogenous capital utilization in a short-run production model. Journal of Econometrics 12, 189-207.

Färe, R., Grosskopf, S., 2003. Non-parametric Productivity Analysis with Undesirable Outputs: Comment. American Journal of Agricultural Economics 85, 1070–74.

Färe, R., Grosskopf, S., 2009. A Comment on Weak Disposability in Nonparametric Production Analysis. American Journal of Agricultural Economics 91, 535-538.

Färe, R., Grosskopf, S., Hernandez-Sancho, F., 2004. Environmental performance: an index number approach. Resource & Energy Economics 26, 343–352.

Färe, R., Grosskopf, S., Tyteca D., 1996. An activity analysis model of the environment performance of firms: application to fossil-fuel-fired electric utilities. Ecological Economics 18, 161–175.

Färe, R., Grosskopf, S., 2004. Modeling undesirable factors in efficiency evaluation: Comment. European Journal of Operational Research 157, 242-245.

Färe, R., Grosskopf, S., Lovell, C.A.K., Pasurka, C., 1989. Multilateral productivity comparisons when some outputs are undesirable. Review of Economics and Statistics 71, 90–98.

Färe, R., Primont, D., 1995. Multi-Output Production and Duality: Theory and Applications. Kluwer Academic Publishers, Boston.

Feldstein, M.S., Foot, D.K., 1971. The other half of gross investment: replacement and modernization expenditures. Review of Economics and Statistics 53, 49-58.

Grossman, G.M., Krueger, A.B., 1995. Economic growth and the environment. Quarterly Journal of Economics 33, 353-377.

Guo, X.D., Zhu, L., Fan, X., Xie, B.C., 2011. Evaluation of potential reductions in carbon emissions in Chinese provinces based on environmental DEA. Energy Policy 39, 2352-2360.

Hailu, A., 2003. Non-parametric Productivity Analysis with Undesirable Outputs: Reply. American Journal of Agricultural Economics 85, 1075–1077.

Hailu A., Veeman, T.S., 2001. Non-parametric Productivity Analysis with Undesirable Outputs: An Application to the Canadian Pulp and Paper Industry. American Journal of Agricultural Economics 83, 605–616.

Halkos, G.E., Tzeremes, N.G., 2009. Exploring the existence of Kuznets curve in countries' environmental efficiency using DEA window analysis. Ecological Economics 68, 2168-2176.

Halkos, G.E., Tzeremes, N.G., 2011. Modelling regional welfare efficiency applying conditional full frontiers. *Spatial Economic Analysis* 6, 451-471.

Halkos, G.E., Tzeremes, N.G., 2012. Measuring German regions' environmental efficiency: a directional distance function approach. Letters in Spatial and Resource Sciences 5, 7-16.

Hall, P., Racine, J.S., Li, Q. 2004. Cross-validation and the estimation of conditional probability densities. Journal of the American Statistical Association 99, 1015–1026.

Hayfield, T., Racine, J.S. 2008. Nonparametric Econometrics: The np Package. Journal of Statistical Software 27, 1-32.

He, J., 2008. China's industrial SO2 emissions and its economic determinants: EKC's reduced vs. structural model and the role of international trade. Environment and Development Economics 14, 227–262.

Hoff, A., 2007. Second stage DEA: comparison of approaches for modelling the DEA score. European Journal of Operational Research 181, 425–435.

Jeong, S.O., Park, B.U., Simar, L., 2010. Nonparametric conditional efficiency measures: asymptotic properties. Annals of Operations Research 173, 105-122.

Knill, C., Lenschow, A., 1998. Coping with Europe: the impact of British and German administrations on the implementation of EU environmental policy. Journal of European Public Policy 5, 595-614.

Korhonen, P., Luptacik, M., 2004. Eco-efficiency analysis of power plants: an extension of data envelopment analysis. European Journal of Operational Research 154, 437–446.

Kuosmanen, T., 2005. Weak disposability in nonparametric production analysis with undesirable outputs. American Journal of Agricultural Economics 87, 1077-1082.

Kuosmanen, T., Podinovski, V., 2009. Weak disposability in nonparametric production analysis: reply to Färe and Grosskopf. American Journal of Agricultural Economics 91, 539-545.

Kuznets, S., 1955. Economic growth and income inequality. American Economic Review 45, 1-28.

Li, Q., Racine, J.S., 2007. Nonparametric econometrics: Theory and practice. Princeton University Press, Oxford.

Macpherson, A.J., Principe, P.P., Smith, E.R., 2010. A directional distance function approach to regional environmental- economic assessments. Ecological Economics 69, 1918-1925.

Managi, S., Jena, P.R., 2008. Environmental productivity and Kuznets curve in India. Ecological Economics 65, 432-440.

Managi, S., 2006. Are there increasing returns to pollution abatement? Empirical analytics of the Environmental Kuznets Curve in pesticides. Ecological Economics 58, 617-636.

Mandal, S.K., Madheswaran, S., 2010. Environmental efficiency of the Indian cement industry: An interstate analysis. Energy Policy 38,1108-1118.

McDonald, J., 2009, Using least squares and tobit in second stage dea efficiency analyses. European Journal of Operational Research 197,792–798.

Nadaraya, E.A., 1964. On Estimating Regression. *Theory of Probability and its Applications* 9, 141–142.

Park, B.U., Simar, L., Zelenyuk, V., 2008. Local likelihood estimation of truncated regression and its partial derivative: Theory and application. Journal of Econometrics 146, 185–2008.

Picazo-Tade, A., García-Reche, A., 2007. What makes environmental performance differ between firms? Empirical evidence from the Spanish tile industry. Environment and Planning A 39, 2232–2247.

Racine, J.S., 1997. Consistent significance testing for nonparametric regression. Journal of Business and Economic Statistics 15, 369–379.

Racine, J.S., 2008. Nonparametric Econometrics: A Primer. Foundation and Trends in Econometrics 3, 1-88.

Racine, J.S., Hart, J.D., Li, Q., 2006. Testing the significance of categorical predictor variables in nonparametric regression models. Econometric Reviews 25, 523–544.

Reinhard, S., Lovell, C.A.K., Thijssen, G., 2000. Environmental efficiency with multiple environmentally detrimental variables estimated with SFA and DEA. European Journal of Operational Research 121, 287–303.

Seiford, L.M., Zhu, J., 1998. Identifying excesses and deficits in chinese industrial productivity (1953–1990): A weighted data envelopment analysis approach. Omega-International Journal of Management Science 26, 269–279.

Seiford, L.M., Zhu, J., 2002. Modeling undesirable factors in efficiency evaluation. European Journal of Operational Research 142, 16–20.

Seiford, L.M., Zhu, J., 2005. A response to comments on modeling undesirable factors in efficiency evaluation. European Journal of Operational Research 161, 579–581.

Selden, T., Song, D., 1994. Environmental quality and development: is there a Kuznets curve for air pollution emissions? Journal of Environmental Economics and Management 27, 147-162.

Shephard, R.W., 1970. Theory of Cost and Production Functions. Princeton University Press, Princeton, NJ.

Shi, G.M., Bi, J., Wang, J.N., 2010. Chinese regional industrial energy efficiency evaluation based on a DEA model of fixing non-energy inputs. Energy Policy 38, 6172-6179.

Simar, L., Wilson, P.W., 2007. Estimation and inference in two-stage, semi-parametric models of productive efficiency. Journal of Econometrics 136:31–64.

Simar, L., Wilson, P.W., 2011. Two-stage DEA: caveat emptor. Journal of Productivity Analysis 36, 205-218.

Simar, L., Vanhems, A., 2012. Probabilistic characterization of directional distances and their robust versions. Journal of Econometrics 166, 342-354.

Taskin, F., Zaim, O., 2001. The role of international trade on environmental efficiency: a DEA approach. Economic Modelling 18, 1-17.

Thrall, R.M., 1996. Duality, classification and slacks in DEA. Annals of Operations Research 66, 109–138.

Tyteca, D., 1997. Linear programming models for the measurement of environmental performance of firms: concepts and empirical results. *Journal* of Productivity Analysis 8, 175–189.

Tyteca, D., 1996. On the Measurement of the Environmental Performance of Firms- A Literature Review and a Productive Efficiency Perspective. Journal of Environmental Management 46, 281- 308.

Wang K., Yu, S., Zhang, W., 2012. China's regional energy and environmental efficiency: A DEA window analysis based dynamic evaluation. Mathematical and Computer Modelling. doi:10.1016/j.mcm.2011.11.067.

Watanabe, M., Tanaka, K., 2007. Efficiency analysis of Chinese industry: A directional distance function approach. Energy Policy 35, 6323-6331.

Watson, G.S., 1964. Smooth regression analysis. Sankhya 26, 359–372.

Yörük, B.K., Zaim, O., 2006. The Kuznets curve and the effect of international regulations on environmental efficiency. Economics Bulletin 17, 1-7.

Zaim, O., 2004. Measuring environmental performance of state manufacturing through changes in pollution intensities: a DEA framework. Ecological Economics, 48, 37-47.

Zaim, O., Taskin, F., 2000a. Environmental efficiency in carbon dioxide emissions in the OECD: A non-parametric approach. Journal of Environmental Management 58, 95–107.

Zaim, O., Taskin, F., 2000b. Searching for a Kuznets curve in environmental efficiency using kernel estimation. Economics Letters 68, 217–223.

Zaim, O., Taskin, F., 2000c. A Kuznets curve in environmental efficiency: an application on OECD countries. Environmental and Resource Economics 17, 21-36.

Zelenyuk, V., Zheka, V., 2006. Corporate governance and firm's efficiency: the case of a transitional country, Ukraine. Journal of Productivity Analysis 25, 143-157.

Zhou, P., Ang, B.W., Poh, K.L., 2006. Slacks-based efficiency measures for modelling environmental performance. Ecological Economics 60, 111–118.

Zhou, P., Poh, K.L., Ang, B.W., 2007. A non-radial DEA approach to measuring environmental performance. European Journal of Operational Research 178, 1–9.

Zofio, J.L., Prieto, A.M., 2001. Environmental efficiency and regulatory standards: the case of CO2 emissions from OECD industries. Resource and Energy Economics 23, 63–83.