Economic growth and environmental efficiency: Evidence from U.S. regions

Halkos, George and Tzeremes, Nickolaos

University of Thessaly, Department of Economics

November 2012

Economic growth and environmental efficiency: Evidence from U.S. regions

By

George E. Halkos* and Nickolaos G. Tzeremes

Laboratory of Operations Research, Department of Economics, University of Thessaly, Korai 43, 38333, Volos, Greece.

Abstract

This paper proposes a conditional directional distance function model in order to examine the link between regional environmental efficiency and GDP per capita levels. As an illustrative example we apply our model to USA regional data revealing an inverted 'U' shape relationship between regional environmental efficiency and per capita income. The results derived from a non-parametric regression indicate a turning point at 49,000 dollars.

Keywords: Regional environmental efficiency; Directional distance function; Conditional measures; U.S. regions.

JEL Codes: C14; Q50; R11.

* **Address for Correspondence:** Laboratory of Operations Research, 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 74772

1. Introduction

Recently, Simar and Vanhems (2012) introduced directional distance functions conditioned to exogenous (environmental) factors. The proposed formulation incorporates into the efficiency measurement an exogenous factor that may influence the production process. This note applies the methodology of conditional directional distance functions into an environmental problem, in order to analyse the impact of economic growth on regional environmental efficiency for the case of the U.S. regions.

Specifically, our paper extents the model originated by Kuosmanen (2005) measuring environmental efficiency. The modified version is based on conditional directional distance functions incorporating the effect of exogenous factors. In our empirical application our proposed model examines the effect of regional economic growth on 51 U.S. regions' environmental efficiency levels.

2. Proposed model

Following several authors (Kuosmanen, 2005; Kuosmanen and Podinovski, 2009; Podinovski and Kousmanen 2011) in every environmental production activity characterised in the context of data envelopment analysis (DEA) terminology there is a vector $\mathbf{v} = (v_1, \ldots v_M) \in \Re_+^M$ indicating desirable (or good) outputs, a vector $\mathbf{w}(w_1, \ldots, w_J) \in \Re_+^J$ indicating the undesirable (or bad) outputs and a vector $\mathbf{x}(x_1, \ldots, x_N) \in \Re_+^N$ indicting the inputs used. Having k regions under consideration the observed activities can be defined as $(\mathbf{v}^k, \mathbf{w}^k, \mathbf{x}^k)$, $k = 1, \ldots K$ and the production technology can be represented as:

$$Y = \left\{ \left(\mathbf{v}, \mathbf{w}, \mathbf{x} \right) \middle| \, \mathbf{x} \text{ can produce } \left(\mathbf{v}, \mathbf{w} \right) \right\} \tag{1}$$

Kuosmanen (2005) developed a model for a technology assuming weak disposability of bad outputs convexity and individual abatement factors θ^k for every observed activity $k = 1,...K$ as:

$$\begin{aligned}
\hat{Y} = \Big\{ \left(\mathbf{v}, \mathbf{w}, \mathbf{x} \right) : &\sum_{k=1}^{K} \theta^k \omega^k v_m^k \geq v_m, \forall m \\
&\sum_{k=1}^{K} \theta^k \omega^k w_j^k = w_j, \forall j \\
&\sum_{k=1}^{K} \omega^k x_n^k \leq x_n, \forall n \\
&\sum_{k=1}^{K} \omega^k = 1 \\
&\omega^k \geq 0, \forall k \\
&0 \leq \theta^k \leq 1, \forall k \Big\}.
\end{aligned} \tag{2}$$

Where variables $\boldsymbol{\omega} = \left(\omega^1, ..., \omega^k \right)$ indicate the intensity weights. Then, in order to linearize (2) we can use the following substitutions:

$$\begin{aligned}
&\lambda^k = \theta^k \omega^k, \qquad \mu^k = \left(1 - \theta^k \right) \omega^k, \forall k, \\
&\text{so that} \qquad \lambda^k + \mu^k = \omega^k
\end{aligned} \tag{3}$$

Then fully linearized version of (2) can be rewritten as:

$$\begin{aligned}
\hat{Y} = \Big\{ \left(\mathbf{v}, \mathbf{w}, \mathbf{x} \right) : &\sum_{k=1}^{K} \lambda^k v_m^k \geq v_m, \forall m \\
&\sum_{k=1}^{K} \lambda^k w_j^k = w_j, \forall j \\
&\sum_{k=1}^{K} \left(\lambda^k + \mu^k \right) x_n^k \leq \left(\lambda^k + \mu^k \right) x_n, \forall n \\
&\sum_{k=1}^{K} \left(\lambda^k + \mu^k \right) = 1 \\
&\lambda^k, \mu^k \geq 0, \forall k \Big\}.
\end{aligned} \tag{4}$$

For given activity of a region $\left(\mathbf{v}^0,\mathbf{w}^0,\mathbf{x}^0\right)$ the output directional distance function allowing a simultaneous increase in good and a reduction of bad output (Chambers et al., 1996, 1998; Chung et al., 1997) can be defined as:

$$D\left(\mathbf{v}^0,\mathbf{w}^0,\mathbf{x}^0;\mathbf{g}^v;\mathbf{g}^w\right)=\sup\left\{\phi\middle|\left(\mathbf{v}^0+\phi\mathbf{g}^v,\mathbf{w}^0-\phi\mathbf{g}^w,\mathbf{x}^0\right)\in Y\right\} \tag{5}$$

Finally, the linear programme calculating regions output directional distance function under the Kuosmanen (2005) technology can be defined as:

$$\begin{aligned}
&\hat{D}=\left(\mathbf{v}^0,\mathbf{w}^0,\mathbf{x}^0;\mathbf{g}^v;\mathbf{g}^w\right)=\max\phi \\
&\text{Subject to} \quad \sum_{k=1}^{K}\lambda^k v_m^k \geq v_m^0+\phi g_m^v, \forall m \\
&\qquad \sum_{k=1}^{K}\lambda^k w_j^k = w_j^0-\phi g_j^w, \forall j \\
&\qquad \sum_{k=1}^{K}\left(\lambda^k+\mu^k\right)x_n^k \leq \left(\lambda^k+\mu^k\right)x_n^0, \forall n \\
&\qquad \sum_{k=1}^{K}\left(\lambda^k+\mu^k\right)=1 \\
&\qquad \lambda^k,\mu^k \geq 0, \forall k.
\end{aligned} \tag{6}$$

Daraio and Simar (2005) extended the probabilistic formulation of the production process first introduced by Cazals et al. (2002)[1]. In our proposed model we define the joint probability measure of our environmental production $\left(\mathbf{v},\mathbf{w},\mathbf{x}\right)$ and the joint probability function of $H_{\mathbf{v},\mathbf{w},\mathbf{x}}(.,.)$ as:

$$H_{\mathbf{v},\mathbf{w},\mathbf{x}}\left(x,v,w\right)=\text{Prob}\left(\mathbf{x}\leq x,\mathbf{v}\geq v,\mathbf{w}\geq w\right) \tag{7}$$

Then the following decomposition can be obtained as:

$$H_{\mathbf{v},\mathbf{w}}\left(x,v,w\right)=\text{Prob}\left(\mathbf{v}\geq v,\mathbf{w}\geq w\middle|\mathbf{x}\leq x\right)\text{Prob}\left(\mathbf{x}\leq x\right)=S_{\mathbf{v},\mathbf{w}|\mathbf{x}}\left(v,w\middle|x\right)F_{\mathbf{x}}\left(x\right) \tag{8}$$

where $F_{\mathbf{x}}\left(x\right)=\text{Prob}\left(\mathbf{x}\leq x\right)$ and $S_{\mathbf{v},\mathbf{w}|\mathbf{x}}\left(v,w\middle|x\right)=\text{Prob}\left(\mathbf{v}\geq v,\mathbf{w}\geq w\middle|\mathbf{x}\leq x\right)$.

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

Moreover, let $\mathbf{z} = (z_1,..., z_r) \in \Re^r$ denote the environmental (exogenous) factors to the production process (in our case is the GDP per capita-GDPPC). Then equation (7) becomes:

$$H_{\mathbf{v},\mathbf{w},\mathbf{x}|\mathbf{z}}\left(x, v, w \middle| z\right) = \text{Prob}\left(\mathbf{x} \le x, \mathbf{v} \ge v, \mathbf{w} \ge w \middle| \mathbf{z} = z\right) \tag{9},$$

which completely characterizes the environmental production process under the effect of an external variable. The following decomposition can then be derived:

$$\begin{aligned} H_{\mathbf{v},\mathbf{w},\mathbf{x}|\mathbf{z}}\left(x, v, w \middle| z\right) &= \text{Prob}\left(\mathbf{v} \ge v, \mathbf{w} \ge w \middle| \mathbf{x} \le x, \mathbf{z} = z\right)\text{Prob}\left(\mathbf{x} \le x \middle| z\right) \\ &= S_{\mathbf{v},\mathbf{w}|\mathbf{x},\mathbf{z}}\left(v, w \middle| x, z\right) F_{\mathbf{x}|\mathbf{z}}\left(x \middle| z\right) \end{aligned} \tag{10}.$$

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

$$\hat{S}_{\mathbf{v},\mathbf{w}|\mathbf{x},\mathbf{z}}\left(v, w \middle| x, z\right) = \frac{\sum_{i=1}^{n} I\left(\mathbf{v}_i \ge v, \mathbf{w}_i \ge w, \mathbf{x}_i \le x\right) K_h\left(\mathbf{z}_i, z\right)}{\sum_{i=1}^{n} I\left(\mathbf{x}_i \le x\right) K_h\left(\mathbf{z}_i, z\right)} \tag{11},$$

where $K_h(\mathbf{z}_i, z) = h^{-1}K((\mathbf{z}_i - z)/h)$ 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)[2].

Recently Simar and Vanhems (2012) developed the probabilistic characterization of directional distance function which according to (5) will take the following form:

$$D\left(\mathbf{v}^0, \mathbf{w}^0, \mathbf{x}^0; \mathbf{g}^v; \mathbf{g}^w\right) = \sup\left\{\phi \middle| H_{\mathbf{v},\mathbf{w},\mathbf{x}}\left(\mathbf{v}^0 + \phi\mathbf{g}^v, \mathbf{w}^0 - \phi\mathbf{g}^w, \mathbf{x}^0\right)\right\} \tag{12}$$

and the conditional directional distance function of $(\mathbf{v}, \mathbf{w}, \mathbf{x})$ conditioned on $\mathbf{z} = z$ can then be defined as:

$$D\left(\mathbf{v}^0, \mathbf{w}^0, \mathbf{x}^0; \mathbf{g}^v; \mathbf{g}^w \middle| \mathbf{z}\right) = \sup\left\{\phi \middle| H_{\mathbf{v},\mathbf{w}|\mathbf{z},\mathbf{x}}\left(\mathbf{v}^0 + \phi\mathbf{g}^v, \mathbf{w}^0 - \phi\mathbf{g}^w, \mathbf{x}^0 \middle| \mathbf{z} = z\right)\right\} \tag{13}.$$

[2] 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).

Finally, the proposed model measuring regions' environmental efficiency[3] based on the Kuosmanen (2005) technology and following the variable returns to scale (Banker et al., 1984) can be calculated as:

$$\hat{D} = \left(\mathbf{v}^0, \mathbf{w}^0, \mathbf{x}^0; \mathbf{g}^v; \mathbf{g}^w \middle| \mathbf{z}\right) = \max \phi$$

$$\begin{aligned}
\text{Subject to} \quad & \sum_{\substack{k=1,\dots,K| \\ |\mathbf{z}_k - z| \le h}} \lambda^k v_m^k \ge v_m^0 + \phi g_m^v, \forall m \\
& \sum_{\substack{k=1,\dots,K| \\ |\mathbf{z}_k - z| \le h}} \lambda^k w_j^k = w_j^0 - \phi g_j^w, \forall j \\
& \sum_{\substack{k=1,\dots,K| \\ |\mathbf{z}_k - z| \le h}} \left(\lambda^k + \mu^k\right) x_n^k \le \left(\lambda^k + \mu^k\right) x_n^0, \forall n \\
& \sum_{\substack{k=1,\dots,K| \\ |\mathbf{z}_k - z| \le h}} \left(\lambda^k + \mu^k\right) = 1 \\
& \lambda^k, \mu^k \ge 0, \forall k.
\end{aligned} \tag{14}$$

In problem (14) we model the direct influence of the exogenous variable $\mathbf{z}$ (in our case GDP per capita) which in turn it shapes the environmental production frontier. Therefore, the efficiency estimates obtained are determined by the inputs (capital stock and total labor force), the good output (regional GDP), the bad output (regional carbon dioxide emission levels) and the exogenous variable (regional GDP per capita) accordingly[4]. As a result the conditional directional distance function is obtained only by points taking their $\mathbf{z}$ value in the neighborhood of z (Daraio and Simar, 2005).

---

[3] Here we are using efficiency estimates rather that inefficiencies by adopting the transformation by Chung et al. (1997) and Champers et al. (1998). According to Podinovski and Kousmanen (2011) the conventional radial Farrell input and output efficiency measures can be obtained as special cases of the directional distance functions.

[4] All the variables used in our empirical application are referring to 2005 and they have extracted from OECD regional database. Moreover, since there is not any data available for U.S. states' capital stock we have used the perpetual inventory method. Therefore states' capital stock can be calculated as: $K_t = I_t + (1-\delta)K_{t-1}$ where K_t is the state's gross capital stock in current year; K_{t-1} is the state's gross capital stock in the previous year; I_t is the state's gross fixed capital formation and δ represents the depreciation rate of capital stock. In our study we have set δ equal to 6%.

In order to identify the effect of per capita regional economic growth ($\mathbf{z}$) on regions 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). When $\mathbf{z}$ is univariate (as in our case), a scatter plot of the ratio $\hat{D} = \left(\mathbf{v}^0, \mathbf{w}^0, \mathbf{x}^0; \mathbf{g}^v; \mathbf{g}^w \middle| \mathbf{z}\right) / \hat{D} = \left(\mathbf{v}^0, \mathbf{w}^0, \mathbf{x}^0; \mathbf{g}^v; \mathbf{g}^w\right)$ against $\mathbf{z}$ and its smooth nonparametric regression line would be able to describe the effect of $\mathbf{z}$ on regions' efficiency levels. Finally, the nonparametric regression smoothing can be presented as:

$$Q = g\left(\mathbf{z}_k\right) + \varepsilon_k, k = 1, \ldots, K \qquad (15),$$

where $Q = \dfrac{\hat{D} = \left(\mathbf{v}^0, \mathbf{w}^0, \mathbf{x}^0; \mathbf{g}^v; \mathbf{g}^w \middle| \mathbf{z}\right)}{\hat{D} = \left(\mathbf{v}^0, \mathbf{w}^0, \mathbf{x}^0; \mathbf{g}^v; \mathbf{g}^w\right)}$, and ε_k is the error term with $E\left(\varepsilon_k \middle| \mathbf{z}_k\right) = 0$, and g is the mean regression function, since $E\left(Q \middle| \mathbf{z}_k\right) = g\left(\mathbf{z}_k\right)$[5].

Since we use an output oriented conditional and unconditional directional distance functions an increasing regression line will indicate a favorable exogenous factor, whereas a decreasing regression line will indicate an unfavorable factor.

3. Empirical findings

Table 1 presents the results of the unconditional (REE) and conditional (REE|z) regional environmental efficiency estimates as derived from our proposed model. The unconditional environmental efficiency results reveal that 11 out of 51 states are reported to be environmentally efficient in terms of carbon dioxide emissions. The descriptive statistics show that all the U.S. regions have similar

[5] In our case we use the Nadaraya (1964) and Watson (1964) nonparametric regression estimator and the least squares cross-validation data driven method (Hall et al., 2004) for the bandwidth selection.

environmental efficiency levels indicated by the low values of standard deviation (0.0081) and a high environmental efficiency mean value (0.9891).

Table 1: Results of the conditional and unconditional environmental efficiency scores

States	REE	REE|z	States	REE	REE|z
Alabama	0.9908	0.2569	*Montana*	0.9940	0.2123
Alaska	0.9965	0.1261	*Nebraska*	0.9824	0.1200
Arizona	0.9945	0.2636	*Nevada*	0.9989	0.1118
Arkansas	0.9869	0.1381	*New Hampshire*	0.9769	0.0990
California	1.0000	1.0000	*New Jersey*	0.9865	0.3533
Colorado	0.9874	0.2315	*New Mexico*	0.9873	0.1248
Connecticut	0.9765	0.1231	*New York*	1.0000	0.6557
Delaware	0.9876	0.0466	*North Carolina*	0.9935	0.3505
District Of Columbia	1.0000	0.0275	*North Dakota*	1.0000	0.5572
Florida	1.0000	0.6263	*Ohio*	0.9877	0.5023
Georgia	0.9886	0.3749	*Oklahoma*	0.9898	0.1959
Hawaii	0.9862	0.0979	*Oregon*	0.9750	0.1339
Idaho	0.9909	0.0722	*Pennsylvania*	0.9913	0.5750
Illinois	0.9897	0.5713	*Rhode Island*	0.9678	0.0966
Indiana	0.9913	0.3707	*South Carolina*	0.9844	0.1930
Iowa	0.9816	0.1654	*South Dakota*	0.9803	0.1764
Kansas	0.9860	0.1628	*Tennessee*	0.9837	0.2684
Kentucky	1.0000	0.2769	*Texas*	1.0000	1.0000
Louisiana	1.0000	0.2907	*Utah*	0.9933	0.1387
Maine	0.9768	0.1022	*Vermont*	1.0000	0.2604
Maryland	0.9845	0.2192	*Virginia*	0.9921	0.2891
Massachusetts	0.9795	0.2517	*Washington*	0.9883	0.2173
Michigan	0.9833	0.4153	*West Virginia*	1.0000	0.1896
Minnesota	0.9817	0.2454	*Wisconsin*	0.9825	0.2582
Mississippi	0.9853	0.1212	*Wyoming*	1.0000	1.0000
Missouri	0.9834	0.3021			
Descriptives	**REE**	**REE|z**			
Mean	0.9891	0.2933			
Std	0.0081	0.2339			
Median	0.9883	0.2315			
Max	1.0000	1.0000			
Mean	0.9678	0.0275			

However when we account for the effect of states' GDP per capita levels, the results are changing significantly. Under the conditional environmental efficiency estimates only 3 states are reported to be environmentally efficient. The mean value of the estimated environmental efficiency results is now significantly lower (0.2933) with a high standard deviation (0.2339).

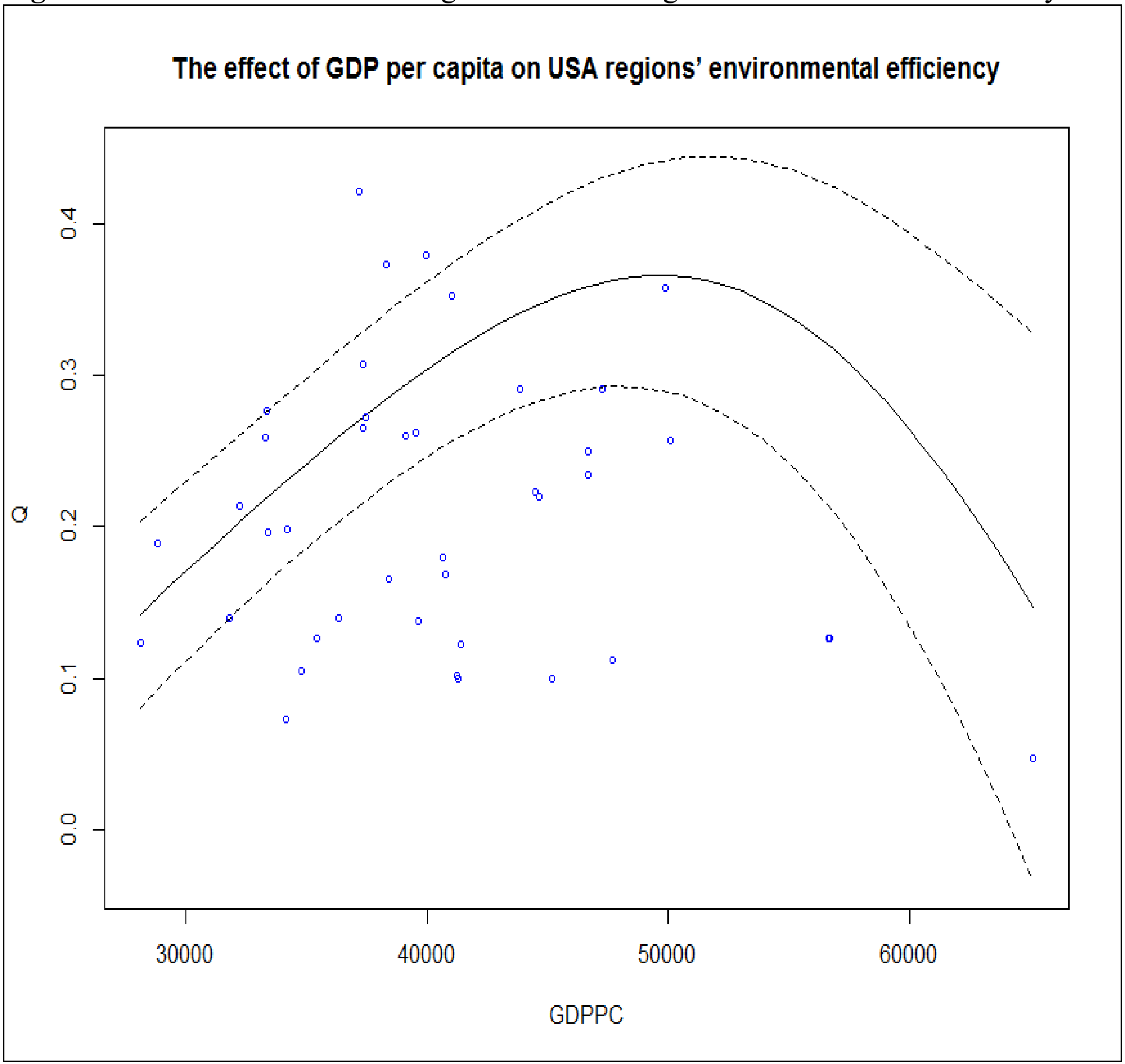
In several cases, states under the effect of economic growth have dramatically decreased their environmental efficiency levels. Under the case of unconditional environmental measures the states of District of Columbia, Florida, Kentucky, Louisiana, New York and North Dakota have been reported to be environmental efficient. However, under the effect of their GDP per capita level their environmental efficiency levels have decreased significantly.

Finally, Figure 1 presents graphically the effect of states' GDP per capita level (GDPPC) on their environmental efficiency levels. As has been analysed previously an increasing nonparametric line indicates a positive effect on regions' environmental efficiency levels whereas a decreasing line indicates a negative effect. The results indicate that the relation of states' economic growth –environmental efficiency levels has an inverted 'U' shape form. This is indicated by an increasing nonparametric regression line up to 49,000$ GDP per capita and then by a decreasing.

4. Conclusions

In this paper, we propose an extension of Kuosmanen's (2005) model incorporating exogenous factors for measuring environmental efficiency levels by applying conditional directional distance functions (Simar and Vanhems, 2012). We apply our model investigating the effect of regional GDP per capita on the 51 U.S. states' environmental efficiency levels in respect to carbon dioxide emissions. The empirical results reveal that under the effect of regional GDP per capita the regions' environmental efficiency levels are decreasing significantly. The nonparametric regression analysis reveals that the examined relationship has an inverted "U" shape form.

Figure 1: Influence of economic growth on U.S. regions' environmental efficiency



References

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., Charnes, A., Cooper, W.W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 30, 1078–1092.

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

Chambers, R.G., Chung, Y., Färe, R., 1996. Benefit and distance functions. Journal of Economic Theory 70, 407–418.

Chambers, R.G., Chung, Y., Färe, R., 1998. Profit, directional distance functions, and Nerlovian efficiency. Journal of Optimization Theory and Applications 98, 351– 364.

Chung, Y., 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.

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.

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

Kuosmanen, T., 2005. Weak disposability in nonparametric productivity 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.

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

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

Podinovski, V.V., Kuosmanen, T. 2011. Modelling weak disposability in data envelopment analysis under relaxed convexity assumptions. European Journal of Operational Research, 577–585.

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

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