Modeling joint production of multiple outputs in StoNED:

Directional distance function approach

Timo Kuosmanen[1*] and Andrew L. Johnson[2, 3]

1) School of Business, Aalto University, 00100 Helsinki, Finland.
2) Department of Industrial and Systems Engineering, Texas A&M University, TX 77840, USA.
3) School of Information Science and Technology, Osaka University, Suita 565-0871, Japan.

Abstract:

Estimation of joint production technologies involving multiple outputs has proved a vexing challenge. Existing methods are unsatisfactory as they either assume away stochastic noise or restrict to functional forms that have incorrect output curvature. The first contribution of this paper is to develop a new probabilistic data generating process that is compatible with the directional distance function. The directional distance function is a very general functional characterization of production technology that has proved useful for modeling joint production of multiple outputs. The second contribution of this paper is to develop a new estimator of the directional distance function that builds upon axiomatic properties and does not require any functional form assumptions. The proposed estimator is a natural extension of stochastic nonparametric envelopment of data (StoNED) framework to multiple output setting. We examine the practical aspects and usefulness of the proposed approach in the context of incentive regulation of the Finnish electricity distribution firms.

Key words: *Data envelopment analysis (DEA), Economies of scope, Efficiency analysis, Frontier estimation, Stochastic nonparametric envelopment of data (StoNED)*

---

[*] Corresponding author: E-mail timo.kuosmanen@aalto.fi, Tel +358943131.

Abbreviations: CNLS = convex nonparametric least squares; DDF = directional distance function; DEA = data envelopment analysis; DGP = data generating process; SFA = stochastic frontier analysis; StoNED = stochastic nonparametric envelopment of data.

1. Introduction

Economies of scope is a fundamental notion in production economics. Economies of scope prevail if joint production of two or more outputs is more cost efficient than producing those outputs separately in specialized production processes (Panzar and Willig, 1981). Recently, Pope and Johnson (2013) introduced the notion of *returns to scope* that does not depend on input prices. According to their terminology, production technology exhibits positive returns to scope if the output sets are convex at all non-negative input levels. For business strategy, the curvature of the output isoquant is a critical determinant of whether it is beneficial to produce multiple products jointly (positive returns to scope) or specialize in a single product (negative returns to scope). Important examples of positive returns to scope include cogeneration (e.g., combined heat and power) and byproducts, which may be desirable outputs (e.g., manure used as fertilizer) or undesirable outputs (e.g., waste and pollution).

In production theory, technology distance functions have been established as valid functional representations of joint production technologies (see, e.g., Färe and Primont, 1995; Chambers et al., 1998). Ability to model joint production consistent with the axioms of the production theory is one of the main advantages of the axiomatic mathematical programming approach that builds upon such classic works as Koopmans (1951), Farrell (1957), Shephard (1970), and Afriat (1972). In operational research, Charnes et al. (1978) further developed this approach and popularized it under the label of data envelopment analysis (DEA). However, the standard DEA models assume all observed data points to be feasible (the envelopment axiom), and hence any random noise due to omitted variables, unobserved heterogeneity, and data errors must be assumed away. This is a shortcoming of the conventional DEA approach in real world applications where data is always more or less noisy.

Stochastic frontier analysis (SFA: Aigner et al. 1977; Meeusen and van den Broeck, 1977) is an econometric approach to model production that takes noise explicitly into account as a random variable. Since Lovell et al. (1994) there has been considerable efforts to extend SFA to multiple output setting using distance functions (see, e.g., recent paper by Kumbhakar, 2013). However, the restrictive functional form assumptions imposed in SFA seem particularly problematic in the context of joint production. By far the most common parametric functional forms used in the estimation of distance functions is translog. Coelli and Perelman (1999) justify the choice of translog by correctly noting that the Cobb-Douglas functional form has “incorrect output curvature.” More specifically, the output sets of the Cobb-Douglas transformation function are neither closed nor convex at any parameter values. Hence, the Cobb-Douglas functional form is clearly invalid representation of a multiple output technology. However, there is no guarantee that the translog functional form would automatically solve the problem, but at least, the translog technology does have closed and convex output sets under certain parameter values (see Perelman and Santini, 2009). It is easy to verify that when the translog functional forms satisfies closedness and convexity, then it violates free disposability of outputs (monotonicity). By the free disposability axiom one should be able to dispose any output variable to zero, but the translog function does not allow this because the logarithm of zero is undefined. In practice, violation of free disposability is problematic because the frontier contains points that are dominated, and hence technically inefficient according to the definition by Koopmans (1951). A related practical problem concerns the modeling of specialized firms that do not produce all of the outputs (consider, e.g., railroad companies that specialize in passenger or freight services). Unfortunately, these problems have not been duly recognized in the SFA literature. For example, Coelli and Perelman (1999) state that they "assume that the technology satisfies the axioms listed in Färe and Primont (1995)," failing to note that the translog function cannot satisfy this set of standard axioms at any parameter values. In our view, the wrong curvature of the translog is a more fundamental problem than the proponents of the parametric approach recognize.

There exist semiparametric SFA approaches that apply local averaging techniques such as kernel regression (Fan et al., 1996; Kneip and Simar, 1996) or local maximum likelihood (Kumbhakar et al., 2007) to avoid the restrictive functional form assumptions of the classical SFA. To extend this approach to distance functions, one would have to be able to impose global properties, at least, the properties required for identification, such as linear homogeneity of the output distance function or the translation property of the directional distance function. As discussed above, it would be also desirable to impose global monotonicity and convexity of output sets. In this respect, our main objections against the translog function discussed above carry over to the semiparametric SFA as well. While it is possible to impose shape restrictions locally (see, e.g., Du et al., 2013), imposing global constraints in local averaging appears to be difficult. Most likely this will require use of similar techniques to those routinely used in the axiomatic mathematical programming approach (see, e.g., Simar and Zelenyuk, 2011; Yagi et al., 2016).

For the reasons discussed above, we consider the axiomatic mathematical programming approach to modeling joint production not only the most promising way forward, but in fact, the *only way* to ensure that the estimated frontier satisfies the key properties such as closedness, free disposability, convexity, and the various homogeneity properties implied by the production theory. In this stream of literature, extending DEA to probabilistic setting has been an active research area since the late 1980s (see, e.g., Olesen and Petersen, 2016, for a recent survey). Important early contributions to this literature include Banker and Maindiratta (1992) and Banker (1993). More recently, Kuosmanen (2006) and Kuosmanen and Kortelainen (2012) introduced *stochastic nonparametric envelopment of data* (StoNED) as a more general framework that combines the classic DEA and SFA models as its special cases (e.g., Kuosmanen and Johnson, 2010; Kuosmanen et al., 2015). StoNED might be equally well called "stochastic DEA" or "nonparametric SFA": the purpose of the new term StoNED is to highlight the generality of the approach and the fact that its intellectual roots are not limited to DEA or SFA but also span such areas as nonparametric tests of optimizing behavior in microeconomic theory (e.g., by Diewert and Parkan, 1983, and Varian 1984, 1985) and convex regression in statistics and operational research (Hildreth, 1954; Dent, 1973; Hanson and Pledger, 1976; Holloway, 1979).

To estimate joint production technologies in the StoNED framework, Kuosmanen (2006) applies the directional distance function (DDF) by Chambers et al. (1996; 1998) that contains the classic radial distance functions as its special cases. Kuosmanen et al. (2015) further elaborate the StoNED estimator of the directional distance function. While extending the StoNED estimator from the single output setting to multiple outputs using distance functions is relatively straightforward, it is not clear what kind of assumptions are required to ensure even the most basic statistical properties of the estimator such as consistency. This question has proved particularly challenging because almost all known probabilistic models of a data generating process (DGP) in the literature assume a single output variable, and only few formal DGPs for multioutput setting have been suggested (e.g., Varian, 1985; Banker and Maindiratta, 1992; Simar 2007; Kuosmanen et al., 2007), none of which are compatible with the DDF.

The main objectives of this paper are two-fold. The first objective is to introduce a novel probabilistic DGP where all inputs and outputs are perturbed in some pre-defined direction. We formally show that the proposed directional DGP is compatible with the directional distance function. The second objective is to develop a consistent nonparametric estimator of the DDF. The proposed estimator is a natural extension of the DDF estimator outlined in Kuosmanen (2006) and Kuosmanen et al. (2015), but in this paper we provide a more detailed, systematic presentation and relax the distributional assumptions associated with the inefficiency and noise terms.

The practical motivation for this study stems from the regulation of electricity distribution firms in Finland (see Kuosmanen, 2012; Kuosmanen et al., 2013). These firms are local monopolies that set their output prices subject to complex dynamic revenue constraints implied by the regulation model. The production process involves multiple inputs (e.g., labor, capital), desirable outputs (e.g., transmitted energy, capacity) and undesirable outputs (e.g., outages, energy loss). The Finnish energy regulator applies the StoNED method as an integral part of incentive regulation to set the acceptable cost level for the regulated firms since 2012 (Kuosmanen, 2012). In this context, there is need for an estimation method that can handle multiple inputs and outputs within the axiomatic framework of production theory and model inefficiency and noise in a probabilistic manner. The purpose of this paper is to develop a method that satisfies both these criteria. Based on the findings of this study, the

Finnish regulator applies the multiple output StoNED approach developed in this paper in the incentive regulation of Finnish electricity distribution firms in years 2016 – 2023.

The electricity distribution regulation application highlights that if the directional distance function is used to estimate the production function and for analysis purposes, the direction consistent with the data generation process may not be the same as the direction used for analysis purposes. Thus, we introduce a two-stage method in which we estimate the production function in the first stage using criteria consistent with the data generation process and in the second stage identify performance benchmarks located on the production function using a direction consistent with the analysis purpose.

The rest of the paper is organized as follows. Section 2 introduces the directional distance function and a probabilistic DGP where all inputs and outputs are perturbed by inefficiency and noise. In Section 3 we show that the proposed DGP is compatible with the DDF representation of technology, and apply the result to develop a regression equation. Orthogonality conditions for identification of the DDF are also formally established. Section 4 develops an axiomatic nonparametric StoNED method for estimating the DDF. Section 5 discusses practical aspects related to the proposed approach in the context of incentive regulation of the Finnish electricity distribution firms. Section 6 draws concluding remarks and discusses some promising avenues for future research. The proofs of all formal propositions are presented in the Appendix, which is available as an online supplement to this paper.

2. Theoretical model

2.1 Benchmark technology

Consider a joint production technology that transforms input vector $\mathbf{x} = (x_1 ... x_m)' \in \Re_+^m$ to an output vector $\mathbf{y} = (y_1 ... y_s)' \in \Re_+^s$. The most obvious representation of the benchmark technology is the production possibility set defined as

$$T = \{(\mathbf{x}, \mathbf{y}) \in \Re_+^{m+s} \mid \mathbf{x} \,\text{can produce}\, \mathbf{y}\}.$$

Set T contains all input-output combinations that are feasible for benchmarking. By the use of the term benchmark technology we want to stress that feasibility can be understood as a *descriptive statement* (if $(\mathbf{x}, \mathbf{y}) \in T$, then $(\mathbf{x}, \mathbf{y})$ is technically feasible) or a *normative statement* (if $(\mathbf{x}, \mathbf{y}) \in T$, then $(\mathbf{x}, \mathbf{y})$ can be used as a benchmark in a performance norm whether or not $(\mathbf{x}, \mathbf{y})$ is technically feasible or not). The normative interpretation is particularly relevant in the context of incentive regulation, which forms the empirical motivation of this study.

The three classical axioms of the benchmark technology T considered in this paper include the following (see, e.g., Koopmans, 1951; Afriat, 1972; Färe and Primont, 1995):

A1. Free disposability of inputs and outputs: $T = T + \Re_+^m \times \Re_-^s$.

A2. Convexity: T = Conv(T).

A3. Constant returns to scale (CRS): $T = \lambda T,\ \lambda > 0$.

The distinction of descriptive versus normative interpretation of the benchmark technology also concerns the axioms. For example, the descriptive interpretation of CRS suggests that proportionate scaling of $(\mathbf{x}, \mathbf{y})$ by any arbitrarily small or large multiplier $\lambda > 0$ results as a technically feasible production plan $(\lambda\mathbf{x}, \lambda\mathbf{y})$. Clearly, the descriptive interpretation of CRS is unrealistic. However, we often find the normative interpretation of CRS appealing: if $(\mathbf{x}, \mathbf{y})$ is a valid benchmark, then so is $(\lambda\mathbf{x}, \lambda\mathbf{y})$. In other words, evaluated firms cannot plea for diseconomies of scale either due to too small or too large scale of operation. In incentive regulation, imposing CRS for the benchmark technology gives firms an incentive to merge or divest, or more generally, operate at the most productive scale size.

2.2 Directional distance function

The directional distance function (DDF) introduced by Chambers et al. (1996, 1998) is defined as

$$\vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}^x, \mathbf{g}^y) = \sup\{\theta \mid (\mathbf{x} - \theta\mathbf{g}^x, \mathbf{y} + \theta\mathbf{g}^y) \in T\},$$

where $(\mathbf{g}^x, \mathbf{g}^y) \in \Re_+^{m+s}$ is a direction vector. The DDF is a general functional representation of the technology: assuming *g*-disposability (see, e.g., Färe et al., 2005), the production possibility set *T* is

$$T = \{(\mathbf{x}, \mathbf{y}) \in \Re_+^{m+s} \mid \vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}^x, \mathbf{g}^y) \geq 0\} \text{ for any } (\mathbf{g}^x, \mathbf{g}^y) \in \Re_+^{m+s}.$$

While the DDF is commonly used as a measure of technical inefficiency, we emphasize that in this paper we are primarily interested in the DDF as a functional representation of production possibilities. The DDF contains the conventional radial input or output distance functions as its special cases. Specifying the direction vector as $(\mathbf{g}^x, \mathbf{g}^y) = (\mathbf{x}, \mathbf{0})$, then the DDF is equal to one minus the input distance function. Similarly, the output distance function is obtained by setting $(\mathbf{g}^x, \mathbf{g}^y) = (\mathbf{0}, \mathbf{y})$.

The DDF inherits the axiomatic properties of the production set *T* (see Chambers et al., 1998, Lemma 2.2). Two fundamental properties that any DDF must satisfy include the following:

A4. Translation property:

$$\vec{D}_T(\mathbf{x} - \alpha\mathbf{g}^x, \mathbf{y} + \alpha\mathbf{g}^y, \mathbf{g}^x, \mathbf{g}^y) = \vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}^x, \mathbf{g}^y) - \alpha.$$

A5. Homogeneity of degree -1 in the direction vector:

$$\vec{D}_T(\mathbf{x},\mathbf{y},\lambda\mathbf{g}^x,\lambda\mathbf{g}^y) = (1/\lambda)\cdot\vec{D}_T(\mathbf{x},\mathbf{y},\mathbf{g}^x,\mathbf{g}^y),\ \lambda > 0 .$$

These properties are crucial for the internal consistency of the DDF. Translation property A4 is an additive analogue of the linear homogeneity property of the input and output distance functions. Other notable properties of the DDF include the following. If inputs and outputs are freely disposable, then the DDF is a continuous function that is monotonically increasing in inputs and monotonically decreasing in outputs. If production possibility set T is convex, then DDF is globally concave in inputs and outputs. If the technology exhibits CRS, then the DDF is linearly homogenous in (**x**,**y**):

$$\vec{D}_T(\lambda\mathbf{x},\lambda\mathbf{y},\mathbf{g}^x,\mathbf{g}^y) = \lambda\cdot\vec{D}_T(\mathbf{x},\mathbf{y},\mathbf{g}^x,\mathbf{g}^y),\ \lambda > 0 .$$

The DDF also has the following property.

Proposition 1: *If inputs and outputs are freely disposable (A1), then the directional distance function* $\vec{D}_T$ *is monotonically decreasing in* $(\mathbf{g}^x,\mathbf{g}^y)$ *for all* $(\mathbf{x},\mathbf{y}) \in T$ *and monotonically increasing in* $(\mathbf{g}^x,\mathbf{g}^y)$ *for all* $(\mathbf{x},\mathbf{y}) \notin T$.

To our knowledge, this monotonicity property has not been formally proved before (Fukuyama, 2003, briefly notes the first part of Proposition 1 without a proof). This result illustrates that changing the direction vector affects the values of $\vec{D}_T$, but $\vec{D}_T$ remains a complete characterization of the technology for any arbitrary direction vector $(\mathbf{g}^x,\mathbf{g}^y) \in \Re_+^{m+s}$. We also note that concavity of $\vec{D}_T$ in (**x**,**y**) under axiom A2 does not generally imply concavity or convexity with respect to the direction vector $(\mathbf{g}^x,\mathbf{g}^y)$.

2.3 Data generating process (DGP)

The previous sub-sections consider the set theoretic and functional representations of the benchmark technology and their axiomatic properties. We consider a sample of n observed input-output vectors denoted by $(\mathbf{x}_i,\mathbf{y}_i)$ indexed by i = 1,2,…,n randomly drawn from a true population. This subsection describes a probabilistic model that generates the observed data and state our assumptions regarding the DGP.

Virtually all behavioral hypotheses known in the theory of the firm assume optimizing behavior. Without restricting to any specific behavioral hypothesis such as profit maximization or cost minimization, we assume that a unique optimal solution to the firm's problem exist, and denote the optimal input-output vector of firm i by $(\mathbf{x}_i^*,\mathbf{y}_i^*)$. That is, given the specific objective function and constraints of firm *i*, $(\mathbf{x}_i^*,\mathbf{y}_i^*)$ is objectively the best solution that a perfectly rational manager with perfect information would always choose. Note that the input-output prices may endogenously depend on the firm's input-output quantities and the prices may be random variables, but if a rational manager solves a well-defined optimization problem (e.g., maximize the expected net present value of future profit flows as in Olley and Pakes, 1996), the optimal solution $(\mathbf{x}_i^*,\mathbf{y}_i^*)$ is both exogenously given and deterministic to firm *i*. We do recognize that real managers are not always perfectly rational or perfectly informed, but at this point, we only assume that a unique deterministic optimal solution $(\mathbf{x}_i^*,\mathbf{y}_i^*)$ exists. Since $(\mathbf{x}_i^*,\mathbf{y}_i^*)$ are vectors of constants, we have $Var(\mathbf{x}_i^*)=\mathbf{0}$ and $Var(\mathbf{y}_i^*)=\mathbf{0}$.

We assume that the objective function is monotonically increasing in outputs **y** and monotonically decreasing in inputs **x**. Together with axiom A1 this implies that the optimal solutions $(\mathbf{x}_i^*,\mathbf{y}_i^*)$ lie on the boundary of the production possibility set *T*, that is,

$$\vec{D}_T(\mathbf{x}_i^*,\mathbf{y}_i^*,\mathbf{g}^x,\mathbf{g}^y)=0 \quad \forall(\mathbf{g}^x,\mathbf{g}^y)>(\mathbf{0},\mathbf{0}), i=1,\ldots,n.$$

It is worth noting that in our DGP the production possibility set *T* is perfectly deterministic, and hence so is its functional representation, the DDF.

As noted above, real managers may have imperfect information about the technology and the (probability density of) input-output prices, real managers make errors in both optimization of production plans as well as implementation of those plans. To model the impacts of all such managerial failures on observed input-output data in a probabilistic manner, we follow the SFA literature and posit a random inefficiency term denoted by u_i. Inefficiency term $u_i \geq 0$ is nonnegative and has a constant finite mean $\mu \geq 0$ and a constant finite variance σ_u^2. Since the real world data is typically subject to omitted variables, unobserved heterogeneity, measurement errors, and other random noise, we also posit another random variable v_i referred to as the noise term. We assume v_i is symmetric with a unique mode at zero and a constant finite variance σ_v^2. We define the composite error term as $\varepsilon_i = u_i + v_i$. Similar to the standard SFA models, the inefficiency and noise terms (and hence the composite errors ε_i) are assumed to be homoskedastic and independent of each other, implying $\sigma_\varepsilon^2 = \sigma_u^2 + \sigma_v^2$. Further, random ε_i are independent of the optimal solutions $(\mathbf{x}_i^*,\mathbf{y}_i^*)$.

In contrast to the conventional SFA models, however, we do not make any specific distributional assumptions regarding inefficiency or noise. Following Hall and Simar (2002), we assume that the asymmetric inefficiency term $u_i \geq 0$ has a density f_u with a jump discontinuity at 0 and expected value μ. The noise term v_i has a unimodal density with a unique mode at zero. Hall and Simar (2002) also make a technical assumption that σ_v^2 approaches to zero asymptotically, which is required for proving consistency of their estimator.

Note that the optimal solutions $(\mathbf{x}_i^*, \mathbf{y}_i^*)$ are vectors in $\Re_+^{m+s}$, but the composite error term ε_i and its components u_i and v_i are scalars. To assign the scalar-valued inefficiency and noise to *m* input and *s* output variables, our DGP employs a direction vector $(\mathbf{g}^x, \mathbf{g}^y)$. Specifically, the optimal solutions $(\mathbf{x}_i^*, \mathbf{y}_i^*)$ are perturbed by a random variable ε_i in the exogenously given direction $(\mathbf{g}^x, \mathbf{g}^y)$, which yields the observed input-output vectors as a result, specifically,

$$\mathbf{x}_i = \mathbf{x}_i^* + \varepsilon_i \mathbf{g}^x \quad \forall i = 1, \ldots, n,$$

$$\mathbf{y}_i = \mathbf{y}_i^* - \varepsilon_i \mathbf{g}^y \quad \forall i = 1, \ldots, n.$$

The use of a direction vector to govern the DGP is a novel idea of this paper. We must stress that in our DGP $(\mathbf{g}^x, \mathbf{g}^y)$ is not an arbitrary parameter vector that a researcher is free to specify, but rather, a parameter vector that governs inefficiency and noise in the probabilistic DGP. Note that if $g_k^x = 0$ for some input *k*, then input *k* is free from inefficiency and noise, and we have $x_{ik} - x_{ik}^* = 0$. The larger the value of g_k^x, the larger the expected value and variance of the deviation $x_{ik} - x_{ik}^*$. Our DGP allows for a possibility that all inputs and outputs are subject to inefficiency and noise, and hence, all inputs and outputs are *endogenous* variables (using the standard econometric terminology).

2.4 Direction Selection for Estimation

Since the observed input-output vectors $(\mathbf{x}_i, \mathbf{y}_i)$ depend on random variables ε_i, $(\mathbf{x}_i, \mathbf{y}_i)$ are random variables themselves. The expected value and variance of $(\mathbf{x}_i, \mathbf{y}_i)$ are (respectively)

$$E(\mathbf{x}_i, \mathbf{y}_i) = (\mathbf{x}_i^* + \mu \mathbf{g}^x, \mathbf{y}_i^* - \mu \mathbf{g}^y),$$

$$Var(\mathbf{x}_i, \mathbf{y}_i) = \sigma_\varepsilon^2 (\mathbf{g}^x, \mathbf{g}^y).$$

Although a researcher cannot directly observe the underlying direction vector governing the DGP, we can use the variances of the input-output variables to inform an empirical specification of the direction vector. Note that in our DGP the variances of all input and output variables are identical and constant across firms because the composite error terms ε_i are scalar, which allows an empirical specification of the direction vector. Some alternatives will be briefly examined next.

To gain intuition, let us first assume $E(\mathbf{x}_i, \mathbf{y}_i)$ to be constant across all firms: consider a sample of profit maximizing firms operating a competitive market where all firms take the same input-output prices as given, such that the optimal solution $(\mathbf{x}_i^*, \mathbf{y}_i^*)$ is exactly the same for all firms. Then given the direction vector $(\mathbf{g}^x, \mathbf{g}^y)$, this implies that $E(\mathbf{x}_i, \mathbf{y}_i)$ is constant across firms. Therefore, the sample variances of input-output variables provide an unbiased and consistent estimator of $\sigma_\varepsilon^2(\mathbf{g}^x,\mathbf{g}^y)$. Note further that since the DDF is homogenous of degree -1 in the direction vector by axiom A5, we can harmlessly rescale the direction vector by an arbitrary scalar. Therefore, we can set the elements of $(\mathbf{g}^x,\mathbf{g}^y)$ equal to the sample variances of the corresponding input-output variables, and simply ignore the unknown variance parameter σ_ε^2 in the specification of the direction vector. To relax the strong assumption of a constant $E(\mathbf{x}_i,\mathbf{y}_i)$ across all firms, one might assume that $E(\mathbf{x}_i,\mathbf{y}_i)$ is constant across subsets of firms in a cross-sectional setting. Of course, the subsets should be large enough to be able to estimate $\sigma_\varepsilon^2(\mathbf{g}^x,\mathbf{g}^y)$ by the sample variances of input-output vectors for each subset.

If one has panel data $(\mathbf{x}_{it},\mathbf{y}_{it}), t=1,\dots,T$, another possibility is to assume that $E(\mathbf{x}_{it},\mathbf{y}_{it})$ is constant over time (at least by approximation). Then, one can calculate the differences from the mean, $\bar{\mathbf{x}}_i = \sum_{t=1}^{T}\mathbf{x}_{it}$ and $\bar{\mathbf{y}}_i = \sum_{t=1}^{T}\mathbf{y}_{it}$, as $(\mathbf{x}_{it}-\bar{\mathbf{x}}_i,\mathbf{y}_{it}-\bar{\mathbf{y}}_i)$, and use sample variances, $\sigma_x^2(\mathbf{g}^x)=\frac{1}{T}\sum_{t=1}^{T}(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)^2$ and $\sigma_y^2(\mathbf{g}^y)=\frac{1}{T}\sum_{t=1}^{T}(\mathbf{y}_{it}-\bar{\mathbf{y}}_i)^2$ of each element of the resulting vector as an unbiased and consistent estimator of $\sigma_\varepsilon^2(\mathbf{g}^x,\mathbf{g}^y)=\left[\sigma_x^2(\mathbf{g}^x),\sigma_y^2(\mathbf{g}^y)\right]$. This empirical specification strategy essentially requires that the expected value of the input-output vector is constant across multiple data points over time such that the sample variance can be used as an estimator of $\sigma_\varepsilon^2(\mathbf{g}^x,\mathbf{g}^y)$. This strategy requires the rather strong assumption that the firms input and output levels are approximately constant over time. Therefore, an alternative approach will be considered in Section 5. However, the question of how best to recover the true direction associated with the proposed DGP in cross-sectional data remains and open and interesting research question.

Finally, some restrictive features of the DGP described above are worth noting. The assumption of a common direction vector $(\mathbf{g}^x,\mathbf{g}^y)$ that is both deterministic and constant across all firms is clearly restrictive. Our defense is that having a formal statistical model with clearly defined assumptions is far better than having no model at all. In fact, only few formal descriptions of a probabilistic DGP with noise are available in the multiple output setting. Banker and Maindiratta (1992) and Simar (2007) consider a radial DGP where the inputs (or outputs) are taken as exogenously given while all the outputs (inputs) are multiplied by a random variable $\exp(\varepsilon_i)$, resulting as radial perturbations along the ray from the origin. Such a radial model of inefficiency and noise is not free of restrictive assumptions (e.g., exogenous inputs, a specific radial direction for inefficiency and noise). Most importantly, the radial multiplicative model of inefficiency and noise is not compatible with the additive, non-radial orientation of the DDF. In our view, the DGP described above is an additive analogue of the radial models by Banker and Maindiratta (1992) and Simar (2007), adapted from the radial framework to the DDF setting.

Varian (1985) and Kuosmanen et al. (2007) consider a more general multivariate DGP where each input and output variable is perturbed by its own random composite error term, that is, there are *m*+*s* composite error terms instead of just one. We note that if the direction vector $(\mathbf{g}^x, \mathbf{g}^y)$ is specified as a random variable that is allowed to have different realizations across firms, then our directional DGP becomes observationally equivalent to the multivariate DGP of Varian (1985) and Kuosmanen et al. (2007). While we admit that the DGP considered in this paper is somewhat restrictive, we note extending the framework to more general settings as an interesting challenge for future research.

3. Econometric model

To establish an explicit connection between the functional representation of technology introduced in Section 2.2 and the DGP introduced in Section 2.3, we first state the following result.

Proposition 2: *If the observed data are generated according to the DGP described in Section 2.3, then the value of the DDF in observed data* $(\mathbf{x}_i, \mathbf{y}_i)$ *point is equal to the realization of the random variable* ε_i *, specifically,*

$$\vec{D}_T(\mathbf{x}_i, \mathbf{y}_i, \mathbf{g}^x, \mathbf{g}^y) = \varepsilon_i \quad \forall i .$$

Interestingly, although the DDF itself is a deterministic function with some known axiomatic properties, its value at a given observed data point is a random variable due to the inefficiency and noise contained in the observed data. This result verifies that the DGP described in Section 2.3 is structurally consistent with the DDF representation of the technology.

To develop the regression equation, we apply the translation property to obtain one of the inputs or outputs as the dependent variable. We could arbitrarily choose any of the input or output variables as the dependent variable, but for the sake of notation, we specify the first output variable y_{1i} as the dependent variable. Note that all inputs and outputs are potentially endogenous.

Assuming $g_1^y \neq 0$, we apply the translation property A4 and set $\alpha = -y_{1i} / g_1^y$ to obtain the following equality

$$\vec{D}_T(\mathbf{x}_i + (y_{1i} / g_1^y)\mathbf{g}^x, \mathbf{y}_i - (y_{1i} / g_1^y)\mathbf{g}^y, \mathbf{g}^x, \mathbf{g}^y) = \vec{D}_T(\mathbf{x}_i, \mathbf{y}_i, \mathbf{g}^x, \mathbf{g}^y) + (y_{1i} / g_1^y) .$$

Applying Proposition 2, and reorganizing terms, we have the regression equation

$$y_{1i}/g_1^y = \vec{D}_T(\mathbf{x}_i + (y_{1i}/g_1^y)\mathbf{g}^x, \mathbf{y}_i - (y_{1i}/g_1^y)\mathbf{g}^y, \mathbf{g}^x, \mathbf{g}^y) - \varepsilon_i \quad (1)$$

The negative sign of the composite error term is due to the fact that the DDF has positive values in the interior of the production possibility set *T* and negative values outside *T*. Note that the explanatory variables of (1) are potentially endogenous because the observed inputs and outputs are perturbed by ε_i. To estimate equation (1), we first need to eliminate this endogeneity.

To simplify the notation, we introduce the following partial differences:

$$\vec{\mathbf{x}}_i = \mathbf{x}_i + (y_{1i}/g_1^y)\mathbf{g}^x,$$

$$\vec{\mathbf{y}}_i = \mathbf{y}_i - (y_{1i}/g_1^y)\mathbf{g}^y.$$

Partial differences are similarly used in many areas of econometrics (consider, e.g., estimation of autoregressive models by feasible generalized least squares). Using these transformations, we can state the regression equation as

$$y_{1i}/g_1^y = \vec{D}_T(\vec{\mathbf{x}}_i, \vec{\mathbf{y}}_i, \mathbf{g}^x, \mathbf{g}^y) - \varepsilon_i.$$

Proposition 3: *If the observed data are generated by the DGP described in Section 2.3, then the transformed input-output variables* $(\vec{\mathbf{x}}_i, \vec{\mathbf{y}}_i)$ *are uncorrelated with the error term* ε_i*, that is,*

$$Cov(\varepsilon_i, \vec{\mathbf{x}}_i) = \mathbf{0}\ \forall i \text{ and } Cov(\varepsilon_i, \vec{\mathbf{y}}_i) = \mathbf{0}\ \forall i.$$

Intuition behind this result is the following. Since y_{1i} contains inefficiency and noise, adding $(y_{1i}/g_1^y)\mathbf{g}^x$ to the observed inputs and by subtracting $(y_{1i}/g_1^y)\mathbf{g}^y$ from the observed outputs will cancel out the inefficiency and noise contained in the other input-output variables (see the proof for details). Hence, it is possible to eliminate the endogeneity problem. To gain intuition, it may be helpful to compare the proposed approach with the two-stage least squares (2SLS) approach where the endogenous variables are first regressed on external instruments, and subsequently, the fitted values of the endogenous variables are used as explanatory variables in the original regression equation. In our setting, the transformed input-output vectors $(\vec{\mathbf{x}}_i, \vec{\mathbf{y}}_i)$ serve the same function as the fitted values of endogenous variables in the 2SLS approach, except that constructing vectors $(\vec{\mathbf{x}}_i, \vec{\mathbf{y}}_i)$ does not require estimation of any auxiliary regression.

4. StoNED estimator

This section describes the StoNED estimator that builds directly on the axiomatic properties of the DDF does not require any functional form assumptions. The proposed estimation is a step-wise method. The first step is to apply convex nonparametric least squares (CNLS) (Kuosmanen, 2008) to estimate the *conditional mean*

$$E(y_{1i} / g_1^y \,|\, \vec{\mathbf{x}}_i, \vec{\mathbf{y}}_i) = \vec{D}_T(\vec{\mathbf{x}}_i, \vec{\mathbf{y}}_i, \mathbf{g}^x, \mathbf{g}^y) - \mu \, .$$

Note that the properties of the DGP imply that the conditional mean is simply the DDF minus a constant μ = $E(u)$. The second step applies kernel deconvolution (Hall and Simar, 2002; Goldenshluger and Tsybakov, 2004) to estimate μ. The third step combines the results of the first two steps to estimate the DDF.

4.1 Convex nonparametric least squares

We first estimate the conditional mean $E(y_{1i} / g_1^y \,|\, \vec{\mathbf{x}}_i, \vec{\mathbf{y}}_i)$ of output 1 by solving the following quadratic programming problem

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\varepsilon}^{\circ}} \sum_{i=1}^{n} (\varepsilon_i^{\circ})^2$$

subject to

$$\begin{aligned}
& y_{1i} / g_1^y = \alpha_i + \boldsymbol{\beta}_i' \vec{\mathbf{x}}_i - \boldsymbol{\gamma}_i' \vec{\mathbf{y}}_i + \varepsilon_i^{\circ} \;\; \forall i \\
& \alpha_i + \boldsymbol{\beta}_i' \vec{\mathbf{x}}_i - \boldsymbol{\gamma}_i' \vec{\mathbf{y}}_i \le \alpha_h + \boldsymbol{\beta}_h' \vec{\mathbf{x}}_i - \boldsymbol{\gamma}_h' \vec{\mathbf{y}}_i \;\; \forall i, h \\
& \boldsymbol{\beta}_i' \mathbf{g}^x + \boldsymbol{\gamma}_i' \mathbf{g}^y \le 1 \;\; \forall i \\
& \boldsymbol{\beta}_i \ge \mathbf{0}, \boldsymbol{\gamma}_i \ge \mathbf{0} \;\; \forall i
\end{aligned}$$

The coefficients $\alpha_i, \boldsymbol{\beta}_i, \boldsymbol{\gamma}_i$ characterize a tangent hyperplane of the estimated DDF for the observation i, and ε_i° is an estimator of $\mu - (v_i + u_i)$. Considering y_{1i} / g_1^y as the dependent variable and the transformed variables ($\vec{\mathbf{x}}_i, -\vec{\mathbf{y}}_i$) as the independent variables, this estimator is directly equivalent to the standard CNLS estimator by Kuosmanen (2008). Statistical consistency of this estimator is proved by Seijo and Sen (2011) and Lim and Glynn (2012). The only additional component to the standard CNLS estimator is the third constraint, which enforces the translation property and monotonicity with respect to the first output variable.

We emphasize that the above CNLS estimator is invariant to the choice of output y_1 as the dependent variable. We can equally well choose some other output or an input as the dependent variable, and obtain exactly the same estimates. The only restriction is that the element of the direction vector corresponding to the dependent variable cannot be equal to zero: we cannot choose an input-output variable that is free from inefficiency and noise as a dependent variable. In applications that we are familiar with it is quite easy to recognize beforehand which inputs and outputs are noisy and which ones are not.

4.2 Estimating the expected inefficiency

The second step is to estimate the expected inefficiency μ. Kuosmanen and Kortelainen (2012) propose to complement the CNLS regression by the method of moments (Aigner et al., 1977) or quasi-likelihood estimation (Fan et al., 1996) that can identify the constant μ based on the skewness of the CNLS residuals. However, these approaches require additional distributional assumptions for both the inefficiency and noise terms (e.g., half-normal inefficiency and normally distributed noise). Further, in these methods the identification of μ is essentially based on the skewness of the residuals (Waldman, 1982). In this paper we use a robust nonparametric estimator that does not depend on the skewness of the residuals or specific distributional assumptions.

Fully nonparametric estimation of the expected inefficiency μ based on the CNLS residuals is possible by applying nonparametric kernel deconvolution, as shown by Hall and Simar (2002). In this approach, the identification of μ is based on the unknown density of the composite error term. Since the CNLS estimator produces residuals e_i° that are consistent estimators of $\varepsilon_i^\circ = \mu - (v_i + u_i)$, we can apply the kernel density estimator for estimating the density function f_{e°. Formally,

$$\hat{f}_{e^\circ}(z) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{z - e_i^\circ}{h}\right),$$

where $K(\cdot)$ is a compactly supported kernel and h is a bandwidth.

Hall and Simar (2002) show that the first derivative of the density function of the composite error term (f'_ε) is proportional to that of the inefficiency term (f'_u) in the neighborhood of μ. This is due to the assumption that f_u has a jump discontinuity at zero. Therefore, a robust nonparametric estimator of expected inefficiency μ is obtained as

$$\hat{\mu} = \arg\max_{z \in \Im} (\hat{f}'_{e^\circ}(z)),$$

where $\Im$ is a closed interval in the right tail of $f_{\varepsilon^\circ}(\cdot)$.

To implement the procedure empirically, a bandwidth must be chosen and $\Im$ must be defined. Different bandwidth selection criteria are known in the literature of deconvolution kernel estimation. Delaigle and Gijbels (2004) discuss alternative bandwidth selection criteria for data contaminated with noise, including the normal reference method, the plug-in method, the cross-validation and the bootstrap method, and compare them using Monte Carlo simulations. They find that the plug-in and the bootstrap methods outperform the cross-validation method.

4.3 Estimating the DDF

Once $\hat{\mu}$ has been obtained, the conditional mean estimated in the first step can be shifted upward to obtain an estimator of the DDF. But first, note that even though the CNLS estimator yields unique predictions $\hat{y}_{1i} / g_1^y = \alpha_i + \boldsymbol{\beta}_i'\vec{\mathbf{x}}_i - \boldsymbol{\gamma}_i'\vec{\mathbf{y}}_i$ in the observed data points, the coefficients $\alpha_i, \boldsymbol{\beta}_i, \boldsymbol{\gamma}_i$ obtained as the optimal solution to the CNLS problem are not necessarily unique [this is directly analogous to the fact that the multiplier weights of data envelopment analysis (DEA) are non-unique]. To resolve the non-uniqueness, Kuosmanen and Kortelainen (2012) appeal to the minimum extrapolation principle and propose to estimate the minimum envelopment of the fitted values of the CNLS estimator. In the similar vein, our estimator of the DDF is given by

$$\hat{D}(\mathbf{x}, \mathbf{y}, \mathbf{g}^x, \mathbf{g}^y) = \min_{\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}} \left\{ \alpha + \boldsymbol{\beta}'\mathbf{x} - \boldsymbol{\gamma}'\mathbf{y} \,\middle|\, \alpha + \boldsymbol{\beta}'\hat{\mathbf{x}}_i^* - \boldsymbol{\gamma}'\hat{\mathbf{y}}_i^* \geq 0 \;\forall i; \boldsymbol{\beta}'\mathbf{g}^x + \boldsymbol{\gamma}'\mathbf{g}^y = 1; \boldsymbol{\beta} \geq 0; \boldsymbol{\gamma} \geq 0 \right\},$$

Where $\hat{\mathbf{x}}_i^* = \mathbf{x}_i - (e_i^\circ + \hat{\mu})\mathbf{g}_i^y$ and $\hat{\mathbf{y}}_i^* = \mathbf{y}_i + (e_i^\circ + \hat{\mu})\mathbf{g}_i^y$ are the estimated projection points on the frontier of *T*. Using the estimated projection points as the reference units, we can compute $\hat{D}(\mathbf{x}, \mathbf{y}, \mathbf{g}^x, \mathbf{g}^y)$ using the standard DEA formulation of the DDF (e.g., Fukuyama, 2003), and it can be solved by linear programming. One can compute all coefficients ($\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}$) by using a convex hull algorithm (e.g., Olesen and Petersen, 2003; Appa and Williams, 2006) to obtain an explicit representation of the estimated DDF as a piece-wise linear function.

It is easy to verify that the estimator $\hat{D}(\mathbf{x}, \mathbf{y}, \mathbf{g}^x, \mathbf{g}^y)$ of the production frontier satisfies free disposability A1, convexity A2, translation property A4, and the homogeneity property A5. The CRS property A3 is imposed by restricting $\alpha_i = 0 \;\forall i$. We stress that the projection points $(\hat{\mathbf{x}}_i^*, \hat{\mathbf{y}}_i^*)$ obtained from the CNLS problem are consistent with these axiomatic properties. The auxiliary DEA step does not influence the DDF estimates of the observed firms: it is only used for computing the minimum envelopment of the fitted values, which allows us to compute the value of the DDF for any real valued production plan $(\mathbf{x}, \mathbf{y})$ and for any arbitrary direction $(\mathbf{g}^x, \mathbf{g}^y) \in \Re_+^{m+s}$.

We are primarily interested in the DDF as a representation of technology. Once the frontier is estimated, it is possible to use the estimated DDF for *ex post* efficiency analysis. One can use the estimated DDF for gauging the distance to the frontier in any direction $(\mathbf{g}^x, \mathbf{g}^y)$, which may be firm-specific. In general, we would expect that the direction vector used for efficiency analysis would be different than the direction vector used in the CNLS estimation of the DDF. The direction vector used in CNLS for estimating the benchmark technology should be specified to match the underlying DGP. However, the direction vector applied in *ex post* performance comparisons can be specified differently to conform to the objectives of the firm managers, the regulator, or other stakeholders, depending on the purposes of efficiency analysis.

Estimating the production frontier and subsequently performing an *ex post* efficiency analysis is a novel proposal of our estimation procedure. Several papers address the issue of selecting the direction vector to estimate a directional distance functions, see for example Daraio and Simar (2014) who proposed a data driven method, Fare et al. (2013) who propose to select the direction $\mathbf{g}^y$ such that the observed data point is as close to the production function as possible when measured in units of $\mathbf{g}^y$ where $\sum_{y=1}^{s}\mathbf{g}^y = 1$,or Zofio et al. (2013) and Atkinson and Tsionas (2016) who propose a method consistent with profit maximization. We see all of these methods as alternative methods to perform an ex post efficiency analysis which could be potentially useful depending on the objective of the efficiency analysis. However, it is not clear they could be use in the estimation of the directional distance production function because they are not consistent with DGP describe in this paper and no DGP has been proposed in the papers in which the estimators are developed.

While the estimated DDF can be useful for relative performance comparisons of firms, we emphasize that the *ex post* distance measures capture both inefficiency and noise because all data points are subject to noise by assumption. Indeed, it is incorrect to interpret a distance from an observed data point to the frontier as an estimate of inefficiency u_i . The density of u_i can be estimated by kernel smoothing (Horrace and Parmeter, 2011), which can be useful for interval estimation and inferences on the aggregate levels of industries or groups of firms. However, consistent point estimators of the firm specific u_i are impossible: without further structure or assumptions the realization of a random variable cannot be estimated based on a single noisy data point. This does not mean that efficiency analysis is meaningless in stochastic setting: for example, estimated frontier production function can be useful for benchmarking purposes. We next consider an application to incentive regulation where the main interest is in the estimation of a cost norm.

5. Application to electricity distribution firms

Benchmark regulation of electricity distribution networks is one of the most significant real world applications of frontier estimation techniques. Several regulators across the world apply the axiomatic deterministic DEA to estimate efficiency improvement targets, and some apply econometric techniques such as SFA (see, e.g., Bogetoft and Otto, 2011, Ch. 10, for a review). Finland was the first country to adopt the StoNED method in use and apply it systematically to real world incentive regulation since 2012 (Kuosmanen, 2012). The application presented in this section was originally developed in 2013 in order to help the Finnish regulator to further improve their benchmarking method. As discussed in more detail below, several methodological developments introduced in the previous sections have been adopted in actual use in the regulation model for Finnish electricity distribution firms for the time period 2016 – 2023. However, we stress that the model presented in this section is not exactly identical to the benchmarking model applied by the Finnish regulator, it has been simplified in a number of ways to allow emphasis on the modeling of multi-output production.

We present a simplified model here mainly for illustrative purposes.

We examine the data set of 89 Finnish electricity distribution firms described in Kuosmanen (2012) (available in its online supplement). We consider the following input and output variables, which are fairly standard in the benchmark regulation of electricity distribution.

Input variables:
x_1 = operational expenditures (OPEX, 1,000 €);
x_2 = capital expenditures (CAPEX, 1,000 €).

Output variables:
y_1 = energy transmitted (GWh);
y_2 = length of lines (km);
y_3 = number of customers.

We can interpret OPEX as a proxy of labor input, and CAPEX as a proxy of capital input. The measurement of labor input in this industry is difficult because a large proportion of work has been outsourced to subcontractors, and hence the labor hours and the wage bills of the firms do not represent the actual labor input. An important unresolved issue in this literature concerns the question of whether the single input variable should be defined as the total expenditures (TOTEX = $x_1 + x_2$) or the operating expenditures (OPEX) (e.g., Jamasb and Pollit, 2007; Bogetoft and Otto, 2011, Ch. 10). Distribution firms can convincingly argue that the capital expenditure (CAPEX), which forms a part of TOTEX, is fixed in the short run. Indeed, investments in the power grid are expected to serve over several decades, which makes it difficult to adjust the CAPEX component in the short run (e.g., in Finland, one regulation period is four years). However, omitting the CAPEX component can give wrong incentives to overinvest since OPEX can be substituted, at least to some extent, by CAPEX. Omitting CAPEX would favor those firms that can invest capital in labor saving technologies to drive down the OPEX component. In Finland, Kuosmanen (2012) recommended the use of TOTEX, but the regulator decided to use OPEX, omitting the CAPEX component.

The use of the DDF that facilitates multiple inputs and multiple outputs can help to resolve the question of OPEX versus TOTEX regulation. For example, a regulator may prefer to measure efficiency in terms of OPEX, treating CAPEX as a fixed input, which can be done by setting the direction vector for efficiency evaluation as $\mathbf{g}^x = (1,0)$ for the inputs. As for the output variables, local distribution monopolies have little effect on the customer demand. For example, the distribution companies have a legal obligation to connect all customers, and they obviously cannot transmit more power than what is demanded by the customers. On the other hand, there are monetary sanctions if distribution firms fail to meet demand. By these arguments, setting direction vector for efficiency evaluation as $\mathbf{g}^y = (0,0,0)$ for the outputs seems an appropriate and meaningful specification from the regulation point of view. This specification allows the regulator to measure the cost saving potential in OPEX, keeping CAPEX and the output variables fixed at their observed level. Indeed, this is the direction vector that the Finnish regulator adopted in use in their benchmarking model since 2016. Although this can be a meaningful direction for the *ex post* performance analysis, we first need to estimate the DDF using observed data that are subject to noise. From the econometric point of view, ignoring noise in all variables except OPEX would be a strong assumption.

In the present application, both input variables are clearly noisy. The main component of OPEX is labor expenses (both in-house and outsourced labor). Thunderstorms, heavy snowfalls, and other random weather events that cause damage to power lines will directly affect the OPEX expenses. CAPEX is measured based on the replacement value of the distribution network using the linear depreciation. Such an accounting measure of capital, which requires monetary valuation of thousands of network components of different vintages, will obviously contain a lot of noise. Therefore, we set both elements of $\mathbf{g}^x$ positive in the estimation of the DDF.

In contrast, the output variables are measured with high precision, and they capture both the actual power transmission (y_1) and the capacity (outputs y_2 and y_3). We also assume that firms take the demand for electric power as given. Therefore, we assume the outputs to be exogenous and specify $\mathbf{g}^y = (0,0,0)$. Thus, we need to take one of the input variables, say OPEX (x_1), as the dependent variable that is normalized and moved to the left-hand side of the regression equation.

Since we have two noisy input variables, we can harmlessly normalize $g_1^x = 1$. The remaining question is then how to specify g_2^x for the CAPEX input? A critical assumption not emphasized thus far is homoscedasticity of the composite error term. This assumption is required in the stepwise estimation strategy for shifting the estimated conditional mean upward to estimate the frontier represented by the DDF. Therefore, our empirical strategy to specify g_2^x is based on the homoscedasticity assumption and a simple grid search strategy.

Appendix 2 (available in the online supplement) describes the CNLS formulation used in this application and describes our empirical strategy of specifying g_2^x. In summary, we compute the CNLS estimates using different values of g_2^x, then apply the White test of heteroscedasticity, and choose g_2^x that yields the lowest value of the White's test statistic. This empirical strategy results as the specification $g_2^x = 0.25$.

Table 1 reports some descriptive statistics of the estimated shadow prices of inputs and outputs. The estimated DDF is a piece-wise linear function consisting of facets characterized by the multipliers **β** and **γ**. Recall we allow coefficients **β** and **γ** to be firm-specific, but in practice, the estimated coefficients are clustered to a smaller number of facets. Note that inefficiency loss is here measured on monetary scale (€). The output coefficients **γ** have compelling price interpretation with well-defined units of measurement whereas input coefficients **β** are unit invariant multipliers. For example, increasing OPEX by one euro increases inefficiency by 86 cents on average.

Table 1: Marginal effects of inputs (β coefficients) and outputs (γ) on the DDF

Variable	mean	st.dev.	min	max
β_1 OPEX	0.857	0.219	0.010	1.000
β_2 CAPEX	0.572	0.878	0.000	3.961
γ_1 energy transmission (c / kWh)	3.529	1.597	0.060	5.955
γ_2 length of lines (€ / km)	410.8	298.5	0.0	1 918
γ_3 customers (€ / customer)	35.01	34.60	0.00	21.72

Next, we apply the kernel density estimator to the CNLS residuals to estimate the expected value of inefficiency (see Section 4.2 for details). Figure 1 plots the estimated density function and its first derivative. The derivative function achieves its maximum at point -1.73 (indicated by the vertical line in Figure 1), and hence our estimate of the expected value of inefficiency is $\hat{\mu} = 1.73$.

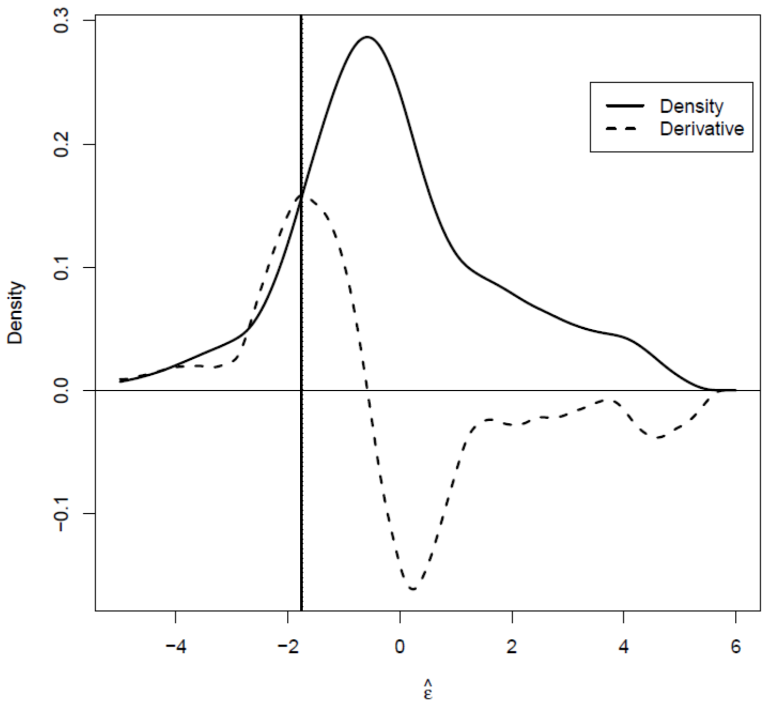

Figure 1: Estimated density of the CNLS residuals (solid line) and its first derivative (broken line). The vertical line indicates the maximum value of the derivative function, which provides

the estimate of the expected inefficiency.

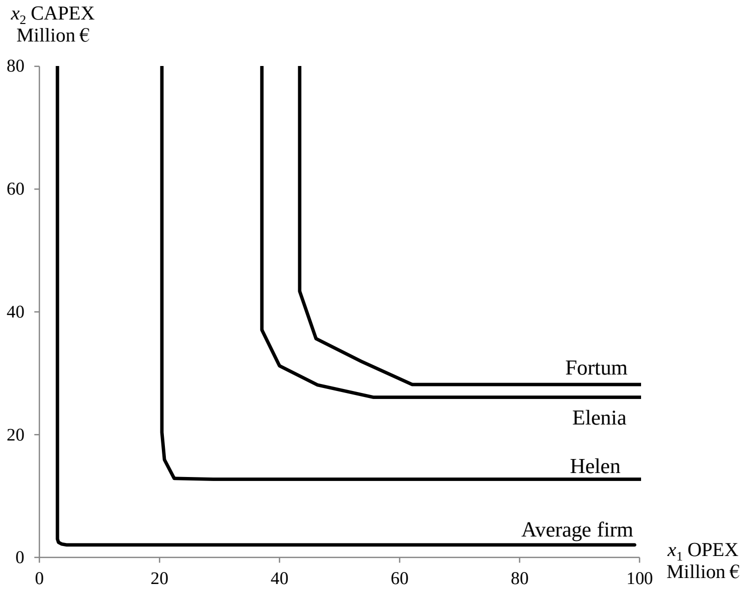

Figure 2: Estimated input isoquants of the average firm and the three largest distribution networks (Fortum, Elenia, and Helen).

Figure 2 plots the input isoquants of the representative distribution firm (i.e., the arithmetic average of the output vectors), and the three largest distribution firms (Fortum, Elenia, and Helen). The horizontal axis is OPEX and the vertical axis is CAPEX: both are measured in Million €. In the multiple output setting, the shape of the input isoquant depends on the output structure of the firm. The estimated isoquants suggest that there are considerable substitution possibilities between OPEX and CAPEX given the output structure of Fortum and Elenia, whereas the output structure of Helen (operating in Helsinki) yields almost a L-shaped (Leontief type) input isoquant.

To assess performance of distribution firms, we next project the observed data points to the frontier using the direction vector $\mathbf{g}^x = (1,0)$, $\mathbf{g}^y = (0,0,0)$, which is the relevant direction for the regulator. Keeping the outputs and CAPEX fixed at their current levels, we first estimate the cost saving potential for each firm as the difference of the observed OPEX and the predicted OPEX according to the StoNED frontier, and then sum over all firms to estimate the total cost-saving potential of the industry. The estimated cost saving potential in OPEX amounts to €44.6 Million, which is 14 percent of the observed total OPEX of the industry. In other words, the overall efficiency of the industry is estimated as 86 percent. Note that the measured distances from observed data points to the estimated frontier contain random noise, but since the noise term has zero mean, the firm-specific noise terms will likely cancel out as the distances of individual firms from the frontier are totaled to assess efficiency at the industry level. If we consider performance in terms of the ratio of estimated and observed OPEX at the firm level (keeping in mind our caveat regarding noise), we find a high dispersion of performance across firms, with the minimum ratio of estimated and observed OPEX equal to 45 percent, and the maximum equal to 140 percent. The arithmetic mean is 81 percent, which is lower than the cost-weighted overall efficiency of 86 percent noted above.

For comparison, we also applied the method of moments estimator assuming half-normal inefficiency and normal noise. This parametric estimator provides the estimated expected value of 2.45 for the inefficiency term, which is considerably higher than the value of 1.72 suggested by the kernel deconvolution method. To put these figures in a proper context, we calculated the total cost saving potential of the industry as explained above, but now using the method of moment estimate. This analysis suggests the total cost-saving potential of €75.5 Million, or 23 percent of the industry OPEX, which is €30.9 Million higher than the corresponding estimates obtained with the kernel deconvolution method. This comparison illustrates the decisive role of the distributional assumptions in regulation. However, sensitivity of the efficiency estimates on the distributional assumptions should not be interpreted as evidence in favor of the deterministic methods. To put the results in the correct perspective, if we assume away noise completely and interpreted all deviations from the frontier as inefficiency, as most regulators currently do, then the estimated saving potential in OPEX is €161.3 Million (or 49.8 percent). This is more than 3.6 times larger than our estimate. It may be also interesting to compare our estimates of the total industry efficiency with those obtained by Kuosmanen et al. (2013) using TOTEX as a single input factor (StoNED €47.5 Million, SFA €93.0 Million, DEA €141.4 Million): our estimate of €44.6 Million comes very close to the previously reported StoNED estimate.

Based on our study, the Finnish energy regulator applies the multiple output StoNED approach for the incentive regulation of electricity distribution firms during the period 2016 – 2023, including CNLS estimation of the conditional mean, kernel deconvolution of the expected inefficiency, and the use of the direction vector $\mathbf{g}^x = (1,0)$, $\mathbf{g}^y = (0,0,0)$, as discussed above. However, we stress again that the illustrative application presented in this section is not exactly identical to the full-scale benchmarking model used in the real world. The benchmarking model used in the real world includes the following important modifications: 1) The frontier is estimated using an unbalanced panel data rather than a single cross section considered in this section. 2) The CAPEX variable is replaced by the replacement value of the capital stock. 3) A hedonic measure of the damage due to interruptions is included in the model as an undesirable output, in addition to the input and output variables considered above. 4) The ratio of connection points to customers is used as a contextual variable in order to better capture observed heterogeneity in the operating environments of firms.

6. Conclusions

The stated objectives of this paper were twofold. Firstly, we introduced a stochastic DGP where all input and output variables are endogenous. To identify the directional distance function, viewed here as a representation of the benchmark technology, we demonstrated how to apply the translation property of the directional distance function to move one of the input or output variables as the dependent variable using a simple data transformation. We showed that the data transformation cancels out inefficiency and noise from the independent variables of the estimation under the stated assumptions. While some of the assumptions in the proposed DGP may appear restrictive, we find it critically important to make the underlying assumptions explicitly. The proposed DGP can be extended to more general settings, but we leave this as an interesting challenge for future research.

Secondly, we developed a new nonparametric axiomatic estimator of the directional distance function that introduces to the StoNED approach some recent developments in the literature of kernel deconvolution. The proposed StoNED approach expands the scope of the previous semi-nonparametric approaches in several dimensions. 1) The proposed approach is compatible with joint production of multiple outputs (returns to scope or economies of scope) using multiple inputs, 2) it has a sound axiomatic foundation, 3) it takes stochastic noise explicitly into account in all input and output variables, and 4) it does not rely on any arbitrary distributional assumptions. These are significant advances from the point of view of practical implementation of the StoNED approach.

We examined the specification of appropriate direction vector through an empirical case of energy regulation in Finland. While we view the choice of the direction vectors as an application specific issue, we believe this application can provide some ideas and insights that are potentially useful in other applications in different industries and at different levels of aggregation.

The present paper focuses on the production side of the economy. However, we believe the insights and techniques developed in this study could be useful in other contexts as well. One possibility is to utilize the intimate connections between the theory of revealed preference and the axiomatic production theory to apply the results and insights of this paper to consumer demand analysis and modeling of household consumption decisions (see, e.g., Cherchye et al., 2007). Technically, the directional distance function we considered is directly analogous to Luenberger's (1992) benefit function. Another possibility is to abstract from the microeconomic theory, and view the distance function simply as a functional representation of dependence between variables. From this perspective, the theoretical results of this paper could be utilized in multivariate regression where all variables are endogenous. We hope this paper open up several interesting avenues of future research.

References

Afriat, S.N. (1972) Efficiency estimation of production functions, *International Economic Review* 13(3): 568-598.

Aigner, D., C.A.K. Lovell, and P. Schmidt (1977) Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics* 6, 21-37.

Appa, G., and H.P. Williams (2006) A new framework for the solution of DEA models, *European Journal of Operational Research* 172(2), 604-615.

Atkinson, S.E., Tsionas, M.G.(2016). Directional distance functions: Optimal endogenous directions. Journal of Econometrics, 190 (2), pp. 301-314.

Banker, R.D. (1993) Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation, *Management Science* 39, 1265-1273.

Banker R.D., and A. Maindiratta (1992). Maximum likelihood estimation of monotone and concave production frontiers. *Journal of Productivity Analysis* 3: 401–415.

Bogetoft, P., and L. Otto (2011) *Benchmarking with DEA, SFA, and R*. International Series in Operations Research and Management Science 157, Springer.

Chambers, R.G., Y.H. Chung, and R. Färe (1996) Benefit and distance functions, *Journal of Economic Theory* 70(2), 407-419.

Chambers, R.G., Y.H. Chung, and R. Färe (1998) Profit, directional distance function, and Nerlovian efficiency, *Journal of Optimization Theory and Application* 98(2), 351-364.

Charnes, A., W.W. Cooper, and E. Rhodes (1978) Measuring the efficiency of decision making units, *European Journal of Operational Research* 2: 429-444.

Cherchye, L., B. De Rock and F. Vermeulen (2007) The collective model of household consumption: A nonparametric characterization, *Econometrica* 75, 553-574.

Coelli, T., and S. Perelman (1999) A comparison of parametric and non-parametric distance functions: With application to European railways, *European Journal of Operational Research* 117, 326-339.

Daraio, C., Simar, L. (2016). Efficiency and benchmarking with directional distances: A data-driven approach, Journal of the Operational Research Society, 67 (7), pp. 928-944.

Delaigle, A. and I. Gijbels (2004) Practical bandwidth selection in deconvolution kernel density estimation, *Computational Statistics and Data Analysis* 45, 249-267.

Diewert, W.E. and C. Parkan (1983) Linear programming tests of regularity conditions for production functions, in W. Eichorn, R. Henn, K. Neumann, and R. W. Shephard (eds.) *Quantitative Studies in Production and Prices*, Physica Verlag.

Fan, Y., Q. Li, and A. Weersink (1996) Semiparametric estimation of stochastic production frontier models, *Journal of Business and Economic Statistics* 4(4), 460-68.

Färe, R., S. Grosskopf, D.-W. Noh, and W. Weber (2005) Characteristics of a polluting technology: theory and practice, *Journal of Econometrics* 126(2), 469-492.

Färe, R., Grosskopf, S., Whittaker, G. (2013). Directional output distance functions: Endogenous directions based on exogenous normalization constraints. Journal of Productivity Analysis, 40 (3), pp. 267-269.

Färe, R., and D. Primont (1995) *Multi-output and duality: Theory and application*, Kluwer, Boston.

Farrell, M.J. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society Series A* 120: 253-281.

Fukuyama, H. (2003) Scale characterizations in a DEA directional technology distance function framework, *European Journal of Operational Research* 144(1), 108-127.

Goldenshluger, A., and A. Tsybakov (2004) Estimating the endpoint of a distribution in the presence of additive observation errors, *Statistics and Probability Letters* 68(1), 39-49.

Hanson, D.L. and G. Pledger (1976) Consistency in concave regression. *Annals of Statistics* 4(6): 1038-1050.

Hall P, and L. Simar (2002) Estimating a changepoint, boundary, or frontier in the presence of observation error, *Journal of the American Statistical Association* 97, 523–534.

Hildreth, C. (1954) Point estimates of ordinates of concave functions, *Journal of the American Statistical Association* 49: 598-619.

Horrace, W.C., and C.F. Parmeter (2011) Semiparametric deconvolution with unknown error variance, *Journal of Productivity Analysis* 35, 129–141

Jamasb, T., and M. Pollitt (2007) Incentive regulation of electricity distribution networks: Lessons of experience from Britain, *Energy Policy* 35(12), 6163–6187.

Koopmans, T.C. (1951) An analysis of production as an efficient combination of activities, pp. 33-97 in Koopmans, T.C. (Ed.) *Activity Analysis of Production and Allocation, Proceeding of a Conference,* John Wiley and Sons Inc., London.

Kumbhakar, S.C. (2013) Specification and estimation of multiple output technologies: A primal approach, *European Journal of Operational Research* 231, 465-473.

Kuosmanen, T. (2006) *Stochastic nonparametric envelopment of data: Combining virtues of SFA and DEA in a unified framework*, MTT Discussion Paper No. 3/2006.

Kuosmanen, T. (2008) Representation theorem for convex nonparametric least squares. *Econometrics Journal* 11, 308-325.

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., A. Saastamoinen, and T. Sipiläinen (2013) What is the best practice for benchmark regulation of electricity distribution? Comparison of DEA, SFA and StoNED methods, *Energy Policy* 61, 740-750.

Kuosmanen, T. and A.L. Johnson (2010) Data envelopment analysis as nonparametric least squares regression, *Operations Research* 58(1), 149-160.

Kuosmanen, T., A.L. Johnson, and A. Saastamoinen (2015) Stochastic nonparametric approach to efficiency analysis: A unified framework, in J. Zhu (Ed.) *Data Envelopment Analysis*, Springer.

Kuosmanen, T., and M. Kortelainen (2012) Stochastic non-smooth envelopment of data: Semi-parametric frontier estimation subject to shape constraints, *Journal of Productivity Analysis* 38(1), 11-28.

Kuosmanen, T., G.T. Post, and S. Scholtes (2007) Testing for productive efficiency in case of errors-in-variables, *Journal of Econometrics* 136, 131-162.

Lim, E., and P.W. Glynn (2012) Consistency of multi-dimensional convex regression, *Operations Research* 60, 196-208.

Lovell, C.A.K., P. Travers, S. Richardson, and L.L. Wood (1994) Resources and functioning: A new view of inequality in Australia. Models and measurement of welfare and inequality. W. Eichorn, Springer Verlag.

Luenberger, D.G. (1992) Benefit functions and duality, *Journal of Mathematical Economics* 21(5), 461-481.

Meeusen, W. and J. Vandenbroeck (1977) Efficiency estimation from Cobb-Douglas production functions with composed error, *International Economic Review* 18(2), 435-445.

Olesen, O.B., and N.C. Petersen (2003) Identification and use of efficient faces and facets in DEA, *Journal of Productivity Analysis* 20(3), 323-360.

Olesen, O.B., and N.C. Petersen (2016) Stochastic Data Envelopment Analysis: A review, *European Journal of Operational Research* 251(1), 2-21.

Olley, S., and A. Pakes (1996) The dynamics of productivity in the telecommunications equipment industry, *Econometrica* 64, 1263–1298.

Panzar, J.C., and R.D. Willig (1981) Economies of scope, *American Economic Review* 71(2), 268-272.

Perelman, S., and D. Santin (2009) How to generate regularly behaved production data? A Monte Carlo experimentation on DEA scale efficiency measurement, *European Journal of Operational Research* 199, 303–310.

Pope, B. and A.L. Johnson (2013) Scope properties: Nonparametric assessment, policy insights and functional estimation, *Journal of Productivity Analysis* 40(2), 239-250.

Seijo, E., and B. Sen (2011) Nonparametric least squares estimation of a multivariate convex regression function, *Annals of Statistics* 39(3), 1633-1657.

Shephard, R.W. (1953) *Cost and production functions*, Princeton University Press, Princeton, N.J..

Shephard, R.W. (1970) *Theory of cost and production functions*, Princeton University Press, Princeton, N.J.

Simar, L. (2007) How to improve the performances of DEA/FDH estimators in the presence of noise? *Journal of Productivity Analysis* 28(3), 183-201.

Simar, L., and V. Zelenyuk (2011) Stochastic FDH/DEA estimators for frontier analysis, *Journal of Productivity Analysis* 36(1), 1-20.

Varian, H.R. (1984). The nonparametric approach to production analysis. *Econometrica* 52, 579-598.

Varian, H.R. (1985) Nonparametric analysis of optimizing behavior with measurement error, *Journal of Econometrics* 30(1/2), 445-458.

Waldman, D. (1982) A stationary point for the stochastic frontier likelihood, *Journal of Econometrics* 18(2), 275-279.

Yagi, D., A.L. Johnson, and T. Kuosmanen (2016) Shape constrained kernel-weighted least squares: Application to production function estimation for Chilean manufacturing industries, arxiv working paper available at: https://arxiv.org/abs/1604.06003.

Zofio J.L., Pastor J.T., Aparicio J. (2013). The directional profit efficiency measure: on why profit efficiency is either technical or allocative. *Journal of Productivity Analysis* 40(3), 257–266.

Online supplement

Appendix 1: Proofs of Propositions

Proposition 1: *If inputs and outputs are freely disposable (A1), then the directional distance function* $\vec{D}_T$ *is monotonically decreasing in* $(\mathbf{g}^x, \mathbf{g}^y)$ *for all* $(\mathbf{x}, \mathbf{y}) \in T$ *and monotonically increasing in* $(\mathbf{g}^x, \mathbf{g}^y)$ *for all* $(\mathbf{x}, \mathbf{y}) \notin T$.

Proof.

Consider first an arbitrary $(\mathbf{x}, \mathbf{y}) \in T$ and a pair of direction vectors $(\mathbf{g}_A^x, \mathbf{g}_A^y) \in \Re_+^{m+s}$ and $(\mathbf{g}_B^x, \mathbf{g}_B^y) \in \Re_+^{m+s}$ such that $(\mathbf{g}_A^x, \mathbf{g}_A^y), (\mathbf{g}_B^x, \mathbf{g}_B^y) \neq (\mathbf{0}, \mathbf{0})$. We need to show that

$$(\mathbf{g}_A^x, \mathbf{g}_A^y) \geq (\mathbf{g}_B^x, \mathbf{g}_B^y) \Rightarrow \vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}_A^x, \mathbf{g}_A^y) \leq \vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}_B^x, \mathbf{g}_B^y).$$

Define the projection points $(\mathbf{x}_A^*, \mathbf{y}_A^*)$ and $(\mathbf{x}_B^*, \mathbf{y}_B^*)$ on the boundary of T as follows:

$$(\mathbf{x}_A^*, \mathbf{y}_A^*) . \vec{D}_T(\mathbf{x}_A^*, \mathbf{y}_A^*, \mathbf{g}_A^x, \mathbf{g}_A^y) = 0$$
$$\mathbf{x}_A^* = \mathbf{x} - \vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}_A^x, \mathbf{g}_A^y)\mathbf{g}_A^x$$
$$\mathbf{y}_A^* = \mathbf{y} + \vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}_A^x, \mathbf{g}_A^y)\mathbf{g}_A^y$$

$$(\mathbf{x}_B^*, \mathbf{y}_B^*) . \vec{D}_T(\mathbf{x}_B^*, \mathbf{y}_B^*, \mathbf{g}_B^x, \mathbf{g}_B^y) = 0$$
$$\mathbf{x}_B^* = \mathbf{x} - \vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}_B^x, \mathbf{g}_B^y)\mathbf{g}_B^x$$
$$\mathbf{y}_B^* = \mathbf{y} + \vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}_B^x, \mathbf{g}_B^y)\mathbf{g}_B^y$$

Next, define a point $(\mathbf{x}_{AB}^*, \mathbf{y}_{AB}^*)$ as

$$\mathbf{x}_{AB}^* = \mathbf{x} - \vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}_A^x, \mathbf{g}_A^y)\mathbf{g}_B^x$$
$$\mathbf{y}_{AB}^* = \mathbf{y} + \vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}_A^x, \mathbf{g}_A^y)\mathbf{g}_B^y$$

Note that point $(\mathbf{x}_{AB}^*, \mathbf{y}_{AB}^*)$ is obtained by projecting (**x**,**y**) in the direction $(\mathbf{g}_B^x, \mathbf{g}_B^y)$ by amount $\vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}_A^x, \mathbf{g}_A^y)$, and it is not necessarily on the frontier. Since $(\mathbf{g}_A^x, \mathbf{g}_A^y) \geq (\mathbf{g}_B^x, \mathbf{g}_B^y)$, point $(\mathbf{x}_A^*, \mathbf{y}_A^*)$ obviously dominates $(\mathbf{x}_{AB}^*, \mathbf{y}_{AB}^*)$ in the sense that

$$\mathbf{x}_A^* \leq \mathbf{x}_{AB}^*$$
$$\mathbf{y}_A^* \geq \mathbf{y}_{AB}^*$$

Monotonicity of $\vec{D}_T$ in (**x**,**y**) requires that $(\mathbf{x}_B^*, \mathbf{y}_B^*)$ must also dominate $(\mathbf{x}_{AB}^*, \mathbf{y}_{AB}^*)$, that is,

$$\mathbf{x}_B^* \leq \mathbf{x}_{AB}^*$$
$$\mathbf{y}_B^* \geq \mathbf{y}_{AB}^*$$

Otherwise the boundary point $(\mathbf{x}_A^*, \mathbf{y}_A^*)$ dominates another boundary point $(\mathbf{x}_B^*, \mathbf{y}_B^*)$, which clearly violates the free disposability axiom A1.

Now, consider the difference of $(\mathbf{x}_B^*, \mathbf{y}_B^*)$ and $(\mathbf{x}_{AB}^*, \mathbf{y}_{AB}^*)$, which can be stated as

$$\mathbf{x}_B^* - \mathbf{x}_{AB}^* = (\vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}_A^x, \mathbf{g}_A^y) - \vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}_B^x, \mathbf{g}_B^y))\mathbf{g}_B^x \leq \mathbf{0}$$

$$\mathbf{y}_B^* - \mathbf{y}_{AB}^* = -(\vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}_A^x, \mathbf{g}_A^y) - \vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}_B^x, \mathbf{g}_B^y))\mathbf{g}_B^y \geq \mathbf{0}$$

Since $(\mathbf{g}_B^x, \mathbf{g}_B^y) \geq (\mathbf{0}, \mathbf{0})$ by assumption, we must have

$$\vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}_A^x, \mathbf{g}_A^y) \leq \vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}_B^x, \mathbf{g}_B^y).$$

Thus, we have shown that $\vec{D}_T$ is monotonically decreasing in $(\mathbf{g}^x, \mathbf{g}^y)$ for all $(\mathbf{x}, \mathbf{y}) \in T$.

Directly analogous argument applies to an arbitrary $(\mathbf{x}, \mathbf{y}) \notin T$. The only difference concerns the signs of inequalities $\mathbf{x}_B^* - \mathbf{x}_{AB}^* \leq \mathbf{0}, \mathbf{y}_B^* - \mathbf{y}_{AB}^* \geq \mathbf{0}$. The signs are reversed because

$$\vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}^x, \mathbf{g}^y) \leq 0 \ \forall (\mathbf{x}, \mathbf{y}) \notin T.$$

While the absolute value $\left|\vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}^x, \mathbf{g}^y)\right|$ is monotonically decreasing in $(\mathbf{g}^x, \mathbf{g}^y)$ for all $(\mathbf{x}, \mathbf{y}) \in \Re_+^{m+s}$, function $\vec{D}_T(\mathbf{x}, \mathbf{y}, \mathbf{g}^x, \mathbf{g}^y)$ is monotonically increasing in $(\mathbf{g}^x, \mathbf{g}^y)$ for all $(\mathbf{x}, \mathbf{y}) \notin T$. □

Proposition 2: *If the observed data are generated according to the DGP described in Section 2.3, then the value of the DDF in observed data* $(\mathbf{x}_i, \mathbf{y}_i)$ *point is equal to the realization of the random variable* ε_i*, specifically,*

$$\vec{D}_T(\mathbf{x}_i, \mathbf{y}_i, \mathbf{g}^x, \mathbf{g}^y) = \varepsilon_i \quad \forall i.$$

Proof.

If the observed data are generated according to the DGP stated in Section 2.3, then the value of the DDF with the observed data of firm *i* is given by

$$\vec{D}_T(\mathbf{x}_i, \mathbf{y}_i, \mathbf{g}^x, \mathbf{g}^y) = \vec{D}_T(\mathbf{x}_i^* + \varepsilon_i \mathbf{g}^x, \mathbf{y}_i^* - \varepsilon_i \mathbf{g}^y, \mathbf{g}^x, \mathbf{g}^y).$$

Using the translation property A4, we have

$$\vec{D}_T(\mathbf{x}_i^* + \varepsilon_i \mathbf{g}^x, \mathbf{y}_i^* - \varepsilon_i \mathbf{g}^y, \mathbf{g}^x, \mathbf{g}^y) = \vec{D}_T(\mathbf{x}_i^*, \mathbf{y}_i^*, \mathbf{g}^x, \mathbf{g}^y) + \varepsilon_i.$$

By definition, the efficient input-output vectors must be on the boundary of the production possibility set *T* and satisfy the condition

$$\vec{D}_T(\mathbf{x}_i^*, \mathbf{y}_i^*, \mathbf{g}^x, \mathbf{g}^y) = 0.$$

Combining the previous steps, we have shown that

$$\overrightarrow{D}_T(\mathbf{x}_i, \mathbf{y}_i, \mathbf{g}^x, \mathbf{g}^y) = \varepsilon_i \quad \forall i \text{ .}$$ □

Proposition 3: *If the observed data are generated by the DGP described in Section 2.3, then the transformed input-output variables* $(\vec{\mathbf{x}}_i, \vec{\mathbf{y}}_i)$ *are uncorrelated with the error term* ε_i *, that is,*

$$Cov(\varepsilon_i, \vec{\mathbf{x}}_i) = \mathbf{0} \ \forall i \quad \text{and} \quad Cov(\varepsilon_i, \vec{\mathbf{y}}_i) = \mathbf{0} \ \forall i \text{ .}$$

Proof.

Properties of the DGP imply that the transformed input vector can be stated as

$$\begin{aligned}\vec{\mathbf{x}}_i &= \mathbf{x}_i + (y_{1i} / g_1^y)\mathbf{g}^x = (\mathbf{x}_i^* + \varepsilon_i \mathbf{g}^x) + ((y_{1i}^* - \varepsilon_i g_1^y) / g_1^y)\mathbf{g}^x \\ &= (\mathbf{x}_i^* + (y_{1i}^* / g_1^y)\mathbf{g}^x) + \varepsilon_i(\mathbf{g}^x - \mathbf{g}^x) \\ &= \mathbf{x}_i^* + (y_{1i}^* / g_1^y)\mathbf{g}^x\end{aligned}$$

Note that the composite error ε_i present in both the elements of $\mathbf{x}_i$ and y_{1i} cancels out completely. Since the optimal solutions $(\mathbf{x}_i^*, \mathbf{y}_i^*)$ are exogenously given constants, we have

$$Cov(\varepsilon_i, \mathbf{x}_i^*) = \mathbf{0},$$

$$Cov(\varepsilon_i, \mathbf{y}_i^*) = \mathbf{0}$$

Further, since the direction vector $(\mathbf{g}^x, \mathbf{g}^y)$ is constant across observations, we obviously have

$$Cov(\varepsilon_i, \mathbf{g}^x) = \mathbf{0},$$

$$Cov(\varepsilon_i, \mathbf{g}^y) = \mathbf{0}$$

Combining these observations, we have demonstrated that

$$Cov(\varepsilon_i, \vec{\mathbf{x}}_i) = Cov(\varepsilon_i, \mathbf{x}_i^* + (y_{1i}^* / g_1^y)\mathbf{g}^x) = \mathbf{0} \ \forall i,$$

which confirms the first part of the proposition.

Consider next the output vector. Using exactly the same arguments as in the case of the input vector, we have

$$\begin{aligned}\vec{\mathbf{y}}_i &= \mathbf{y}_i - (y_{1i} / g_1^y)\mathbf{g}^y \\ &= (\mathbf{y}_i^* - \varepsilon_i \mathbf{g}^y) - ((y_{1i}^* - \varepsilon_i g_1^y) / g_1^y)\mathbf{g}^y \\ &= (\mathbf{y}_i^* - (y_{1i}^* / g_1^y)\mathbf{g}_i^y) - \varepsilon_i(\mathbf{g}^y - \mathbf{g}^y) \\ &= \mathbf{y}_i^* - (y_{1i}^* / g_1^y)\mathbf{g}_i^y\end{aligned}$$

As in the case of the input vector, the composite error term ε_i present in the elements of both $\mathbf{y}_i$ and in y_{1i} cancels out completely. We have already note that $Cov(\varepsilon_i, \mathbf{y}_i^*) = \mathbf{0}$ and $Cov(\varepsilon_i, \mathbf{g}^y) = \mathbf{0}$, which imply that the second part of the proposition must hold:

$$Cov(\varepsilon_i, \vec{\mathbf{y}}_i) = Cov(\varepsilon_i, \mathbf{y}_i^* - (y_{1i}^* / g_1^y)\mathbf{g}_i^y) = \mathbf{0} \ \forall i$$ □

Appendix 2: Additional materials related to the application

Normalizing an input variable as the dependent variable, the regression equation becomes

$$x_{1i} / g_1^x = -\vec{D}_T(\mathbf{x}_i - (x_{1i} / g_1^x)\mathbf{g}^x, \mathbf{y}_i + (x_{1i} / g_1^x)\mathbf{g}^y, \mathbf{g}^x, \mathbf{g}^y) + \varepsilon_i \,.$$

Note that the signs of the DDF and the composite error term change compared to the situation where an output variable is normalized as the dependent variable. Function $-\vec{D}_T(\cdot)$ is monotonically decreasing in inputs, monotonically increasing in outputs, and convex in inputs and outputs.

The CNLS formulation presented in the paper does not make any assumption about the returns to scale (i.e., variable returns to scale are implicitly assumed). The previous study by Kuosmanen (2012), using this same data with TOTEX as the single input variable, argues that the CRS axiom A3 is justified from the regulation point of view. Further, the null hypothesis of CRS could not be rejected in empirical specification tests. In this study we impose CRS by dividing all input and output variables by output y_1. In practice, this implies that distance to frontier is measured in €/GWh, the inputs are expressed as €/GWh, and the outputs are the line km per GWh and customers per GWh. This specification also helps to alleviate heteroscedasticity due to the large differences in firm size.

Applying the modifications discussed above, the CNLS problem becomes

$$\min_{\alpha,\beta_2,\gamma_2,\gamma_3,\varepsilon^\circ} \sum_{i=1}^{n} (\varepsilon_i^\circ)^2$$

$$x_{1i} / y_{1i} = \alpha_i + \gamma_{2i}(y_{2i} / y_{1i}) + \gamma_{3i}(y_{3i} / y_{1i}) - \beta_{2i}(x_{2i} / y_{1i} - g_2^x(x_{1i} / y_{1i})) + \varepsilon_i^\circ \;\; \forall i$$

subject to

$$\alpha_i + \gamma_{2i}(y_{2i} / y_{1i}) + \gamma_{3i}(y_{3i} / y_{1i}) - \beta_{2i}(x_{2i} / y_{1i} - g_2^x(x_{1i} / y_{1i}))$$
$$\geq \alpha_h + \gamma_{2h}(y_{2i} / y_{1i}) + \gamma_{3h}(y_{3i} / y_{1i}) - \beta_{2h}(x_{2i} / y_{1i} - g_2^x(x_{1i} / y_{1i})) \;\; \forall i, h$$
$$g_2^x \beta_{2i} \leq 1$$
$$\alpha_i \geq 0, \beta_{2i} \geq 0, \gamma_{2i} \geq 0, \gamma_{3i} \geq 0 \;\; \forall i$$

The subscripts of the gamma and beta coefficients refer to the shadow prices of the corresponding output and input variables (i.e., y_2, y_3, x_2). Coefficient α_i is the shadow price of the transmission output (y_1). The shadow price of OPEX (x_1) is $(1 - g_2^x \beta_{2i})$.

We apply the CNLS regression using a grid of initial values for g_2^x, say $\mathbf{g}_0^x = (1,0)$, $\mathbf{g}_1^x = (1,0.1)$, $\mathbf{g}_1^x = (1,0.2)$, and so forth. We then apply the White test of heteroscedasticity to the CNLS residuals using the transformed variables $(y_{2i} / y_{1i}, y_{3i} / y_{1i}, x_{2i} / y_{1i} - g_2^x(x_{1i} / y_{1i}))$, their squared values, and their cross products as regressors. Based on the estimates of the White test statistic (nR^2), we can introduce a finer grid of values for g_2^x, and compute the estimator and the associated White test statistics. Our main aim is to identify at least one direction for which the null hypothesis of homoscedasticity can be maintained according to the White test. Finding the "optimal" direction that minimizes the White test statistic is of secondary importance, but it can be used as an empirical criterion for the specification of the direction vector.

Figure A plots the values of the White test statistic as the value of the direction g_2^x increases from zero to one. We obtain the lowest value of the test statistic by setting $g_2^x = 0.25$. We first used grid search with increments of 0.1, and then performed a more detailed search with the interval [0.2, 0.3], estimating the model using 21 different direction vectors in total. We find that all test statistic values associated with $g_2^x < 0.7$ are below the critical value of the White test at the conventional levels of significance (the critical values of the White test at 1%, 5%, and 10% significance levels are indicated in Figure A). Thus, we will set the direction vector as $\mathbf{g}^x = (1, 0.25)$.

The search of the most homoscedastic direction also revealed that the skewness of CNLS residuals is positive at all values of the direction vector considered, as expected.

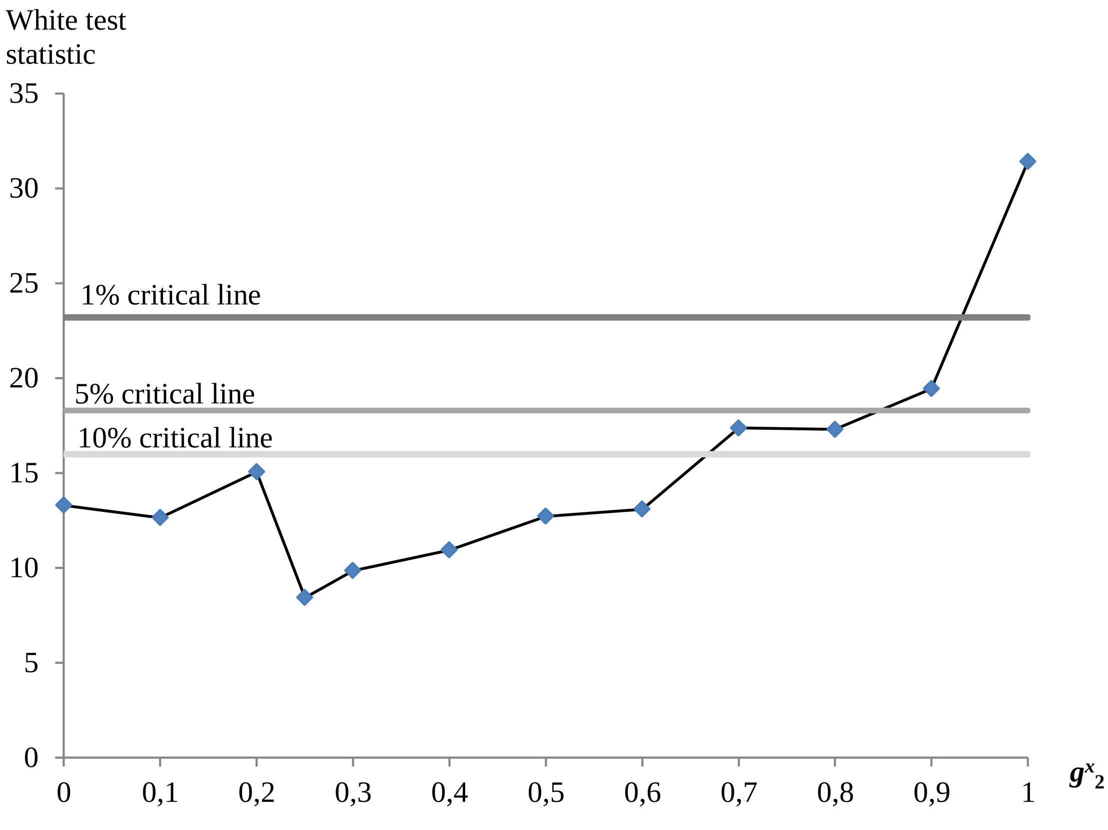

Figure A: Values of the White test statistic as the function of g_2^x.