



This is an Accepted Manuscript version of the article published originally by Springer accepted for publication in the journal:

Journal of Productivity Analysis

This version may differ from the original in pagination and typographic details. When using, please cite the original.

**AUTHOR(S)**

Kuosmanen, T., & Dai, S.

**TITLE**

Modeling Economies of Scope in Joint Production: Convex Regression of Input Distance Function

**YEAR**

2024

**DOI**

10.1007/s11123-024-00739-x

**CITATION**

Kuosmanen, T., & Dai, S. (2024). *Modeling Economies of Scope in Joint Production: Convex Regression of Input Distance Function*. Journal of Productivity Analysis. <https://doi.org/10.1007/s11123-024-00739-x>

**VERSION**

Accepted Manuscript

**LICENSE**

In Copyright © 2024 Springer

# Modeling economies of scope in joint production: Convex regression of input distance function

Timo Kuosmanen<sup>1</sup>, Sheng Dai<sup>2\*</sup>

<sup>1</sup> *Department of Economics, Turku School of Economics, University of Turku*

<sup>2</sup> *School of Economics, Zhongnan University of Economics and Law, 430073 Wuhan, China*

## Abstract

Modeling of joint production has proved a vexing problem. This paper develops a radial convex nonparametric least squares (CNLS) approach to estimate the input distance function with multiple outputs. We document the correct input distance function transformation and prove that the necessary orthogonality conditions can be satisfied in radial CNLS. A Monte Carlo study is performed to compare the finite sample performance of radial CNLS and other deterministic and stochastic frontier approaches in terms of the input distance function estimation. We apply our novel approach to the Finnish electricity distribution network regulation and empirically confirm that the input isoquants become more curved. In addition, we introduce the weight restriction to radial CNLS to mitigate the potential overfitting and increase the out-of-sample performance in energy regulation.

**Keywords:** Production, Convex regression, Multiple outputs, Input distance function, Energy regulation

**JEL Codes:** C14, C44, D24

---

\*Corresponding author.

*E-mail addresses:* timo.kuosmanen@utu.fi (T. Kuosmanen), sheng.dai@zuel.edu.cn (S. Dai).

# 1 Introduction

While the conventional microeconomic theory takes a single-product firm as a norm, there is growing empirical evidence that multiproduct firms are more common than usually assumed. Bernard et al. (2010) were among the first to argue that extensive reallocation of resources occurs through product switching by multiproduct firms: product switching is frequent, widespread, and influential in determining both aggregate and firm outcomes. A recent empirical study by Kuosmanen and Valmari (2023) suggest that approximately one-half of the Finnish manufacturing firms are multi-product firms, and approximately 20 percent of firms operate in multiple industry divisions.

Modeling the joint production of multiple outputs has proved a more vexing problem than most authors realize. In the deterministic setting where the input-output data contain no noise, the technology distance functions such as the input and output distance functions by Shephard (1970) and the directional distance functions (DDF) by Chambers et al. (1996, 1998) provide theoretically sound representations of the technology, which can be conveniently estimated by linear programming techniques known as data envelopment analysis (DEA; Charnes et al., 1978).<sup>1</sup> By contrast, the situation becomes more difficult in the stochastic case, where data are perturbed by random noise. Most approaches known in the literature are based on hidden assumptions that are highly restrictive and unlikely to hold.

In mainstream economics, a commonly used approach to modeling multiproduct firms is to estimate a system of single-output production functions, dividing the firm's total inputs by outputs based on the revenue shares of outputs (Foster et al., 2008). However, such a proportional assignment of ratios of inputs to outputs seems completely arbitrary: a profit-maximizing multiproduct firm would demand inputs proportionate to the revenue shares of outputs only under very restrictive assumptions, which are unlikely to hold in the real world. De Loecker et al. (2016) propose an improved version of this approach, where they use the data of single-product firms to estimate production functions and then interpolate the estimated functions to model multiproduct firms. However, the linear interpolation approach assumes away possible synergies between multiple products: the resulting output isoquants

---

<sup>1</sup>The distance functions are reciprocals of Farrell's (1957) radial measures of technical efficiency. However, the distance functions are representations of technology, which can be used for analyzing marginal properties such as elasticities of substitution, transformation, or scale, even if one is not primarily interested in efficiency measurement.

are linear by construction. This seems counterintuitive due to economies of scope, in which physical synergies imply strictly convex output sets. Indeed, synergies are often seen as the main reason why firms find it beneficial to produce multiple products jointly rather than specialize in a single product (e.g., Panzar and Willig, 1981). Therefore, assuming away the possible physical synergies from the outset seems restrictive.

In the econometric literature of stochastic frontier analysis (SFA), the modeling of joint production is usually based on Shephard's (1970) input and output distance functions (Lovell et al., 1994; Coelli and Perelman, 1999; Kumbhakar, 2013). While the distance functions provide a general and theoretically sound framework for modeling joint production, the parametric functional forms usually employed in SFA tend to violate the theoretical properties of the distance function (Färe and Primont, 1995). For instance, it is easy to show that the Cobb-Douglas functional form for the distance function cannot support convex output sets at any parameter values, in fact, the output sets are not even bounded. The commonly used translog functional form is flexible enough to support convex output sets, but in that case, the output sets bend backward, violating the free disposability property. As a result, the translog distance functions cannot satisfy both convexity and free disposability axioms.<sup>2</sup>

To our knowledge, the only theoretically sound approach to estimating technology distance functions of multiproduct firms is to resort to the shape-constrained nonparametric regression, following Kuosmanen and Johnson (2017). Convex regression is a fully nonparametric method imposing shape constraints such as monotonicity, convexity, and homogeneity. The desired axiomatic properties of distance functions are thus guaranteed to hold. To disentangle technical inefficiency from random noise, Kuosmanen and Johnson (2017) employ the nonparametric kernel deconvolution method suggested by Hall and Simar (2002). If the parametric distributional assumptions on inefficiency and/or noise are imposed, one can combine convex regression with the SFA techniques, as first outlined by Kuosmanen and Kortelainen (2012), who refer to the blended application of nonparametric and parametric techniques as stochastic nonparametric envelopment of data (StoNED).

Thus far, the convex regression and StoNED estimation of distance functions have focused on either DDF by Kuosmanen and Johnson (2017) or a prior transformation on output

---

<sup>2</sup>We refer to global convexity and free disposability axioms in the axiomatic production theory (e.g., Färe and Primont, 1995).

variables by Schaefer and Clermont (2018). DDF is a general and theoretically valid representation of the technology, but its additive structure implies a somewhat limiting additive structure for the composite error term. In econometrics, a multiplicative error term where the size of the error is proportionate to the output is often found useful in applications. While a prior transformation on output variables is a meaningful attempt, the necessary orthogonality condition can not be satisfied in Schaefer and Clermont (2018), as demonstrated below.

The purpose of this paper is to present a theoretically sound approach to model economies of scope in joint production using the convex regression approach. Our first methodological contribution is to develop a radial convex nonparametric least squares (CNLS) formulation to estimate the input distance function with the multiple input-multiple output specification. The orthogonality condition in radial CNLS will be stated and proven. We also extend the proposed radial CNLS approach to estimate output distance functions and the expected inefficiency.

Our second contribution is to compare the finite sample performance of radial CNLS versus the standard DEA and SFA estimators in terms of the input distance function estimation with multiple outputs in the controlled environment of the Monte Carlo study. The simulations reveal the superior performance of the radial CNLS estimator in the joint production setting involving multiple outputs.

The third contribution is to empirically apply the radial CNLS estimator to the Finnish electricity distribution firms' dataset. In contrast to the earlier DDF approach (Kuosmanen and Johnson, 2017) or the CNLS with input requirement function approach (Kuosmanen and Johnson, 2020), the proposed radial CNLS approach does not rely on any prior specification of the direction vector. To reduce the potential overfitting in shadow prices and increase the out-of-sample performance in energy regulation, we introduce the weight restriction constraints to radial CNLS. The results confirm that the input isoquants become more curved, implying better substitution possibilities, if we use two inputs rather than just project in the direction of one input as in the input requirement function. These results are highly relevant for the incentive regulation of electricity distribution firms in the real world.

The rest of this paper is organized as follows. Section 2 introduces the theory on the input distance function and the regression model. Section 3 presents a naive approach to

estimating the input distance function and proposes the radial CNLS approach. The possible extensions are discussed in Section 4. The Monte Carlo study and the empirical application of Finland’s electricity distribution network regulation are demonstrated in Sections 5 and 6. Section 7 concludes this paper with suggested avenues for future research. The formal proofs of Theorems and additional tables and figures are provided in the Online Supplement.

## 2 Theory

### 2.1 Input distance function

Multi-output technology can be generally represented by the production possibility set.

$$\mathcal{T} = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}_+^{m+s} \mid \mathbf{x} \text{ can produce } \mathbf{y}\} \quad (1)$$

where we assume that  $\mathcal{T}$  is a closed and nonempty set satisfying free disposability (monotonicity) and convexity axioms. See Färe and Primont (1995) for further detailed discussions on the axiomatic production theory.

The Shephard’s input distance function  $DI : \mathbb{R}_+^m \times \mathbb{R}_+^s \rightarrow \mathbb{R}_+^1 \cup \{+\infty\}$  is defined as (Shephard, 1970):

$$DI(\mathbf{x}, \mathbf{y}) = \sup\{\phi \mid (\mathbf{x}/\phi, \mathbf{y}) \in \mathcal{T}\} \quad (2)$$

where the construction of input distance function  $DI$  does not depend on behavioral hypotheses such as profit maximization or cost minimization.

Similar to Shephard (1970), we consider  $DI$  as a general functional representation of multi-output technology: If  $DI(\mathbf{x}, \mathbf{y}) = 1$ , then production plan  $(\mathbf{x}, \mathbf{y})$  is technically feasible and efficient. If  $DI(\mathbf{x}, \mathbf{y}) > 1$ , then  $(\mathbf{x}, \mathbf{y})$  is technically feasible but inefficient. If  $DI(\mathbf{x}, \mathbf{y}) < 1$ , then  $(\mathbf{x}, \mathbf{y})$  is technically infeasible. Therefore,  $(\mathbf{x}, \mathbf{y})$  for which  $DI(\mathbf{x}, \mathbf{y}) \geq 1$  are within the production possibility set and  $(\mathbf{x}, \mathbf{y})$  for which  $DI(\mathbf{x}, \mathbf{y}) = 1$  are its efficient subset.

Linear homogeneity of  $DI$  is a fundamental property that does not relate to the properties of  $\mathcal{T}$ , but arises from the definition of  $DI$ . The linear homogeneity implies that

$$DI(\lambda\mathbf{x}, \mathbf{y}) = \lambda DI(\mathbf{x}, \mathbf{y}) \quad \forall \lambda > 0, (\mathbf{x}, \mathbf{y}) \in \mathcal{T} \quad (3)$$

Consider a sample of multiproduct firms where at least some firms produce more than just a single output. The sample can also include firms that specialize in just one or few

products, but in contrast to De Loecker et al. (2016), our empirical strategy does not rely on the existence of such firms. We assume that each firm takes the output demands as given, and there is no noise in the observed outputs  $\mathbf{y} \in \mathbb{R}_+^s$ . The firm managers optimize some meaningful objective functions, e.g., minimizing cost. Note that cost minimization is equivalent to profit maximization when the output demands are taken as given. The observed inputs  $\mathbf{x} \in \mathbb{R}_+^m$  are subject to random perturbances  $\varepsilon_i$ , and the firm manager then adjusts their input demands according to

$$\mathbf{x}_i = \exp(\varepsilon_i)\mathbf{x}^* \quad (4)$$

where  $\mathbf{x}^*$  is the optimal (cost minimizing) input vector.

The linear homogeneity property (3) helps develop the following convex regression model of the input distance function. The value of the input distance function  $DI(\mathbf{x}_i, \mathbf{y}_i)$  is then completely determined by the error term  $\varepsilon_i$  (see Theorem 1).

**Theorem 1.** *If the observed data are generated according to the data generating process (DGP) described by Eq. (4), then the value of the input distance function is equal to*

$$DI(\mathbf{x}_i, \mathbf{y}_i) = \exp(\varepsilon_i)$$

*Proof.* See Appendix A. ■

This result is analogous to Proposition 2 in Kuosmanen and Johnson (2017), adapted to the present setting of  $DI$ . Note that even though the distance function is a deterministic representation of a deterministic technology, its value is a random variable when the observed data are perturbed by random noise and inefficiency.

## 2.2 Regression model

To derive the regression equation, we can utilize the homogeneity property (3) and normalize by  $x_{1i}$  to obtain<sup>3</sup>

$$DI(\mathbf{x}_i/x_{1i}, \mathbf{y}_i) = (1/x_{1i}) \cdot DI(\mathbf{x}_i, \mathbf{y}_i) \quad (5)$$

where  $\mathbf{x}/x_1 = (1, x_2/x_1, \dots, x_m/x_1)'$ . After taking the logs, we have  $\ln x_{1i} = -\ln DI(\mathbf{x}_i/x_{1i}, \mathbf{y}_i) + \ln DI(\mathbf{x}_i, \mathbf{y}_i)$ .

---

<sup>3</sup>Note that any one of the input variables can be selected in the distance function normalization.

Using Theorem 1, we substitute  $\ln DI(\mathbf{x}_i, \mathbf{y}_i)$  with the error term  $\varepsilon_i$  and obtain

$$\ln x_{1i} = -\ln DI(\mathbf{x}_i/x_{1i}, \mathbf{y}_i) + \varepsilon_i \quad (6)$$

A similar regression equation is commonly used in the SFA literature (e.g., Kumbhakar, 2013). The main difference is that we derive the regression equation without imposing any particular functional form for  $DI$ . Note that the parametric functional forms usually employed in SFA tend to violate the theoretical properties of the distance function such as monotonicity and convexity of the output sets. In the SFA literature,  $\varepsilon_i$  consists of asymmetric inefficiency  $u \sim N^+(0, \sigma_u^2)$  and symmetric noise  $v \sim N(0, \sigma_v^2)$ . For simplicity, we here mainly focus on the conventional case of a symmetric error term, but the extension to stochastic frontier setting with an asymmetric composite error term is briefly examined in section 4.2 below.

### 3 Convex regression

#### 3.1 A naive approach

To estimate the regression equation (6) developed in the previous section by convex regression, it would be tempting to formulate the CNLS problem as follows

$$\begin{aligned} \min_{\chi, \alpha, \beta, \gamma, \varepsilon} \quad & \frac{1}{2} \sum_{i=1}^N \varepsilon_i^2 & (7) \\ \text{s.t.} \quad & \ln x_{1i} = -\ln(\chi_i) + \varepsilon_i & \forall i \\ & \chi_i = \alpha_i + \beta'_i(\mathbf{x}_i/x_{1i}) - \gamma'_i \mathbf{y}_i & \forall i \\ & \chi_i \leq \alpha_h + \beta'_h(\mathbf{x}_i/x_{1i}) - \gamma'_h \mathbf{y}_i & \forall i, h \\ & \beta_i \geq \mathbf{0}, \gamma_i \geq \mathbf{0}, \chi_i \geq 0 & \forall i \end{aligned}$$

The first two constraints of (7) are used to characterize the regression equation (6). The decision variable  $\chi_i$  denotes the distance function  $DI(\mathbf{x}_i/x_{1i}, \mathbf{y}_i)$ . The second constraint can thus be interpreted as a multivariate linear regression among the normalized inputs, outputs, and distance function. The third constraint imposes the concavity on the distance function. The last set of constraints guarantees the monotonicity of the distance function.

The sample orthogonality conditions of the CNLS estimator (7) are formally stated in the following Theorem 2.

**Theorem 2.** *The CNLS residuals obtained as the optimal solution to problem (7) satisfy the following conditions*

$$\begin{aligned} \sum_{i=1}^n \frac{\hat{\varepsilon}_i}{\hat{\chi}_i} &= 0; \\ \sum_{i=1}^n \left( \frac{x_{ik}}{x_{1i}} \frac{\hat{\varepsilon}_i}{\hat{\chi}_i} - \eta_{ik} \right) &= 0, \end{aligned}$$

where  $\eta \geq 0$  denotes the shadow price of the monotonicity constraint.

*Proof.* See Appendix A. ■

The first equation in Theorem 2 reveals that the CNLS residuals do not necessarily sum to zero, but the residuals scaled by the predicted  $x_1$  must sum to zero. The second equation can be interpreted as the orthogonality condition with respect to inputs. Note that the usual orthogonality condition immediately fails when the monotonicity constraint is binding at least for one observation (i.e.,  $\eta_{ik} > 0$ , for some  $i, k$ ). If we relax monotonicity (i.e.,  $\eta_{ik} = 0, \forall i, k$ ) or assume the constant return to scale, then the usual orthogonality still does not hold, but it would hold “approximately” with respect to the scaled residuals  $\hat{\varepsilon}_i/\hat{\chi}_i$ .

In the parametric regression, the linear homogeneity of  $DI$  is satisfied by construction because the regression residuals  $\hat{\varepsilon}$  satisfy the following sample orthogonality condition

$$\sum_{i=1}^n (\mathbf{x}_i/x_{1i}) \cdot \hat{\varepsilon}_i = 0$$

which is the sample counterpart of the population orthogonality condition

$$\mathbb{E}[(\mathbf{x}_i/x_{1i}) \cdot \varepsilon_i] = 0.$$

This guarantees the equiproportionate radial orientation of  $DI$  in the parametric models. However, but as noted above, the CNLS residuals do not generally satisfy these orthogonality conditions by construction. Therefore, problem (7) is henceforth referred to as a naive CNLS, and we need to find a way of imposing the orthogonality for the input ratios  $(\mathbf{x}_i/x_{1i})$ .

We note that Schaefer and Clermont (2018) propose an extended CNLS approach to estimate the ray production function in the multiple input-multiple output specification. This approach relies on the polar coordinate transformation to introduce the angle-related variables and generates a “multiple input-single output” model (i.e., a prior transformation

on output variables). While it is a meaningful attempt to model multiple outputs in the framework of CNLS, similar to problem (7), the proposed formulation by Schaefer and Clermont (2018) does not impose the orthogonality conditions and therefore fails to satisfy the linear homogeneity property of the radial distance functions.

### 3.2 Proposed approach

To gain intuition, consider first the special case of two inputs. We utilize the homogeneity property and normalize inputs by the Cobb-Douglas function  $(x_{1i}^{1-d}x_{2i}^d)$ ,  $0 < d < 1$ , to obtain

$$DI(\mathbf{x}_i/(x_{1i}^{1-d}x_{2i}^d), \mathbf{y}_i) = (x_{1i}^{d-1}x_{2i}^{-d}) \cdot DI(\mathbf{x}_i, \mathbf{y}_i) \quad (8)$$

where we can use any arbitrary parameter value  $d$  (e.g.,  $d = 0.5$ ). We will later ensure that the specific choice of  $d$  does not matter.

Taking logs and reorganizing, we now have the following regression equation

$$\ln x_{1i} = -\ln DI(\mathbf{x}_i/(x_{1i}^{1-d}x_{2i}^d), \mathbf{y}_i) + d(\ln x_{1i} - \ln x_{2i}) + \varepsilon_i \quad (9)$$

To estimate the equation (9), we propose the following radial CNLS approach to model multiple input-multiple output setting

$$\begin{aligned} \min_{\chi, \alpha, \beta, \gamma, \delta, \varepsilon} \quad & \frac{1}{2} \sum_{i=1}^N \varepsilon_i^2 & (10) \\ \text{s.t.} \quad & \ln x_{1i} = -\ln(\chi_i) + \delta(\ln x_{1i} - \ln x_{2i}) + \varepsilon_i & \forall i \\ & \chi_i = \alpha_i + \beta'_i(\mathbf{x}_i/(x_{1i}^{1-d}x_{2i}^d)) - \gamma'_i\mathbf{y}_i & \forall i \\ & \chi_i \leq \alpha_h + \beta'_h(\mathbf{x}_i/(x_{1i}^{1-d}x_{2i}^d)) - \gamma'_h\mathbf{y}_i & \forall i, h \\ & \beta_i \geq \mathbf{0}, \gamma_i \geq \mathbf{0}, \chi_i \geq 0 & \forall i \end{aligned}$$

where problem (10) is a semi-nonparametric partial linear formulation building on Johnson and Kuosmanen (2011, 2012). Similar to problem (7), the first two sets of constraints to represent the regression equation (9), the last two sets of constraints impose the concavity and monotonicity on the input distance function, respectively.

In contrast to parameter  $d$  that we simply postulate in Eq. (8),  $\delta$  is a decision variable to be estimated. Theorem 3 shows that the semi-nonparametric partial linear formulation satisfies the orthogonality condition and hence  $\delta$  can be a free decision variable in problem (10).

**Theorem 3.** *The radial CNLS residuals obtained as the optimal solution to problem (10) satisfy the following sample orthogonality condition*

$$\sum_{i=1}^n (\ln x_{1i} - \ln x_{2i}) \cdot \hat{\varepsilon}_i = 0$$

*Proof.* See Appendix A. ■

Theorem 3 shows that problem (10) effectively enforces the equiproportionate radial orientation, in other words, the linear homogeneity, which is an essential property of  $DI$ . We thus can consistently estimate the regression equation (9) using the nonparametric radial CNLS formulation (10).

Regarding the choice of parameter  $d$ , it is clearly possible or perhaps even likely that the coefficient  $\delta$  obtained as the optimal solution to problem (10) differs from our arbitrary specification of  $d$ . Fortunately, it is unnecessary to guess the optimal value of  $d$  beforehand. If we solve the problem (10) for multiple different values of  $d$ , it is easy to verify that the parameter estimates of  $\delta$  change, however, the residuals  $\hat{\varepsilon}_i$  remain exactly the same irrespective of the choice of  $d$ . It turns out that  $\delta$  is an over-identified parameter value that does not influence the optimal solution. More specifically, we need to consider  $-\ln(\chi_i) + \delta(\ln x_{1i} - \ln x_{2i})$  as the estimate of  $DI$ , not only the nonparametric part  $-\ln(\chi_i)$ . Note that the sum of two concave functions is itself a concave function. As the value of parameter  $\delta$  changes, so does that of the nonparametric part, but their sum remains the same. This can be referred to as partial identification in the present context (see, e.g., Manski, 2009; Tamer, 2010).

To find a unique estimate of  $DI$  in any given point  $(\mathbf{x}_i, \mathbf{y}_i)$ , we can solve the following linear programming (LP) problem, directly analogous to the multiplier formulation of DEA

$$\begin{aligned} \widehat{DI}(\mathbf{x}_i, \mathbf{y}_i) &= \max \boldsymbol{\gamma}' \mathbf{y}_i & (11) \\ \text{s.t. } \boldsymbol{\beta}' \exp(\hat{\varepsilon}_i) \mathbf{x}_i &= 1 & \forall i \\ \alpha_i + \boldsymbol{\beta}' \exp(\hat{\varepsilon}_h) \mathbf{x}_h - \boldsymbol{\gamma}' \mathbf{y}_h &\geq 0 & \forall i, h \\ \boldsymbol{\gamma} \geq \mathbf{0}, \boldsymbol{\beta} \geq \mathbf{0} & & \forall i \end{aligned}$$

The adjustment by the CNLS residuals  $\hat{\varepsilon}_i$  of problem (10) effectively projects all observations to the estimated production frontier (in this case the average practice technology).

The optimal solution to this LP problem provides the shadow prices  $\boldsymbol{\gamma}$  and  $\boldsymbol{\beta}$  associated with the minimum extrapolation technology.

To summarize, we stress that variable  $\delta$  is key to ensure that the sample orthogonality condition holds (see Theorem 3), so it needs to be included in the estimation of the model (10). But having solved (10) to obtain  $\hat{\varepsilon}_i$ , we proceed to remove it by solving the LP problem (11). Therefore, our final estimate of DI does not include or depend on variable  $\delta$ .

## 4 Extensions

In this section, we consider the general case with more than two inputs, connect with the efficiency analysis, and estimate the output distance function.

### 4.1 More than two inputs

To generalize the results of the previous section, consider the general case of  $M$  inputs (i.e.,  $M > 2$ ). Let us normalize  $DI$  by the Cobb-Douglas function

$$(x_{1i}^{1/M} x_{2i}^{1/M} \dots x_{Mi}^{1/M}) = (x_{1i}^{1-(M-1)/M} x_{2i}^{1/M} \dots x_{Mi}^{1/M}).$$

In this case, the regression equation (6) becomes

$$\ln x_{1i} = -\ln DI(\mathbf{x}_i / (x_{1i}^{1/M} x_{2i}^{1/M} \dots x_{Mi}^{1/M}), \mathbf{y}_i) + \frac{1}{M} \sum_{m=2}^M (\ln x_{1i} - \ln x_{mi}) + \varepsilon_i \quad (12)$$

and the radial CNLS estimator can be stated as

$$\begin{aligned} \min_{\chi, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \delta, \varepsilon} \quad & \frac{1}{2} \sum_{i=1}^N \varepsilon_i^2 & (13) \\ \text{s.t.} \quad & \ln x_{1i} = -\ln(\chi_i) + \sum_{m=2}^M \delta_m (\ln x_{1i} - \ln x_{mi}) + \varepsilon_i \quad \forall i \\ & \chi_i = \alpha_i + \boldsymbol{\beta}'_i(\mathbf{x}_i / (x_{1i}^{1/M} x_{2i}^{1/M} \dots x_{Mi}^{1/M})) - \boldsymbol{\gamma}'_i \mathbf{y}_i \quad \forall i \\ & \chi_i \leq \alpha_h + \boldsymbol{\beta}'_h(\mathbf{x}_i / (x_{1i}^{1/M} x_{2i}^{1/M} \dots x_{Mi}^{1/M})) - \boldsymbol{\gamma}'_h \mathbf{y}_i \quad \forall i, h \\ & \boldsymbol{\beta}_i \geq \mathbf{0}, \boldsymbol{\gamma}_i \geq \mathbf{0}, \chi_i \geq 0 \quad \forall i \end{aligned}$$

This formulation is a more general CNLS estimation of the input distance function with multiple outputs. That is, problem (10) is a special case of problem (13), where  $M = 2$ .

Having solved the radial CNLS problem (13), the same LP problem (11) can be applied to obtain the unique estimate of  $DI$ .

## 4.2 Efficiency analysis

Given the residuals  $\hat{\varepsilon}_i$  estimated by radial CNLS (10), it is possible to estimate the expected inefficiency using the nonparametric kernel deconvolution or if one imposes further parametric distributional assumptions by quasi-likelihood or the method of moments. Following Kuosmanen and Kortelainen (2012) and Kuosmanen and Johnson (2017), we in this section briefly summarize how to estimate the unconditional expected inefficiency  $\mathbb{E}(u)$  and the conditional expected inefficiency  $\mathbb{E}(u_i | \hat{\varepsilon}_i)$ .

Nonparametric kernel deconvolution is a full nonparametric estimation of the expected inefficiency  $\mathbb{E}(u) = \mu$ . For the input distance function, the residuals  $\hat{\varepsilon}_i$  are the consistent estimators of  $\hat{\varepsilon}_i = \mu + (v_i - u_i)$ . The density function of  $\hat{\varepsilon}_i$  is defined as

$$\hat{f}_{\hat{\varepsilon}}(x) = (nh)^{-1} \sum_{i=1}^n K\left(\frac{x - \hat{\varepsilon}_i}{h}\right)$$

where  $K(\cdot)$  is a standard kernel function,  $h$  is the bandwidth,  $x$  is the projected data of  $\hat{\varepsilon}$ , and  $n$  is the number of observations. The first derivative of the density function of the composite error term ( $f'_{\hat{\varepsilon}}$ ) is proportional to that of the inefficiency term ( $f'_u$ ) in the neighborhood of  $\mu$  (Hall and Simar, 2002). Therefore,  $\hat{\mu} = \arg \max_{x \in C} (\hat{f}'_{\hat{\varepsilon}}(x))$  provides a nonparametric estimation of the expected inefficiency  $\mu$ , where  $C$  is a closed interval in the right tail of  $f_{\varepsilon}(\cdot)$ .

Alternatively, the method of moments and the quasi-likelihood approaches for the residual decomposition build upon additional parametric distributional assumptions (e.g., the half-normal inefficiency and normal noise). For the method of moments approach, we first calculate the second and third central moments of the residuals  $\hat{\varepsilon}_i$  by

$$\hat{M}_2 = \sum_{i=1}^n (\hat{\varepsilon}_i - \bar{\varepsilon})^2 / n$$

$$\hat{M}_3 = \sum_{i=1}^n (\hat{\varepsilon}_i - \bar{\varepsilon})^3 / n$$

Under the parametric distributional assumptions, we have  $M_3 = (2/\pi)^{1/2}(1 - 4/\pi)\sigma_u^2$  and

$M_2 = \sigma_v^2 + \sigma_u^2(\pi - 2)/\pi$ . Given the estimated  $\hat{M}_2$  and  $\hat{M}_3$ , we can calculate  $\sigma_u$  and  $\sigma_v$  by

$$\hat{\sigma}_u = \sqrt[3]{\frac{\hat{M}_3}{\left(\sqrt{\frac{2}{\pi}}\right)\left[1 - \frac{4}{\pi}\right]}}$$

$$\hat{\sigma}_v = \sqrt[2]{\hat{M}_2 - \left[\frac{\pi - 2}{\pi}\right]\hat{\sigma}_u^2}$$

The quasi-likelihood estimation is an alternative to compute  $\hat{\sigma}_u$  and  $\hat{\sigma}_v$ . We apply the standard maximum likelihood method to estimate the following likelihood function

$$\ln L(\lambda) = -n \ln(\hat{\sigma}) + \sum \ln \Phi \left[ \frac{-\varepsilon_i \lambda}{\hat{\sigma}} \right] - \frac{1}{2\hat{\sigma}^2} \sum \varepsilon_i^2$$

where

$$\varepsilon_i = \hat{\varepsilon}_i - (\sqrt{2\lambda\hat{\sigma}})/[\pi(1 + \lambda^2)]^{1/2}$$

$$\hat{\sigma} = \left\{ \frac{1}{n} \sum (\hat{\varepsilon}_i)^2 / \left[ 1 - \frac{2\lambda^2}{\pi(1 + \lambda^2)} \right] \right\}$$

The likelihood function  $\ln L(\lambda)$  consists of a single parameter  $\lambda$ .  $\Phi$  denotes the cumulative distribution function of normal distribution. After obtaining  $\hat{\sigma}_u$  by these two parametric approaches, we can compute the expected inefficiency,  $\mu = (2/\pi)^{1/2}\hat{\sigma}_u$ .

Jondrow et al. (1982) propose a widely applied estimator for calculating the conditional expected inefficiency  $\mathbb{E}[u_i | \hat{\varepsilon}_i]$ . Specifically, for the input distance function, the conditional expected value of inefficiency  $\mathbb{E}[u_i | \varepsilon_i]$  is formulated as

$$\mathbb{E}[u_i | \hat{\varepsilon}_i] = \mu_{*i} + \sigma_* \left[ \frac{\phi(-\mu_{*i}/\sigma_*)}{1 - \Phi(-\mu_{*i}/\sigma_*)} \right] = \sigma_* \left[ \frac{\phi(\hat{\varepsilon}_i \lambda / \sigma)}{1 - \Phi(\hat{\varepsilon}_i \lambda / \sigma)} - \frac{\hat{\varepsilon}_i \lambda}{\sigma} \right]$$

where  $\mu_{*i} = -\hat{\varepsilon}_i \sigma_u^2 / \sigma^2$ ,  $\sigma_*^2 = \sigma_u^2 \sigma_v^2 / \sigma^2$ ,  $\lambda = \sigma_u / \sigma_v$ , and  $\sigma^2 = \sigma_u^2 + \sigma_v^2$ .  $\phi$  and  $\Phi$  are the standard normal density function and its cumulative distribution function, respectively. Note that the conditional mean inefficiency in Jondrow et al. (1982) could be further extended to conditional quantile inefficiency.

Finally, some practical deficiencies in those decomposition methods are worth noting. The parametric approaches require a priori distribution assumptions on noise and inefficiency, which may seem unrealistic in real-world applications. Moreover, the method of moments

method may yield a positive third central moment in estimating the production function.<sup>4</sup> Regarding the nonparametric kernel density estimation, it may not be suitable when the variance of random noise is relatively large; the estimator is inconsistent when the variance of the random noise does not approach zero (Hall and Simar, 2002). In addition, the kernel density method estimates the unconditional expectation of inefficiency: further distribution assumptions are still required if one wishes to estimate firm-level inefficiencies (Dai, 2016).

### 4.3 Output distance function

Consider now the general output distance function

$$DO(\mathbf{x}, \mathbf{y}) = \inf\{\phi \mid (\mathbf{x}, \mathbf{y}/\phi) \in \mathcal{T}\} \quad (14)$$

where  $DO(\mathbf{x}, \mathbf{y})$  is homogeneous of degree 1 in  $\mathbf{y}$ , implying that

$$DO(\mathbf{x}, \lambda \mathbf{y}) = \lambda DO(\mathbf{x}, \mathbf{y}) \quad \forall \lambda > 0, (\mathbf{x}, \mathbf{y}) \in \mathcal{T}.$$

Given the case of  $S$  outputs (i.e.,  $S \geq 2$ ) and the linear homogeneity, similar to equation (12), the regression equation can be formulated as

$$\ln y_{1i} = -\ln DO(\mathbf{x}_i, \mathbf{y}_i / (y_{1i}^{\frac{1}{S}} y_{2i}^{\frac{1}{S}} \dots y_{Si}^{\frac{1}{S}})) + \frac{1}{S} \sum_{s=2}^S (\ln y_{1i} - \ln y_{si}) + \varepsilon_i \quad (15)$$

We then estimate the output distance function using the following radial CNLS formulation

$$\begin{aligned} \min_{\chi, \alpha, \beta, \gamma, \delta, \varepsilon} \quad & \frac{1}{2} \sum_{i=1}^N \varepsilon_i^2 & (16) \\ \text{s.t.} \quad & \ln y_{1i} = -\ln(\chi_i) + \sum_{s=2}^S \delta_s (\ln y_{1i} - \ln y_{si}) + \varepsilon_i \quad \forall i \\ & \chi_i = \alpha_i + \gamma'_i(\mathbf{y}_i / (y_{1i}^{\frac{1}{S}} y_{2i}^{\frac{1}{S}} \dots y_{Si}^{\frac{1}{S}})) - \beta'_i \mathbf{x}_i \quad \forall i \\ & \chi_i \geq \alpha_h + \gamma'_h(\mathbf{y}_i / (y_{1i}^{\frac{1}{S}} y_{2i}^{\frac{1}{S}} \dots y_{Si}^{\frac{1}{S}})) - \beta'_h \mathbf{x}_i \quad \forall i, h \\ & \beta_i \geq \mathbf{0}, \gamma_i \geq \mathbf{0}, \chi_i \geq 0 \quad \forall i \end{aligned}$$

---

<sup>4</sup>To address this question, we could either apply a bootstrapping method when using the standard distribution (Simar and Wilson, 2010) or consider alternative distributions that allow for positive skewness (Almanidis and Sickles, 2012).

where the first two constraints characterize the regression model (15). In contrast to problem (10), the third set of constraints of problem (16) impose the convexity of the output distance function. The last set of constraints also ensures the monotonicity of the output distance function.

The main challenge here is that the output isoquants of the nonparametric part  $-\ln(\chi_i)$  are concave by construction, whereas those of the parametric part  $\sum_{s=2}^S \delta_s (\ln y_{1i} - \ln y_{si})$  are convex by default. In contrast to the input distance function where both parts have the same curvature, in this case, there is no guarantee that the estimated output sets satisfy convexity.

If the data are well-behaved, the nonparametric part is flexible enough to offset the wrong curvature of the Cobb-Douglas part. In other words, the formulation (16) could still satisfy the convexity of the output sets, however, this cannot be guaranteed always to hold. Note that possible violations of convexity of the output sets can always be fixed in the next step, where we apply the LP problem (11) to estimate the output distance function. We leave a more detailed examination of the output distance function as an interesting avenue for future research.

## 5 Monte Carlo study

In the Monte Carlo study, our main objective is to investigate the finite sample performance of the proposed radial CNLS approach and compare it with the naive CNLS approach and other commonly seen deterministic and stochastic frontier estimation approaches in estimating the input distance function.

### 5.1 Setup

We generate the two input-two output production datasets by using the following two sets of DGPs (Färe et al., 2010)

$$\begin{aligned} \text{DGP I : } P(\mathbf{x}) &= \{(y_1, y_2) : y_2 = \beta_0 + \beta_1 y_1 + \beta_2 y_1^2 + \beta_3 y_1^3 + \beta_4 y_1^4 + x_1^{0.9} x_2^{0.8} + \varepsilon_i\} \\ \text{DGP II : } P(\mathbf{x}) &= \{(y_1, y_2) : y_2 = \exp(\beta_0 + \beta_1 \ln(y_1) + \beta_2 [\ln(y_1)]^2 + \beta_3 [\ln(y_1)]^3 + \\ &\quad \beta_4 [\ln(y_1)]^4 + x_1^{0.9} x_2^{0.8}) + \varepsilon_i\} \end{aligned}$$

where the inputs  $\mathbf{x}$  in each DGP are independently and randomly drawn from the uniform distribution,  $U[0, 1]$ , and the error term  $\varepsilon$  has three different specifications:  $\varepsilon = v$  (i.e., only noise),  $\varepsilon = -u$  (i.e., only inefficiency), and  $\varepsilon = v - u$  (i.e., both noise and inefficiency). We generate noise  $v \sim N(0, \sigma_v^2)$  and inefficiency  $u \sim N^+(0, \sigma_u^2)$ .

The first output  $y_1$  in DGP I is randomly generated from the following two different gamma distributions

$$\text{Type-A : } y_1 \sim \Gamma(\alpha = 5, \beta = 0.5);$$

$$\text{Type-B : } y_1 \sim \Gamma(\alpha = 18, \beta = 0.25)$$

where  $\alpha$  and  $\beta$  are the shape and rate parameters of the gamma distribution. For DGP II, the first output  $y_1$  is drawn from  $U[e^{0.7}, e^{1.4}]$ .

Given the different true parameter combinations ( $\beta_k, k = 0, \dots, 4$ ) in Table 1, DGPs I and II consist of three polynomial and translog technologies, respectively. We thus consider 9 models with DGPs I and II in the experiments. Note that Model 3 in each DGP makes the true functions in terms of outputs more concave than the other two models, and Type-B in DGP I generates a more balanced outputs dataset (i.e., the value of generated  $y_1$  is close to that of  $y_2$ ).

**Table 1.** The true parameters used in DGPs I and II.

	DGP I			DGP II		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
$\beta_0$	10.70	10.10	9.60	3.000	2.845	2.690
$\beta_1$	-0.91	-0.72	-0.54	-3.500	-3.400	-3.300
$\beta_2$	$0.50 \times 10^{-5}$	$0.50 \times 10^{-4}$	$0.10 \times 10^{-2}$	3.900	4.000	4.100
$\beta_3$	$0.10 \times 10^{-4}$	$0.10 \times 10^{-3}$	$0.10 \times 10^{-2}$	-1.500	-1.475	-1.415
$\beta_4$	$-0.45 \times 10^{-3}$	$-0.12 \times 10^{-2}$	$-0.24 \times 10^{-2}$	-0.140	-0.220	-0.330

*Source:* Färe et al. (2010).

In all experiments that follow, we resort to the Python/pyStoNED package (Dai et al., 2021) to solve the CNLS and DEA models and the Python/pySFA package to estimate the SFA models.<sup>5</sup> Note that the linear and nonlinear programming problems are solved by the off-the-shelf solvers MOSEK (9.3.11) and KNITRO (13.2), respectively. All experiments are

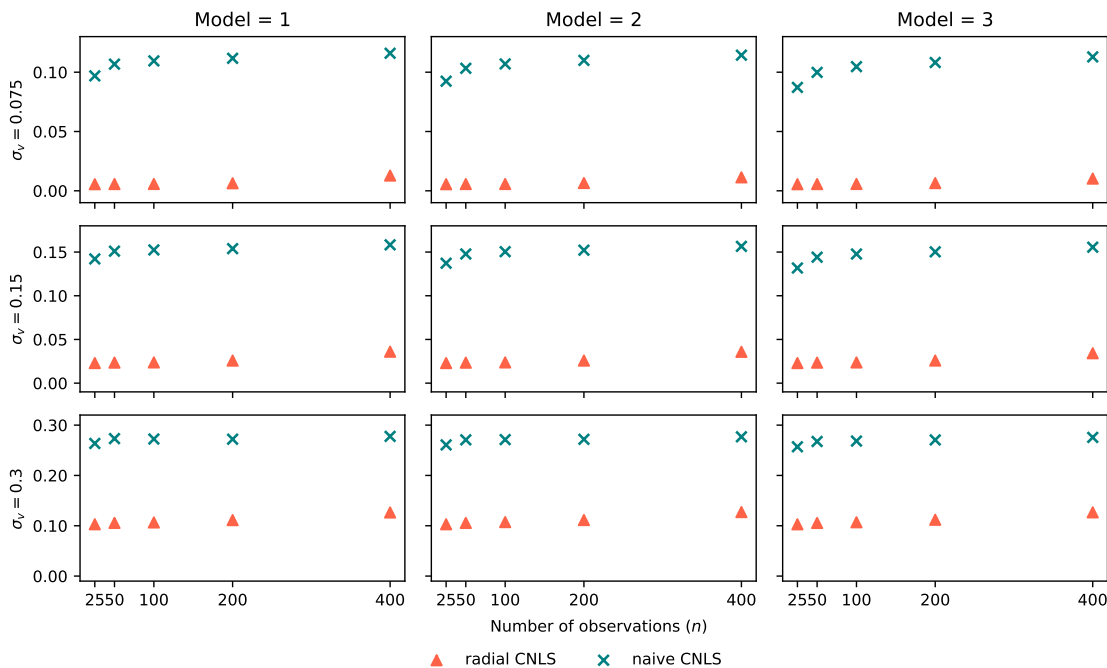
<sup>5</sup>A Python Package for Stochastic Frontier Analysis (pySFA): <https://github.com/gEAPA/pySFA>.

run on Finland’s high-performance computing cluster Puhti with Xeon @2.1 GHz processors, 2 CPUs, and 5 GB of RAM per task.

## 5.2 Experiment 1

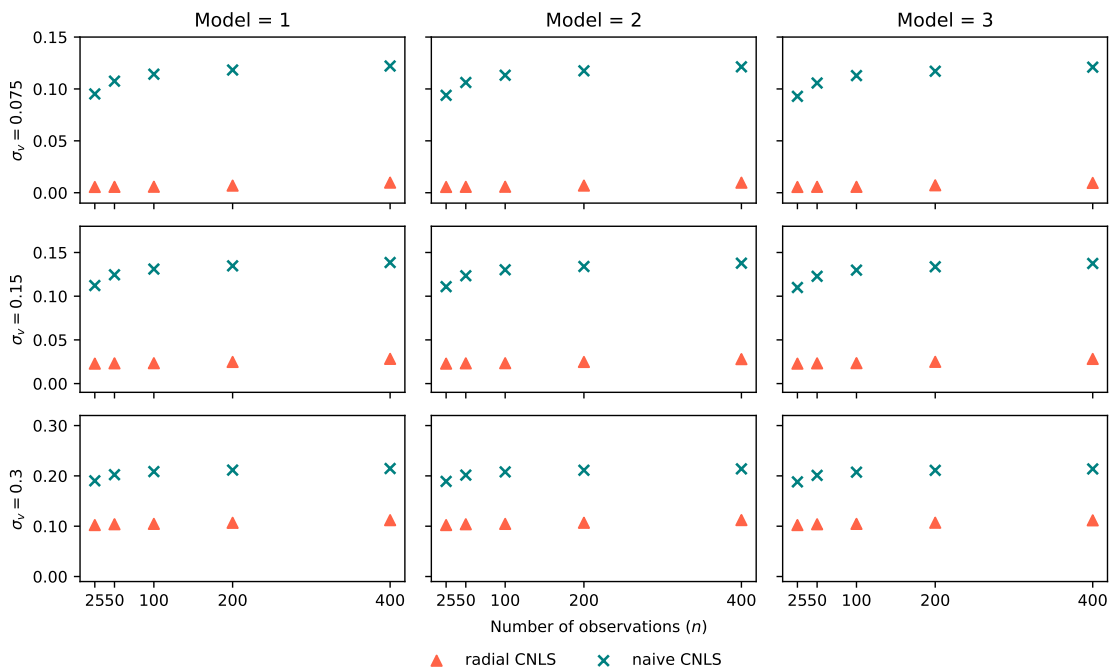
In experiment 1 we compare the performance of the proposed radial CNLS formulations (10) and (13) with the naive CNLS formulation (7) in the absence of inefficiency (i.e.,  $\varepsilon = v$ ). We thus consider 135 scenarios with different numbers of observations,  $n \in \{25, 50, 100, 200, 400\}$ , and the variation of noise levels,  $\sigma_v \in \{0.075, 0.15, 0.3\}$ . To evaluate their finite sample performance, each scenario is run 500 times to calculate mean square error (MSE) and mean absolute deviation (MAD) (see the Online Supplement) in terms of input distance function estimation.

Figs. 1 and 2 depict the MSE results of radial CNLS and naive CNLS in DGP I with Type-A and DGP II. Radial CNLS outperforms naive CNLS in all scenarios considered, suggesting that radial CNLS is a more efficient method for estimating the input distance function. In contrast to arbitrarily choosing the input, normalizing the input distance function can satisfy the orthogonality conditions, which probably helps fit the true function with smaller expected errors. Several other interesting findings are also observed.



**Fig. 1.** MSE results of radial CNLS and naive CNLS in DGP I with Type-A.

First, the curvature of the generated production function is likely to affect the performance of estimators. For instance, compared to the other two models in DGPs I and II, Model 3 has the smallest MSE values due to that the generated production function is more concave in terms of the outputs. Second, the MSE values increase as the noise becomes large. It is clearly evident from Figs. 1 and 2 that the estimation of the input distance function becomes worse when  $\sigma_v$  increases. Third, compared to DGP I, DGP II has a smaller MSE difference between the two estimators. This is because the generated dataset by DGP II is more balanced (see Figure 1 in Färe et al., 2010), which dampens the disadvantage of the naive CNLS approach. Further, Figs.B1 and B2 in Appendix B demonstrate the MAD results of radial CNLS and naive CNLS in two sets of DGPs, showing similar findings as in the MSE comparison and supporting the main conclusion of experiment 1.



**Fig. 2.** MSE results of radial CNLS and naive CNLS in DGP II.

Given the potential influence of the curvature of the true output isoquant, we further compare two CNLS approaches based on DGP I with Type-B, where a more balanced dataset in terms of outputs is generated. As shown in Table 2, the MSE values of naive CNLS are always larger than those of radial CNLS. As a robustness check, the smaller variances of radial CNLS over naive CNLS also confirm its better performance.

**Table 2.** DGP I with Type-B ( $n = 400$ ).

Model	$\sigma_v$	MSE		MAD		Variance	
		radial CNLS	naive CNLS	radial CNLS	naive CNLS	radial CNLS	naive CNLS
1	0.075	0.0122	0.1147	0.0725	0.2772	0.0121	0.1061
	0.15	0.0358	0.1570	0.1375	0.3227	0.0355	0.1478
	0.3	0.1266	0.2770	0.2662	0.4227	0.1245	0.2711
2	0.075	0.0101	0.1124	0.0684	0.2730	0.0099	0.0874
	0.15	0.0331	0.1547	0.1337	0.3192	0.0325	0.1632
	0.3	0.1242	0.2749	0.2634	0.4206	0.1226	0.2693
3	0.075	0.0088	0.1104	0.0656	0.2692	0.0088	0.1028
	0.15	0.0310	0.1530	0.1309	0.3165	0.0311	0.1451
	0.3	0.1222	0.2730	0.2618	0.4187	0.1203	0.2679

### 5.3 Experiment 2

We proceed to compare the performance among radial CNLS, DEA, and SFA. Following Schaefer and Clermont (2018), we consider 180 scenarios in experiment 2 with different numbers of observations,  $n \in \{25, 50, 100, 200, 400\}$ , the noise-to-signal ratios,  $\rho_{nts} \in \{0, 0.5, 1, 2\}$ , where  $\rho_{nts} = \sigma_v/\sigma_u$ , and  $\sigma_u = 0.15$ . Each scenario is also run 500 times to compute the MSE and MAD metrics. Note that the input-oriented DEA model with variable returns to scale and SFA with Cobb Douglas function (SFA-CD) and translog function (SFA-TL) specifications are considered in the benchmarking procedures.

Tables 3 and B1 in Appendix B demonstrate the performance comparison of radial CNLS, DEA, and SFA with  $n = 400$ . This experiment further confirms the superiority of the proposed radial CNLS approach in estimating input distance function compared to other available methods. Specifically, the radial CNLS approach has the lowest MSE and MAD values in all scenarios. Notably, in the case of the DGPs with inefficiency only, radial CNLS also works best compared to its counterparts. The MSE and MAD values of the SFA-CD approach are larger than those of others and increase much more sharply at the high noise-to-signal ratio of  $\rho_{nts} = 2$  (cf. Schaefer and Clermont, 2018). This is because when estimating the Cobb-Douglas function using SFA, the considered DGPs are misspecified, especially in polynomial technologies. Moreover, the parametric functional forms used in SFA are likely to violate the theoretical properties of the distance function (e.g., monotonicity and convexity of the output sets), deteriorating the performance of SFA. The results also show that radial CNLS and DEA are relatively robust to polynomial and translog technologies.

Furthermore, as expected, the values of MSE and MAD increase as the noise increases (see, e.g., Henningsen et al., 2015; Schaefer and Clermont, 2018; Ahn et al., 2023).

**Table 3.** Performance of radial CNLS, DEA, and SFA with  $n = 400$ .

DGP	Model	$\rho$	MSE				MAD			
			CNLS	DEA	SFA-TL	SFA-CD	CNLS	DEA	SFA-TL	SFA-CD
I (Type-A)	1	0	0.013	0.083	1.189	44.739	0.082	0.218	0.479	1.531
		0.5	0.017	0.140	2.294	48.761	0.100	0.317	0.589	1.552
		1	0.032	0.222	2.443	16.478	0.140	0.430	0.709	1.364
		2	0.099	0.380	8.255	64.802	0.240	0.574	1.145	2.003
	2	0	0.012	0.075	6.225	60.838	0.082	0.205	0.642	1.923
		0.5	0.017	0.127	6.404	60.657	0.099	0.298	0.747	2.000
		1	0.033	0.209	3.853	12.270	0.142	0.414	0.846	1.332
		2	0.099	0.369	3.986	60.876	0.241	0.566	1.025	1.883
	3	0	0.012	0.069	1.036	59.532	0.081	0.194	0.536	1.935
		0.5	0.017	0.117	1.656	61.946	0.099	0.283	0.671	2.089
		1	0.032	0.197	2.092	112.842	0.140	0.400	0.790	2.263
		2	0.099	0.353	2.159	65.421	0.240	0.553	0.812	1.663
II	1	0	0.013	0.077	0.336	6.708	0.083	0.235	0.339	0.852
		0.5	0.017	0.083	0.463	6.344	0.100	0.246	0.374	0.827
		1	0.032	0.101	0.434	2.897	0.140	0.272	0.378	0.757
		2	0.095	0.175	0.819	5.775	0.237	0.348	0.518	0.929
	2	0	0.013	0.072	0.339	18.427	0.082	0.226	0.329	0.952
		0.5	0.017	0.078	0.403	4.658	0.100	0.236	0.346	0.781
		1	0.032	0.095	0.429	6.117	0.140	0.262	0.371	0.807
		2	0.094	0.168	0.690	6.425	0.237	0.340	0.495	0.925
	3	0	0.013	0.069	0.378	6.503	0.083	0.221	0.345	0.880
		0.5	0.018	0.075	0.348	5.897	0.102	0.231	0.334	0.832
		1	0.032	0.092	0.365	5.813	0.140	0.258	0.352	0.822
		2	0.095	0.164	0.755	10.009	0.238	0.336	0.503	0.989

## 6 Application

The most significant real-world application of CNLS has thus far been incentive regulation of electricity distribution firms (Kuosmanen, 2012). This literature introduces the CNLS estimator that combines the virtues of DEA and SFA to the Finnish electricity distribution network regulation. But most existing applications of CNLS or other frontier estimation techniques focus on either variable cost or total cost and hence have demonstrable shortcomings (Kuosmanen and Johnson, 2020). To consider variable cost and fixed cost simultaneously in the benchmark regulation, Kuosmanen and Johnson (2017) propose a new CNLS approach with DDF (i.e., DDF CNLS), and Kuosmanen and Johnson (2020) and Kuosmanen et al.

(2022) further develop a CNLS approach to incorporate the input requirement function (i.e., IRF CNLS).

However, the DDF CNLS approach requires a prespecified direction vector, and the IRF CNLS approach needs to project in the direction of one input. The estimates by DDF CNLS or IRF CNLS are not invariant to the direction vector or the input choice.<sup>6</sup> An appealing feature of the radial CNLS is that it treats all inputs symmetrically, using a proportionate scaling.

## 6.1 Model specification

Following Kuosmanen and Johnson (2020), we consider the following the multiplicative cost frontier model with contextual variables

$$\mathbf{x}_{i,t} = C(\mathbf{y}_{i,t}, \mathbf{b}_{i,t}) \exp(\boldsymbol{\lambda} \mathbf{z}_{i,t} + \varepsilon_{i,t}) \quad (17)$$

where  $C$  is a non-decreasing convex conical cost function.  $\mathbf{x}$ ,  $\mathbf{y}$ , and  $\mathbf{b}$  are inputs, desirable outputs, and undesirable outputs, respectively.  $\mathbf{z}$  represents a set of contextual variables, and  $\boldsymbol{\lambda}$  denotes the associated marginal coefficients.  $\varepsilon$  is a composite error term involving inefficiency and random effects. Accordingly, the input, output, and contextual variables of the cost frontier model (17) are specified as follows.

### Inputs ( $\mathbf{x}$ )

$x_1$  = Fixed cost, capital stock (regulatory asset value, NKA, €)

$x_2$  = Variable cost, controllable operational expenditure (KOPEX, €)

### Outputs

Desirable outputs ( $\mathbf{y}$ )

$y_1$  = Energy supply (GWh, weighted by voltage)

$y_2$  = Network length (km)

$y_3$  = Number of use points

Undesirable output ( $\mathbf{b}$ )

$b_1$  = Outages (hedonic damage cost, €)

### Contextual variables ( $\mathbf{z}$ )

---

<sup>6</sup>Note that econometric estimation will not give the same results in all specifications simply because of the fact the dependent (endogenous) variable is not the same in each specification (Parmeter and Kumbhakar, 2014).

$z_1 =$  Connection points / Use points

$z_2 =$  Energy loss (%)

While radial CNLS (10) is ready for estimating the multiple input-multiple output cost frontier model (17), it might overfit training data and predictably perform poorly on testing data.<sup>7</sup> Considering that the benchmark regulation generally involves future economic incentives, we set a lower bound and upper bound on each subgradient to control their magnitudes to increase the performance of radial CNLS (Kuosmanen et al., 2022). Other restrictions on fitted subgradients for reducing overfitting include the Lipchitz norm (e.g., Mazumder et al., 2019) and  $L_2$  norm (e.g., Dai et al., 2023). One could even respecify the loss function to mitigate the effect of overfitting (e.g., Liao et al., 2023).

After taking into account these specifications and choosing  $x_1$  as numeraire, we rephrase radial CNLS (10) to estimate cost frontier model (17) and have the following nonlinear programming problem

$$\begin{aligned}
\min_{\chi, \beta, \gamma, \lambda, \delta, \mu, \varepsilon} \quad & \sum_{t=1}^T \sum_{i=1}^N \varepsilon_{it}^2 & (18) \\
\text{s.t.} \quad & \ln x_{1it} = -\ln(\chi_{it}) + \delta(\ln x_{1it} - \ln x_{2it}) + \boldsymbol{\lambda}' \mathbf{z}_{it} + \varepsilon_{it} & \forall i, t \\
& \chi_{it} = \boldsymbol{\beta}'_{i,t}(\mathbf{x}_{it}/(x_{1it}^{1/2} x_{2it}^{1/2})) + \mu_{it} b_{1it} - \boldsymbol{\gamma}'_{it} \mathbf{y}_{it} & \forall i, t \\
& \chi_{it} \leq \boldsymbol{\beta}'_{hs}(\mathbf{x}_{it}/(x_{1it}^{1/2} x_{2it}^{1/2})) + \mu_{hs} b_{1it} - \boldsymbol{\gamma}'_{hs} \mathbf{y}_{it} & \forall i, h; \forall t, s \\
& \boldsymbol{\beta}_{lo} \leq \boldsymbol{\beta}_{it} \leq \boldsymbol{\beta}_{up}, \quad \boldsymbol{\gamma}_{lo} \leq \boldsymbol{\gamma}_{it} \leq \boldsymbol{\gamma}_{up}, \quad \mu_{lo} \leq \mu_{it} \leq \mu_{up} & \forall i, t \\
& \boldsymbol{\beta}_{it} \geq \mathbf{0}, \quad \boldsymbol{\gamma}_{it} \geq \mathbf{0}, \quad \chi_{it} \geq 0 & \forall i, t
\end{aligned}$$

where the fourth set of constraints refers to the weight restrictions on subgradients, and  $\mu_{it}$  denotes the marginal coefficient of the outages. The lower and upper bounds (e.g.,  $\boldsymbol{\beta}_{lo}$  and  $\boldsymbol{\beta}_{up}$ ) in problem (18) can be determined by either a data-driven approach (e.g., cross-validation) or decision-makers and/or stakeholders. A major practical advantage of the latter is that the weight restrictions can be easily communicated to decision-makers (e.g., the regulator) and stakeholders (e.g., the regulated firms and their customers) as they can literally see the weight restrictions themselves and comment if those are too loose, too restrictive, or just fine.

---

<sup>7</sup>Overfitting is a longstanding problem in convex regression and other general nonparametric regression, where the subgradients may become very large at the boundary of the convex hull of the design points (Liao et al., 2023).

In practice, we first solve problem (18) without weight restrictions (i.e., the fourth sets of constraints) to obtain the subgradients estimates (i.e.,  $\hat{\beta}$ ,  $\hat{\gamma}$ , and  $\hat{\mu}$ ) and then take 10% and 90% quartiles of each subgradient as its lower and upper bounds to reestimate problem (18).

## 6.2 Empirical results

In this section we discuss what degree of difference in the estimates with and without weight restrictions and compare radial CNLS with its alternative. We thus apply the proposed radial CNLS approach to a panel of 77 Finnish electricity distribution firms in the years 2008–2020.<sup>8</sup> The descriptive statistics for inputs, outputs, and contextual variables are summarized in Table B2 (Appendix B). See Kuosmanen and Johnson (2020) for a detailed introduction to these selected variables.

Table 4 summarizes the results estimated by the basic radial CNLS model and weight-restricted radial CNLS model. Note that the estimated input distance function is a piece-wise linear function consisting of hyperplanes characterized by subgradients  $\beta_i$ ,  $\gamma_i$ , and  $\mu_i$ . When weight constraints are introduced to the radial CNLS model, the average estimates for the nonparametric part dramatically decrease, and the corresponding standard deviations also decline. This implies that a small part of firms has extremely estimated shadow prices (i.e.,  $\hat{\beta}_i$ ,  $\hat{\gamma}_i$ , or  $\hat{\mu}_i$ ), which can severely affect the fitted input distance function in a testing set and hence deteriorate the performance of economic incentives in energy regulation.

By comparing the 10%, 50%, and 90% quartile estimates, we observe that the changes in estimated shadow prices remain rather marginal, suggesting that most of the firms would not be highly affected by additional weight constraints. Furthermore, the estimated residual comparison also shows little impact of weight restrictions on the firm’s relative performance in that most firms cluster around the 45-degree line (see Fig. 3). That is, while the applied weight-restricted CNLS model will affect the estimated input distance function, it can avoid extreme shadow prices and show its robustness in inefficiency estimation and endogeneity bias.

---

<sup>8</sup>The original dataset is applied to carry out the incentive regulation for Finnish electricity distribution networks (Kuosmanen et al., 2022), and its earlier version has been widely used in, e.g., Kuosmanen (2012), Kuosmanen et al. (2013), and Kuosmanen and Johnson (2020).

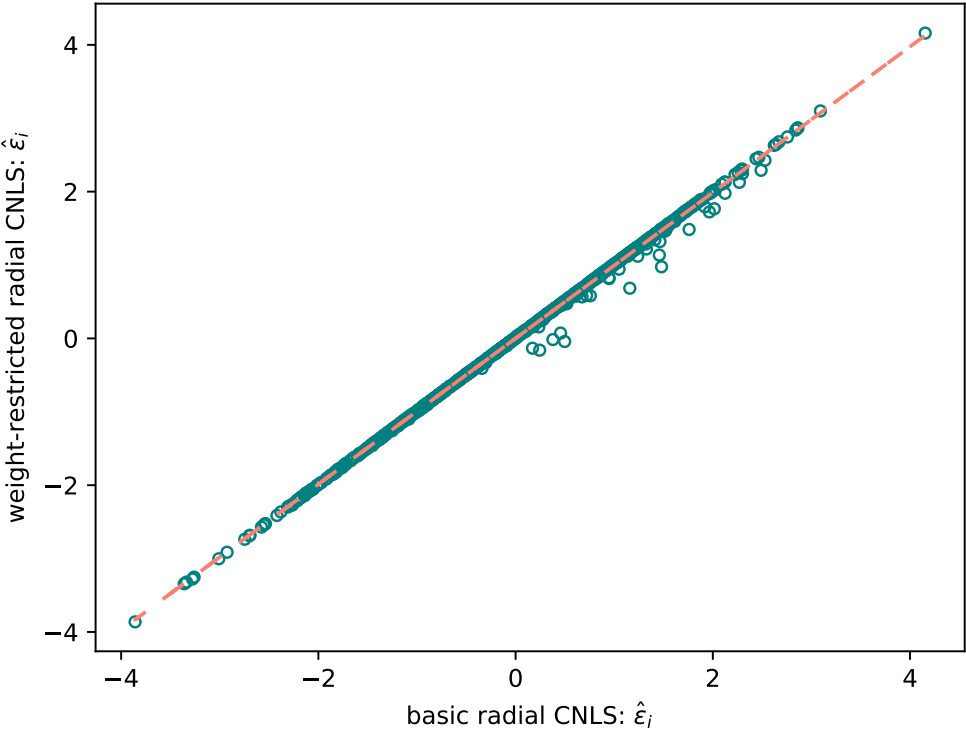
**Table 4.** Radial CNLS estimates for the Finnish electricity distribution firms.

	Mean	Std. Dev.	Min.	p10	p50	p90	Max.
Basic radial CNLS model: Problem (18) without weight constraints							
<i>Nonparametric part</i>							
$\hat{\beta}_1$	95.293	749.698	0.000	0.599	0.601	0.605	11511.607
$\hat{\beta}_2$	329.330	2413.736	0.000	0.000	0.000	0.088	22924.910
$\hat{\gamma}_1$ (€ cents/kWh)	28046.330	294777.200	0.000	0.001	0.186	382.347	5635193.000
$\hat{\gamma}_2$ (€/km)	3937.605	41107.930	0.000	0.001	22.846	49.378	752614.085
$\hat{\gamma}_3$ (€/customer)	553.660	6175.607	0.000	0.000	0.672	3.978	130694.500
$\hat{\mu}$	49027.060	675278.700	-139973.600	2.840	53.010	224.093	14844690.000
<i>Parametric part</i>							
$\hat{\delta}$				4.10			
$\hat{\lambda}_1$				-0.71			
$\hat{\lambda}_2$				4.11			
Weight-restricted radial CNLS model: Problem (18)							
<i>Nonparametric part</i>							
$\hat{\beta}_1$	0.601	0.002	0.599	0.599	0.601	0.603	0.605
$\hat{\beta}_2$	0.011	0.025	0.000	0.000	0.000	0.050	0.088
$\hat{\gamma}_1$ (€ cents/kWh)	80.904	134.517	0.001	0.007	2.057	364.635	382.347
$\hat{\gamma}_2$ (€/km)	21.685	16.134	0.001	0.010	22.817	46.901	49.378
$\hat{\gamma}_3$ (€/customer)	1.540	1.654	0.000	0.000	0.543	3.816	3.978
$\hat{\mu}$	64.726	61.222	2.840	2.881	64.738	141.897	224.093
<i>Parametric part</i>							
$\hat{\delta}$				4.10			
$\hat{\lambda}_1$				-0.71			
$\hat{\lambda}_2$				3.98			

It is worth noting that the minimal estimated  $\hat{\mu}_i$  is negative in the basic radial CNLS model but is positive in the weight-restricted radial CNLS model. Unsurprisingly, the marginal effect of outages at the firm level can be positive or negative in the context of the benchmark regulation (Kuusmanen and Johnson, 2020). Recall that there is huge heterogeneity between Finnish electricity distribution firms in terms of variable and fixed costs as reflected by their standard deviation (see Table B2 in Appendix B). Firms with low outages can use a high level of quality and reliability of distribution as a competitive advantage in the yardstick competition. For these firms, the shadow price of outages is negative. Firms that face exceptionally severe weather shocks have higher operational costs due to large outages, deducing the positive shadow price of outages. This is because the standard compensations paid to customers for interruptions are also included in the variable cost. In practice, the threshold value for a positive outage effect is very high because the outage

values must be exceptionally high compared to other firms such that the firm’s operations appear competitive in the benchmark regulation between firms. Therefore, the impact of outages on the distribution firms is a “U-shaped” curve.

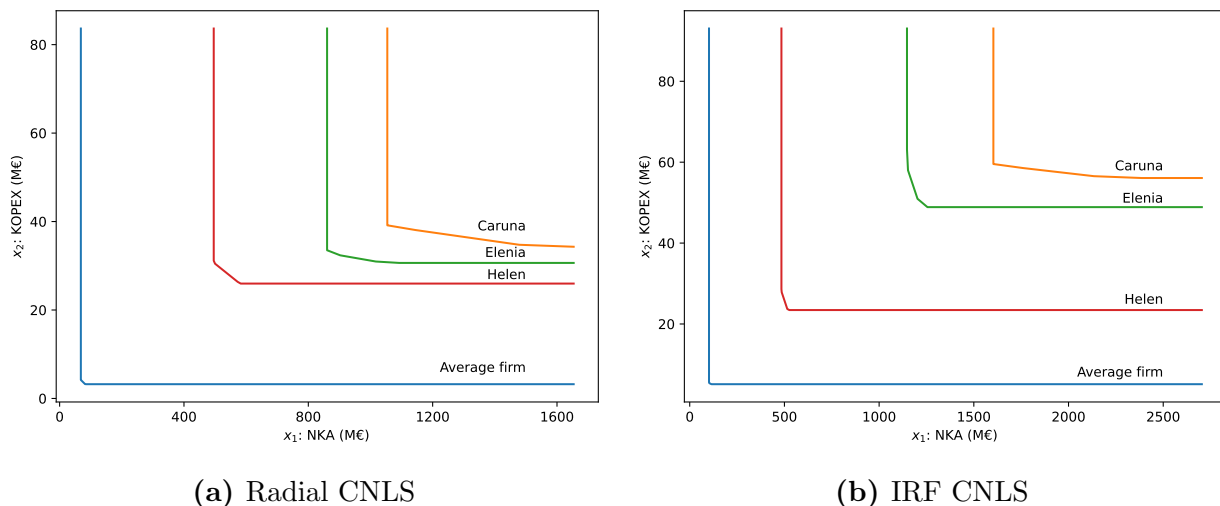
Furthermore, additional weight constraints also influence the estimated coefficient of the second contextual variable, energy loss. In the weight-restricted radial CNLS model, the value of the estimated coefficient of energy loss is 3.98, slightly lower than the value of 4.11 obtained from the basic radial CNLS model. Since the share of energy loss is on average higher in sparsely populated areas than in densely populated areas, the distorting effect of possible efficiency differences can be partially canceled out from the energy loss estimate with the help of weight restrictions.



**Fig. 3.** Estimated residuals: radial CNLS and weighted-restricted radial CNLS.

Fig. 4 demonstrates the estimated input isoquants of the three largest distribution firms (Caruna, Elenia, and Helen) and the average level of all firms based on radial CNLS and IRF CNLS estimates. The horizontal axis is fixed cost (NKA), and the vertical axis is variable cost (KOPEX): both are rescaled in Million €. Note that the shape of the input isoquant

is determined by the output structure of the firm in the multiple output setting. Given the output structure of the three largest distribution firms, both two subfigures illustrate the considerable substitution possibilities between NKA and KOPEX. But the right figure shows that the average firm yields almost a Leontief-type input isoquant given the averaged output structure.



**Fig. 4.** Estimated input isoquants of the average firm and the three largest firms.

Furthermore, it can be seen from Fig. 4 that the estimated input isoquants become somewhat different when we use two inputs in the radial CNLS model rather than projecting in the direction of one input in the IRF CNLS model. However, from the figure alone, we cannot determine which approach is better. To empirically test which specification yields more accurate predictions, one possibility is to partition the sample into a training set and a test set to compare the out-of-sample MSE (Kuusmanen et al., 2022).

## 7 Conclusions

Modeling of joint production is a more vexing problem than most authors realize. The axiomatic nonparametric approach is the only theoretically sound alternative when data are perturbed by random noise, and the parametric approach is likely to violate the theoretical properties of the distance function. Modeling radial input and output distance functions in CNLS/StoNED has also been an enigma for almost two decades. Finally, in this paper the

problem has been solved in the case of the input distance function, but the output distance function partially remains.

This paper proposes a new radial CNLS approach to estimate the input distance function in the multiple input-multiple output specification. We demonstrate how to transform the input distance function correctly and show that the orthogonality conditions can be satisfied in the developed approach. We then discuss the possible extensions from the other three aspects: considering the general case with more than two inputs, connecting with the efficiency analysis, and estimating the output distance function.

Our simulations compare the finite sample performance of radial CNLS and other deterministic and stochastic frontier approaches regarding the input distance function estimation. Radial CNLS outperforms the naive CNLS, DEA, and SFA approaches in all scenarios considered, suggesting that radial CNLS is a more efficient method for estimating the input distance function.

The developed approach is further applied to the Finnish electricity distribution network regulation. To reduce the potential overfitting and increase the out-of-sample performance in energy regulation, we introduce weight restrictions on shadow prices to the framework of radial CNLS. In contrast to other traditional model specifications (e.g., DDF and IRF), radial CNLS is more robust to input selection. The illustrated input isoquants suggest more substitution possibilities.

Future research could further examine joint production from the perspectives of nonparametric identification (see, e.g., Matzkin, 2007; Chiappori et al., 2015; Heckman et al., 2010) and partial identification (see, e.g., Manski, 2009; Tamer, 2010). Axioms of production theory (linear homogeneity) provide moment conditions (i.e., generalized method of moments), but the model remains inherently underidentified. Underidentification is conventionally seen as a failure in econometrics (Arellano et al., 2012), but in many relevant applications, partial identification may be sufficient for the task at hand. One possible avenue of future research is to impose moment conditions supported by theory (linear homogeneity) and apply additional statistical criteria (e.g., maximum likelihood and least squares) to achieve (partial) identification.

## Acknowledgments

The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources. Sheng Dai gratefully acknowledges financial support from the OP Group Research Foundation [grant no. 20230008] and the Turku University Foundation [grant no. 081520].

## References

- Ahn, H., Clermont, M. and Langner, J. (2023). Comparative performance analysis of frontier-based efficiency measurement methods—A Monte Carlo simulation, *European Journal of Operational Research* **307**: 294–312.
- Almanidis, P. and Sickles, R. C. (2012). The skewness issue in stochastic frontiers models: Fact or fiction?, *Exploring research frontiers in contemporary statistics and econometrics: A festschrift for Léopold Simar*, Springer Verlag, Berlin, pp. 201–227.
- Arellano, M., Hansen, L. P. and Sentana, E. (2012). Underidentification?, *Journal of Econometrics* **170**: 256–280.
- Bernard, A. B., Redding, S. J. and Schott, P. K. (2010). Multiple-product firms and product switching, *American Economic Review* **100**: 70–97.
- Chambers, R. G., Chung, Y. and Färe, R. (1996). Benefit and distance functions, *Journal of Economic Theory* **70**: 407–419.
- Chambers, R. G., Chung, Y. and Färe, R. (1998). Profit, directional distance functions, and Nerlovian efficiency, *Journal of Optimization Theory and Applications* **98**: 351–364.
- Charnes, A., Cooper, W. and Rhodes, E. (1978). Measuring the efficiency of decision making units, *European Journal of Operational Research* **2**: 429–444.
- Chiappori, P.-A., Komunjer, I. and Kristensen, D. (2015). Nonparametric identification and estimation of transformation models, *Journal of Econometrics* **188**: 22–39.
- Coelli, T. and Perelman, S. (1999). A comparison of parametric and non-parametric distance functions: With application to European railways, *European Journal of Operational Research* **117**: 326–339.
- Dai, S., Fang, Y. H., Lee, C. Y. and Kuosmanen, T. (2021). pyStoNED: A Python package for convex regression and frontier estimation, *arXiv preprint arXiv:2109.12962*.
- Dai, S., Kuosmanen, T. and Zhou, X. (2023). Non-crossing convex quantile regression, *Economics Letters* **233**: 111396.
- Dai, X. (2016). Non-parametric efficiency estimation using Richardson-Lucy blind deconvolution, *European Journal of Operational Research* **248**: 731–739.
- De Loecker, J., Goldberg, P. K., Khandelwal, A. K. and Pavcnik, N. (2016). Prices, markups, and trade reform, *Econometrica* **84**: 445–510.

- Färe, R., Martins-Filho, C. and Vardanyan, M. (2010). On functional form representation of multi-output production technologies, *Journal of Productivity Analysis* **33**: 81–96.
- Färe, R. and Primont, D. (1995). *Multi-output production and duality: Theory and applications*, Kluwer Academic Publishers, Boston.
- Farrell, M. J. (1957). The measurement of productive efficiency, *Journal of the Royal Statistical Society. Series A (General)* **120**: 253–290.
- Foster, L., Haltiwanger, J. and Syverson, C. (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?, *American Economic Review* **98**: 394–425.
- Hall, P. and Simar, L. (2002). Estimating a change point, boundary, or frontier in the presence of observation error, *Journal of the American Statistical Association* **97**: 523–534.
- Heckman, J. J., Matzkin, R. L. and Nesheim, L. (2010). Nonparametric identification and estimation of nonadditive hedonic models, *Econometrica* **78**: 1569–1591.
- Henningsen, G., Henningsen, A. and Jensen, U. (2015). A Monte Carlo study on multiple output stochastic frontiers: a comparison of two approaches, *Journal of Productivity Analysis* **44**: 309–320.
- Johnson, A. L. and Kuosmanen, T. (2011). One-stage estimation of the effects of operational conditions and practices on productive performance: Asymptotically normal and efficient, root- $n$  consistent StoNEZD method, *Journal of Productivity Analysis* **36**: 219–230.
- Johnson, A. L. and Kuosmanen, T. (2012). One-stage and two-stage DEA estimation of the effects of contextual variables, *European Journal of Operational Research* **220**: 559–570.
- Jondrow, J., Lovell, C. A. K., Materov, I. S. and Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model, *Journal of Econometrics* **19**: 233–238.
- Kumbhakar, S. C. (2013). Specification and estimation of multiple output technologies: A primal approach, *European Journal of Operational Research* **231**: 465–473.
- Kuosmanen, N. and Valmari, N. (2023). Renewal of companies through product switching, *ETLA Working Papers, No. 104*, The Research Institute of the Finnish Economy (ETLA), Helsinki.
- Kuosmanen, T. (2012). Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model, *Energy Economics* **34**: 2189–2199.
- Kuosmanen, T. and Johnson, A. L. (2017). Modeling joint production of multiple outputs in StoNED: Directional distance function approach, *European Journal of Operational Research* **262**: 792–801.
- Kuosmanen, T. and Johnson, A. L. (2020). Conditional yardstick competition in energy regulation, *The Energy Journal* **41**: 67–92.

- Kuosmanen, T. and Kortelainen, M. (2012). Stochastic non-smooth envelopment of data: Semi-parametric frontier estimation subject to shape constraints, *Journal of Productivity Analysis* **38**: 11–28.
- Kuosmanen, T., Kuosmanen, N. and Dai, S. (2022). Kohtuullinen muuttuva kustannus sähkön jakeluverkkoyhtiöiden valvontamallissa: Ehdotus tehostamiskannustimen kehittämiseksi 6. ja 7. valvontajaksoilla vuosina 2024–2031 (in Finnish), *Technical report*. Available from: [energiavirasto.fi](http://energiavirasto.fi), accessed 18 March 2023.
- Kuosmanen, T., Saastamoinen, A. and Sipiläinen, T. (2013). What is the best practice for benchmark regulation of electricity distribution? Comparison of DEA, SFA and StoNED methods, *Energy Policy* **61**: 740–750.
- Liao, Z., Dai, S. and Kuosmanen, T. (2023). Convex support vector regression, *European Journal of Operational Research* **313**: 858–870.
- Lovell, C. K., Travers, P., Richardson, S. and Wood, L. (1994). Resources and functionings: A new view of inequality in Australia, in W. Eichhorn (ed.), *Models and Measurement of Welfare and Inequality*, Springer, Berlin, pp. 787–807.
- Manski, C. F. (2009). *Identification for prediction and decision*, Harvard University Press, Cambridge, MA.
- Matzkin, R. L. (2007). Nonparametric identification, *Handbook of econometrics* **6**: 5307–5368.
- Mazumder, R., Choudhury, A., Iyengar, G. and Sen, B. (2019). A computational framework for multivariate convex regression and its variants, *Journal of the American Statistical Association* **114**: 318–331.
- Panzar, J. C. and Willig, R. D. (1981). Economies of scope, *The American Economic Review* **71**: 268–272.
- Parmeter, C. F. and Kumbhakar, S. C. (2014). Efficiency analysis: A primer on recent advances, *Foundations and Trends in Econometrics* **7**: 191–385.
- Schaefer, J. and Clermont, M. (2018). Stochastic non-smooth envelopment of data for multi-dimensional output, *Journal of Productivity Analysis* **50**: 139–154.
- Shephard, R. W. (1970). *Theory of cost and production functions*, Princeton University Press, Princeton.
- Simar, L. and Wilson, P. W. (2010). Inferences from cross-sectional, stochastic frontier models, *Econometric Reviews* **29**: 62–98.
- Tamer, E. (2010). Partial identification in econometrics, *Annual Review of Economics* **2**: 167–195.

# Online Supplement

## Appendix A: Proofs of Theorems

**Theorem 1.** *If the observed data are generated according to the data generating process (DGP) described, then the value of the input distance function is equal to*

$$DI(\mathbf{x}_i, \mathbf{y}_i) = \exp(\varepsilon_i)$$

*Proof.* If the observed data are generated according to the DGP stated in Section 2.1, then we have

$$DI(\mathbf{x}, \mathbf{y}) = DI(\exp(\varepsilon_i)\mathbf{x}^*, \mathbf{y})$$

By utilizing the linear homogeneity of input distance function  $DI$ , the above input distance function  $DI(\exp(\varepsilon_i)\mathbf{x}^*, \mathbf{y})$  can be reorganized as

$$DI(\mathbf{x}, \mathbf{y}) = DI(\exp(\varepsilon_i)\mathbf{x}^*, \mathbf{y}) = \exp(\varepsilon_i)DI(\mathbf{x}^*, \mathbf{y})$$

By definition, the efficient input-output vectors must be on the boundary of the production possibility set  $\mathcal{T}$  and satisfy the condition  $DI(\mathbf{x}^*, \mathbf{y}) \equiv 1$ . We thus have

$$DI(\mathbf{x}, \mathbf{y}) = \exp(\varepsilon_i)$$

■

**Theorem 2.** *The CNLS residuals obtained as the optimal solution to problem (7) satisfy the following conditions*

$$\begin{aligned} \sum_{i=1}^n \frac{\hat{\varepsilon}_i}{\hat{\chi}_i} &= 0 \\ \sum_{i=1}^n \left( \frac{x_{ik} \hat{\varepsilon}_i}{x_{1i} \hat{\chi}_i} - \eta_{ik} \right) &= 0 \end{aligned}$$

*Proof.* The Lagrangian function of problem (7) can be formulated

$$\begin{aligned} L(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\rho}, v) &= \frac{1}{2} \sum_{i=1}^n \varepsilon_i^2 + \sum_{i=1}^n \mu_i (\ln x_{1i} + \ln(\chi_i) - \varepsilon_i) \\ &+ \sum_{i=1}^n \sum_{h=1}^n \lambda_{ih} (\alpha_i + \boldsymbol{\beta}'_i(\mathbf{x}_i/x_{1i}) - \boldsymbol{\gamma}'_i \mathbf{y}_i - \alpha_h - \boldsymbol{\beta}'_h(\mathbf{x}_i/x_{1i}) + \boldsymbol{\gamma}'_h \mathbf{y}_i) \\ &- \sum_{i=1}^n \boldsymbol{\eta}'_i \boldsymbol{\beta}_i - \sum_{i=1}^n \boldsymbol{\rho}'_i \boldsymbol{\gamma}_i - \sum_{i=1}^n v_i \chi_i \end{aligned}$$

where  $\mu_i$ ,  $\lambda_{ih}$ ,  $\eta_i$ ,  $\rho_i$ , and  $v_i$  are the Lagrangian multipliers of the corresponding constraints. The optimal solution of problem (7) satisfies the first order conditions with respect to  $\varepsilon_i$ ,  $\alpha_i$ , and  $\beta_{ik}$

$$\begin{aligned}\frac{\partial L}{\partial \varepsilon_i} &= \hat{\varepsilon}_i - \mu_i = 0 \quad \forall i \\ \frac{\partial L}{\partial \alpha_i} &= \mu_i \frac{1}{\hat{\chi}_i} + \sum_{h=1}^n \lambda_{ih} - \sum_{h=1}^n \lambda_{hi} = 0 \quad \forall i \\ \frac{\partial L}{\partial \beta_{ik}} &= \mu_i \frac{x_{ik}}{x_{1i}} \frac{1}{\hat{\chi}_i} + \sum_{h=1}^n \lambda_{ih} \frac{x_{ik}}{x_{1i}} - \sum_{h=1}^n \lambda_{hi} \frac{x_{hk}}{x_{1i}} - \eta_{ik} = 0 \quad \forall i, k\end{aligned}$$

Using the first equation to substitute  $\mu_i$  by  $\hat{\varepsilon}_i$  in the second equation, we have

$$\hat{\varepsilon}_i = \hat{\chi}_i \left[ \sum_{h=1}^n \lambda_{hi} - \sum_{h=1}^n \lambda_{ih} \right].$$

The expression in brackets is equal to zero by construction. This implies that the sum of residuals estimated by problem (7) is not necessarily zero, but the sum of the ratio of residuals to the fitted values is zero:

$$\sum_{i=1}^n \frac{\hat{\varepsilon}_i}{\hat{\chi}_i} = 0.$$

Applying the same operations to the third first-order condition, we have

$$\sum_{i=1}^n \left( \frac{x_{ik}}{x_{1i}} \frac{\hat{\varepsilon}_i}{\hat{\chi}_i} - \eta_{ik} \right) = 0.$$

■

**Theorem 3.** *The radial CNLS residuals obtained as the optimal solution to problem (10) satisfy the following sample orthogonality condition*

$$\sum_{i=1}^n (\ln x_{1i} - \ln x_{2i}) \cdot \hat{\varepsilon}_i = 0$$

*Proof.* The Lagrangian function of problem (10) can be formulated

$$\begin{aligned}L(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\rho}, \boldsymbol{v}) &= \frac{1}{2} \sum_{i=1}^n \varepsilon_i^2 + \sum_{i=1}^n \mu_i (\ln x_{1i} + \ln(\chi_i) - \delta(\ln x_{1i} - \ln x_{2i}) - \varepsilon_i) \\ &+ \sum_{i=1}^n \sum_{h=1}^n \lambda_{ih} (\alpha_i + \boldsymbol{\beta}'_i(\mathbf{x}_i / (x_{1i}^{1-d} x_{2i}^d)) - \boldsymbol{\gamma}'_i \mathbf{y}_i - \alpha_h - \boldsymbol{\beta}'_h(\mathbf{x}_i / (x_{1i}^{1-d} x_{2i}^d)) + \boldsymbol{\gamma}'_h \mathbf{y}_i) \\ &- \sum_{i=1}^n \boldsymbol{\eta}'_i \boldsymbol{\beta}_i - \sum_{i=1}^n \boldsymbol{\rho}'_i \boldsymbol{\gamma}_i - \sum_{i=1}^n v_i \chi_i\end{aligned}$$

where  $\mu_i$ ,  $\lambda_{ih}$ ,  $\eta_i$ ,  $\rho_i$ , and  $v_i$  are the Lagrangian multipliers of the corresponding constraints. The optimal solution of problem (10) satisfies the first order conditions with respect to  $\varepsilon_i$  and  $\delta$

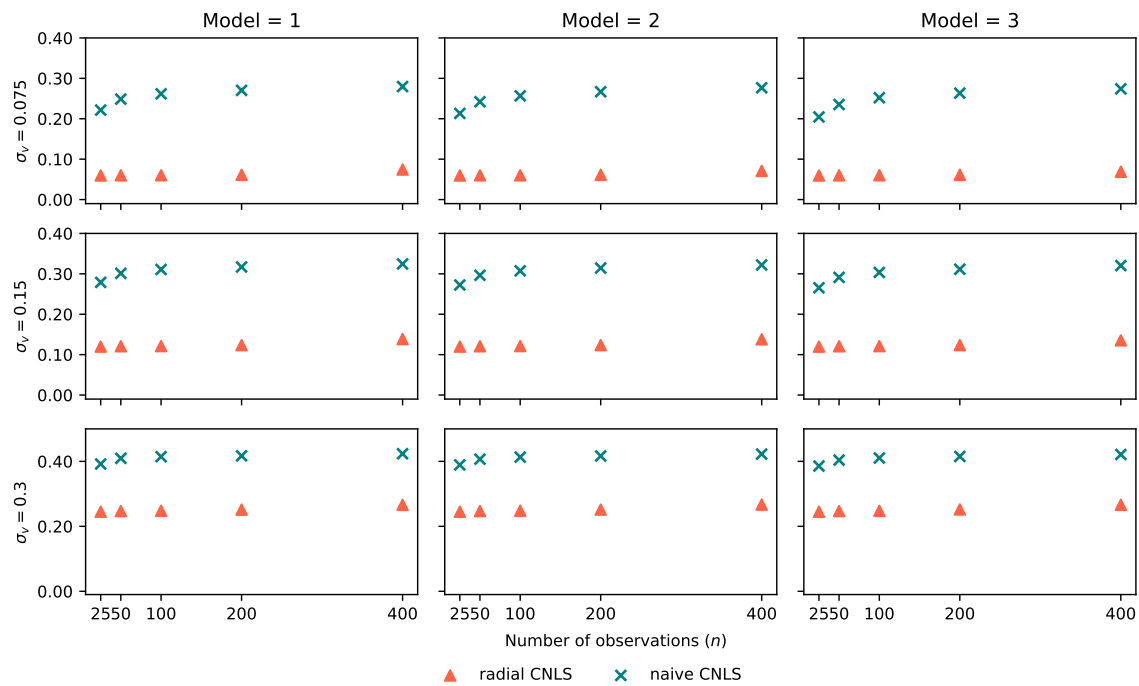
$$\begin{aligned}\frac{\partial L}{\partial \varepsilon_i} &= \hat{\varepsilon}_i - \mu_i = 0 \quad \forall i \\ \frac{\partial L}{\partial \delta} &= \sum_{i=1}^n \mu_i (\ln x_{1i} - \ln x_{2i}) = 0\end{aligned}$$

Combining these two first-order conditions, we thus have

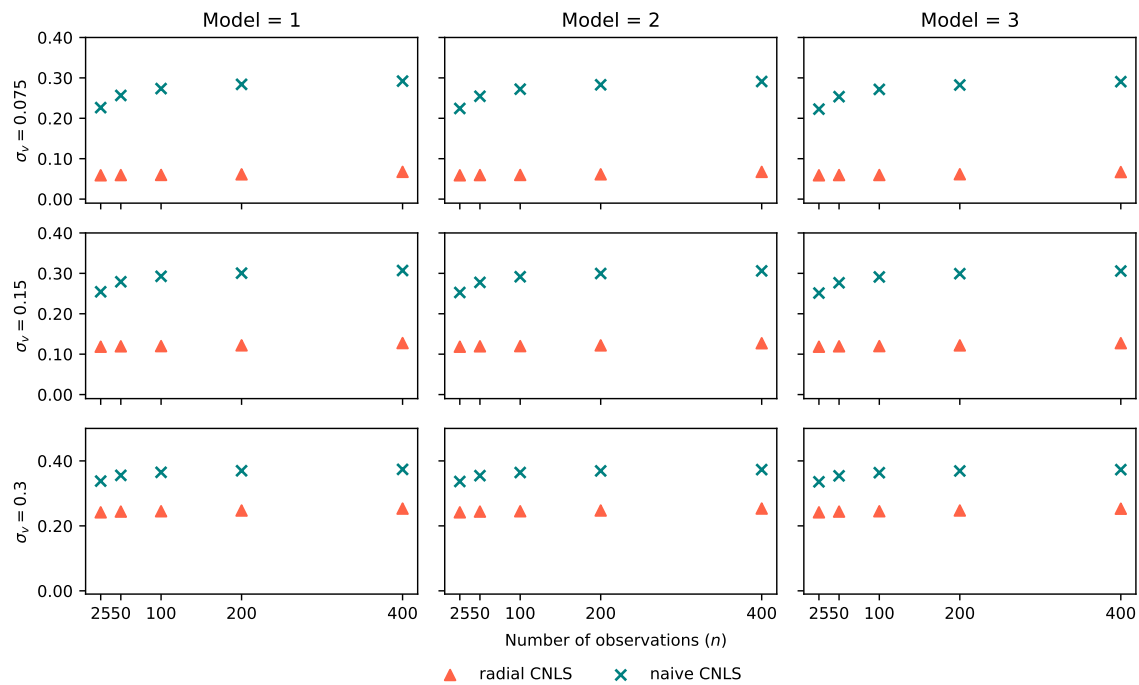
$$\sum_{i=1}^n (\ln x_{1i} - \ln x_{2i}) \cdot \hat{\varepsilon}_i = 0$$

■

## Appendix B: Additional tables and figures



**Fig. B1.** Mean absolute errors of radial CNLS and naive CNLS in DGP I with Type-A.



**Fig. B2.** Mean absolute errors of radial CNLS and naive CNLS in DGP II.

**Table B1.** Performance of radial CNLS, DEA, and SFA with  $n = 400$  in DGP I with Type-B.

Model	$\rho$	Mean Squared Error				Mean Absolute Error			
		CNLS	DEA	SFA-TL	SFA-CD	CNLS	DEA	SFA-TL	SFA-CD
1	0	0.013	0.076	0.508	12.350	0.082	0.234	0.189	0.859
	0.5	0.017	0.081	0.499	11.941	0.100	0.242	0.195	0.862
	1	0.030	0.097	0.500	12.779	0.138	0.266	0.210	0.862
	2	0.094	0.169	0.797	15.031	0.237	0.341	0.301	0.895
2	0	0.012	0.070	0.531	13.279	0.082	0.223	0.200	0.933
	0.5	0.017	0.076	0.668	13.204	0.101	0.233	0.213	0.924
	1	0.030	0.092	0.812	15.970	0.138	0.257	0.242	0.933
	2	0.094	0.164	1.742	16.856	0.238	0.335	0.388	0.969
3	0	0.012	0.068	1.328	15.575	0.082	0.218	0.263	1.028
	0.5	0.017	0.073	1.472	15.852	0.100	0.228	0.282	1.015
	1	0.030	0.089	1.701	15.810	0.138	0.253	0.329	1.012
	2	0.094	0.160	2.734	17.903	0.237	0.331	0.500	1.041

**Table B2.** The descriptive statistics for inputs, outputs, and contextual variables.

Variable	Mean	Std. Dev.	Min.	Max.	Obs.
Capital stock ( $x_1$ )	126253.5	284279.9	1540.6	2664448.0	1001
Controllable operational expenditure ( $x_2$ )	5864.9	11156.8	115.2	114993.4	1001
Energy supply ( $y_1$ )	569.3	1095.1	15.8	7528.9	1001
Network length ( $y_2$ )	5063.1	11832.3	133.6	80278.2	1001
Number of use points ( $y_3$ )	44721.0	83271.3	741.0	479365.0	1001
Outages ( $b_1$ )	2202.7	8520.3	2.3	173568.5	1001
Connection points / Use points ( $z_1$ )	0.63	0.24	0.09	1.00	1001
Energy loss ( $z_1$ )	0.04	0.03	0.00	0.24	1001