GRIPS Policy Information Center Research Report : I-2002-0004 Rev.

# Nonparametric and Parametric Measures of Scale Elasticity: A Comparative Evaluation

By

**Kaoru Tone**[a], **Biresh K. Sahoo**[b] and **Miki Tsutsui**[a,c]

[a] *National Graduate Institute for Policy Studies*
*2-2 Wakamatsu-cho, Shinjuku-ku, Tokyo 162-8677, Japan*
E-mail: tone@grips.ac.jp

[b] *Amrita School of Business, Amrita Vishwa Vidyapeetham*
*Ettimadai Post, Coimbatore 641 105, India*
E-mail: s_biresh@ettimadai.amrita.edu

[c] *Central Research Institute of Electric Power Industry*
*1-6-1 Ohtemachi, Chiyoda-ku, Tokyo 100-8126, Japan*
E-mail: miki@criepi.denken.or.jp

First Draft: January 17, 2005

Research supported by Grant-in-Aid for Scientific Research (C)
Japan Society for the Promotion of Science

# Nonparametric and Parametric Measures of Scale Elasticity: A Comparative Evaluation

Kaoru Tone,* Biresh K. Sahoo † and Miki Tsutsui‡

**Abstract**

In any competitive set up, policy recommendations based on elasticity parameters have assumed greater significance for the firm's financial viability and success. So there is a need to examine with more prudence the elasticity estimates obtained from various parametric as well nonparametric methods. The main aim of this paper is first to critically examine these methods in order to shed light on what seems to be missing, and then to proceed by developing an empirically demanding ecompassing measure of scale elasticity. All these measures, which are finally applied to a panel data of US electric companies, constitute the empirical premise of this paper.
**Keywords**: DEA; Translog cost function; Cost efficiency; Scale elasticity.

## 1 Introduction

Intensifying pressures in competitive environment have motivated many industries to build larger operating units to achieve the widely advantages of 'scale economies.' This is apparent not only in manufacturing industries but also in regulated/state-owned industries such as electricity, water, telecom, etc., and public sector units such as hospitals and schools. This reflects the spread of faith in the underlying benefits of 'scale increases' in the minds of

*Corresponding author. tone@grips.ac.jp
†s_biresh@ettimadai.amrita.edu
‡miki@criepi.denken.or.jp

economists, engineers, industrial managers, and governments. From a policy point of view, the estimation of scale elasticity (returns to scale, RTS) parameter is of particular importance concerning whether there is any scope for increased productivity by expanding/contracting, whether minimum efficient scales will allow competitive markets to be established, and if the existing size distribution of firms is consistent with a competitive market outcome. So there is a need that these estimates should be examined with more prudence for the firm's financial viability and success.

We find in the economics literature (Färe *et al.*, 1988) that there are two approaches to the estimation of scale elasticity: the neoclassical approach and axiomatic approach. The former approach (whose estimation method is parametric econometric approach) gives us quantitative measure of scale economies whereas the latter approach (whose estimation method is the non-parametric data envelopment analysis (DEA) by Charnes *et al.*, 1978) yields qualitative information on scale economies. Recently, we find in the literature that DEA models also generate quantitative information of scale economies (Banker *et al.*, 1996b, Førsund, 1996, Sueyoshi, 1997). Both the methods have become important analytical tools in the empirical evaluation of scale elasticity. The main purpose of this paper is to empirically examine the nature of scale properties in both the methods, then to point out their relative strengths and weaknesses, and finally to propose an alternative based on the premise that other estimates can be illusory.

The choice of generation division of the U.S. electric industry for the empirical comparison of scale elasticity estimates between these two methods undertaken in this study, is not only made for its importance but also designed to be illustrative of the many potential applications elsewhere in other divisions too. The existence of monopoly structure in this industry had long been debated just after the appearance of Christensen and Greene (1976)'s study, which showed that economies of scale at the generation division had been exhausted in several U.S. electric power companies. However, in comparison to the generation division, it had also been presumed that economies of scale had still remained in the network activities. These observations prompted policy makers bringing deregulation into this industry, introducing competition into the generation division, and unbundling of vertical integrated structure of the electric power industry.[1]. Much of earlier

[1]Even knowing the fact that economies of scale are neither necessary nor sufficient

literature have too examined the existence of not only the economies of scale in generation division, but also economies of vertical integration.

This paper unfolds as follows. In Section 2, we first introduce the non-parametric DEA models for the qualitative evaluation of RTS, discuss the quantitative evaluation of elasticity, then point out their limitations and suggest an alternative method in the same existing DEA framework. Section 3 deals with discussion of quantitative evaluations of elasticity in the recent parametric models. For illustrative empirical comparison, we employ the panel data of 18 US Electric Companies over a period of eight years to demonstrate the strengths and weaknesses of each model in Section 4. Section 5 offers some concluding remarks.

# 2 Nonparametric DEA Models

Throughout this paper, we deal with $n$ decision making units (DMUs)/firms, each uses $m$ inputs to produce $s$ outputs. For each DMU$_o$ ($o = 1, \ldots, n$), we denote respectively the input/output vectors by $\boldsymbol{x}_o \in R^m$ and $\boldsymbol{y}_o \in R^s$. The input/output matrices are defined by $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in R^{m \times n}$ and $Y = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) \in R^{s \times n}$. We assume that $X > O$ and $Y > O$.

## 2.1 Technology and Scale Elasticity

The technology ($T$), which converts inputs into outputs at any given point of time is defined as the set of all feasible input-output combinations,

$$\boldsymbol{T} \equiv \{(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x} \text{ can produce } \boldsymbol{y}\}.$$

The standard neoclassical characterization of production function for multiple outputs and multiple inputs is the transformation function $\psi(\boldsymbol{x}, \boldsymbol{y})$, which exhibits the following properties:

$$\psi(\boldsymbol{x}, \boldsymbol{y}) = 0, \ \frac{\partial \psi(\boldsymbol{x}, \boldsymbol{y})}{\partial y_r} < 0 \ (\forall r) \text{ and } \frac{\partial \psi(\boldsymbol{x}, \boldsymbol{y})}{\partial x_i} > 0 \ (\forall i).$$

Alternatively, the technology can be described by its input set

$$L(\boldsymbol{y}) \equiv \{\boldsymbol{x} \mid (\boldsymbol{x}, \boldsymbol{y}) \in T\} \text{ for all } \boldsymbol{y},$$

condition for the existence of natural monopoly or by its output set

$$P(\boldsymbol{x}) \equiv \{\boldsymbol{y} \mid (\boldsymbol{x}, \boldsymbol{y}) \in T\} \text{ for all } \boldsymbol{x}.$$

Following Shephard (1970), the output distance function is defined as

$$D_o(\boldsymbol{x}, \boldsymbol{y}) \equiv \inf\{\delta \mid \boldsymbol{y}/\delta \in P(\boldsymbol{x}), \delta > 0\}.$$

For any output vector $\boldsymbol{y}$, $\boldsymbol{y}/D_0(\boldsymbol{x}, \boldsymbol{y})$ is the largest output quantity vector on the ray from the origin through $\boldsymbol{y}$ that can be produced from $\boldsymbol{x}$. Assuming free disposability, the following holds true:

$$\boldsymbol{y} \in P(\boldsymbol{x}) \text{ if and only if } D_0(\boldsymbol{x}, \boldsymbol{y}) \leq 1.$$

Thus, $D_0(\boldsymbol{x}, \boldsymbol{y})$ provides a representation of the technology.

The returns to scale (RTS) or scale elasticity in production ($\rho_p$) or degree of scale economies (DSE) or *Passus Coefficient*, is defined as the ratio of the maximum proportional ($\beta$) expansion of outputs to a given proportional ($\mu$) expansion of inputs. So differentiating the transformation function $\psi(\mu\boldsymbol{x}, \beta\boldsymbol{y}) = 0$ w.r.t. scaling factor $\mu$, and then equating it with zero yields the following local scale elasticity measure[2]:

$$\rho_p(\boldsymbol{x}, \boldsymbol{y}) \equiv -\sum_i^m x_i \frac{\partial \psi}{\partial x_i} \Big/ \sum_r^s y_r \frac{\partial \psi}{\partial y_r}.$$

However, in case of single input and single output technology, $\rho_p$ is simply expressed as the ratio of marginal product (MP)[$=dy/dx$] to average product (AP) [$=y/x$], i.e.,

$$\rho_p(x, y) \equiv \frac{\text{MP}}{\text{AP}} = \frac{dy/dx}{y/x}.$$

The scale elasticity also reflects the sensitivity of the output distance function with respect to changes in the input quantity vector where $\psi(\boldsymbol{x}, \boldsymbol{y}) = D_o(\boldsymbol{x}, \boldsymbol{y}) - 1 = 0$ (Färe *et al.*, 1986, and Ray, 1999). Assuming $D_o(\boldsymbol{x}, \boldsymbol{y})$ to be continuously differentiable, $\rho_p$ is then defined by

$$\rho_p(\boldsymbol{x}, \boldsymbol{y}) \equiv -\frac{\sum_{i=1}^m x_i \frac{\partial D_o(\boldsymbol{x}, \boldsymbol{y})}{\partial x_i}}{D_o(\boldsymbol{x}, \boldsymbol{y})}.$$

[2]See Hanoch (1970), Starrett (1977), Panzar and Willig (1977) and Baumol *et al.* (1982) for the detailed discussion.

For a neoclassical 'S-shaped production function' (or *Regular Ultra Passum Law* (RUPL) in the words of Frisch, 1965), $\rho_p(x, y)$ takes on values ranging from 'greater than one' for sub-optimal output levels, through 'one' at the optimal scale level, and to values 'less than one' at the super-optimal output levels. So the production function satisfies RUPL if $\partial\rho_p/\partial y < 0$ and $\partial\rho_p/\partial x < 0$ (Førsund and Hjalmarsson, 2004). RTS are increasing, constant and decreasing if $\rho_p > 1$, $\rho_p = 1$, and $\rho_p < 1$ respectively.

Following Baumol *et al.* (1988), the dual measure of scale elasticity[3] in terms of cost ($\rho_c$), is defined in multiple input and multiple output environment as

$$\rho_c \equiv C(\boldsymbol{w}, \boldsymbol{y}) \Big/ \sum_{r=1}^{s} y_r \frac{\partial C(\boldsymbol{w}, \boldsymbol{y})}{\partial y_r},$$

where $C(\boldsymbol{w}, \boldsymbol{y}) \equiv \min_{\boldsymbol{x}}\{\boldsymbol{w}.\boldsymbol{x} \mid \boldsymbol{x} \in L(\boldsymbol{y})\}$ is the minimum cost of producing output vector $\boldsymbol{y}$ when input price vector is $\boldsymbol{w}$. However, $\rho_c$ can be expressed as the ratio of average cost to marginal cost in the case of single output. RTS are increasing, constant or decreasing depending upon whether $\rho_c > 1$, $\rho_c = 1$, or $\rho_c < 1$ respectively.

## 2.2 Qualitative Information on RTS

The CCR output oriented model (Charnes *et al.*, 1978), which is based on the assumption of constant returns to scale, is used to qualitatively describe local RTS for $\text{DMU}_o$.

$$\begin{aligned} &[\text{CCR-O}] && \max\ \theta \\ &\text{subject to} && \sum_{j=1}^{n} x_{ij}\lambda_j \le x_o\ (i = 1, \ldots, m) \\ &&& -\sum_{j=1}^{n} y_{rj}\lambda_j + \theta y_o \le 0\ (r = 1, \ldots, s) \\ &&& \lambda_j \ge 0.\ (\forall j) \end{aligned}$$

If $\sum_{j=1}^{n} \lambda_j = 1$ in any alternate optima, then constant returns to scale (CRS) prevails on $\text{DMU}_o$; if $\sum_{j=1}^{n} \lambda_j < 1$ for all alternate optima, then increasing

[3] In the production economics literature, the reciprocal of scale elasticity is described as cost elasticity (Chambers, 1988 and Varian, 1992). It is also sometimes referred to as an index of *cost flexibility*, which serves to explain the mark-up in a 'quasi-competitive' benchmark setting (Baumol *et al.*, 1988).

returns to scale (IRS) prevails; and if $\sum_{j=1}^{n} \lambda_j > 1$ for all alternate optima, then decreasing returns to scale (DRS) prevails.

The dual of the BCC model (Banker *et al.*, 1984), which is based on the assumption of variable returns to scale (VRS), is also used for obtaining the qualitative information on local RTS for $\text{DMU}_o$.

$$\begin{aligned} &[\text{BCC-O}] \quad && \min\ \phi = \sum_{i=1}^{m} v_i x_{io} + v_o \\ &\text{subject to} && -\sum_{r=1}^{s} u_r y_{rj} + \sum_{i=1}^{m} v_i x_{ij} + v_o \geq 0, (j = 1, \ldots, n) \\ & && \sum_{r=1}^{s} u_r y_{ro} = 1 \\ & && u_r, v_i \geq 0,\ \text{and } v_o:\ \text{free.} \end{aligned}$$

If $v_o^* = 0$ ($*$ represents optimal value) in any alternate optimal then CRS prevails on $\text{DMU}_o$, if $v_o^* < 0$ in all alternate optimal then IRS prevails, and if $v_o^* > 0$ in all alternate optimal then DRS prevails on $\text{DMU}_o$.

Färe *et al.* (1985) introduced the following 'scale efficiency index' (SEI) method, which is based on non-increasing returns to scale (NIRS), to determine the nature of local RTS for $\text{DMU}_o$ as follows:

$$\begin{aligned} &[\text{SEI-O}] \quad && \max\ f \\ &\text{subject to} && \sum_{j=1}^{n} x_{ij}\lambda_j \leq x_o\ (i = 1, \ldots, m) \\ & && -\sum_{j=1}^{n} y_{rj}\lambda_j + f y_o \leq 0\ (r = 1, \ldots, s) \\ & && \sum_{j=1}^{n} \lambda_j \leq 1 \\ & && \lambda_j \geq 0.\ (\forall j) \end{aligned}$$

If $\theta^* = \phi^*$, then $\text{DMU}_o$ exhibits CRS; otherwise if $\theta^* < \phi^*$, then $\text{DMU}_o$ exhibits DRS iff $\phi^* > f^*$, and $\text{DMU}_o$ exhibits IRS iff $\phi^* = f^*$.

These three different RTS methods are equivalent to estimate RTS parameter (Banker *et al.*, 1996b and Färe and Grosskopf, 1994). In empirical applications, one, however, finds that the CCR and BCC RTS methods may fail when DEA models have alternate optima. However, the scale efficiency index method does not suffer from the above problem, and hence is found robust. An elaborate discussion on the qualitative evaluation of RTS of different DEA models is found in Lõthgren and Tambour (1996) and Tone and Sahoo (2003, 2004, 2004a,b).

In the light of all possible multiple optima problem in the CCR and BCC methods, Banker and Thrall (1992) generalized by introducing new variables $v_o^+$ and $v_o^-$, which represent optimal solutions obtained by solving the dual of the output-oriented BCC model, with one more constraint $\sum_{i=1}^m v_i x_{io} + v_o = 1$ and replacing the objective function in this model by either $v_o^+ = \max v_o$ or $v_o^- = \min v_o$. They show that IRS operates iff $v_o^+ \geq v_o^- > 0$, DRS operates iff $0 > v_o^+ \geq v_o^-$ and CRS operates iff $v_o^+ \geq 0 \geq v_o^-$.

Banker *et al.* (1996b) point out that the concept of RTS is unambiguous only at point on the efficient facets of production technology. So the RTS for the inefficient units may depend upon whether the efficiency estimation is made through an input-oriented or output-oriented manner. A detailed method of doing so is found in the studies of Banker *et al.* (1996a), Tone (1996) and Cooper *et al.* (1999).

## 2.3 Quantitative Information on RTS

In this subsection we first discuss the quantitative evaluation of scale elasticity in both primal and dual environments, then point out their limitations, and finally suggest an alternative measure to get rid of such limitations.

### 2.3.1 Scale Elasticity in Primal Environment

If a DMU$_o$ is efficient in [BCC-O], then it holds that

$$-\sum_{r=1}^{s} u_r^* y_{ro} + \sum_{i=1}^{m} v_i^* x_{io} + v_o^* = 0$$

In order to unify multiple outputs and multiple inputs, let us define a scalar output $y$ and scalar input $x$ respectively as

$$y = \sum_{r=1}^{s} u_r^* y_{ro}, \text{ and } x = \sum_{i=1}^{m} v_i^* x_{io}.$$

Then, we have output ($y$) to input ($x$) relationship as

$$y = x + v_o^*.$$

From this equation, we define MP as

$$\text{MP} = \frac{dy}{dx} = 1,$$

and AP as

$$\text{AP} = \frac{y}{x} = \frac{1}{x} = \frac{1}{1 - v_o^*}, \text{ since } y = \sum_{r=1}^{s} u_r^* y_{ro} = 1.$$

Now, the scale elasticity in production ($\rho_p$)[4] is defined as

$$\rho_p = \frac{\text{MP}}{\text{AP}} = 1 - v_o^*.$$

However, if DMU$_o$ is inefficient, then $\rho_p$ equals $(1 - \frac{1}{\phi^*}.v_o^*)$. RTS are increasing, constant and decreasing if $v_o^* < 0$, $v_o^* = 0$ and $v_o^* > 0$ respectively.

To note here that as pointed out by Førsund and Hjalmarsson (2004), the scale elasticity in production, $\rho_p$ does not satisfy fully the requirement of RUPL as

$$\frac{\partial \rho_p(x, y)}{\partial x_{io}} = -v_o^* \frac{\partial (1/\phi)}{\partial x_{io}} = (1/\phi^2) v_o^* v_i, \ i = 1, \ldots, m.$$

IRS ($v_o^* < 0$) implies decreasing production elasticity in accordance with RUPL, while DRS ($v_o^* > 0$) implies an increasing $\rho_p$, thus violating the law.

### 2.3.2 Scale Elasticity in Dual Environment

Sueyoshi (1997) used the following dual of the VRS cost DEA model[5]

$$[\text{COST}] \qquad \gamma^* = \max \sum_{r=1}^{s} u_r y_{ro} + \omega_o$$

[4]Several authors (Førsund, 1996; Sueyoshi, 1999; and Fukuyama, 2001) have derived this same scale elasticity formula in different ways. However, we need to mention here that our approach towards the derivation of scale elasticity is much simpler.

[5]We call [COST] DEA model as the 'classical' cost efficiency model. On the DEA measure of cost efficiency, refer to Farrell (1957), Färe *et al.* (1985, 1994), Byrnes and Valdmans (1994), Coelli *et al.* (1998), Cooper *et al.* (1999) and Sueyoshi (1999). This

$$\text{subject to} \quad -\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{s} u_r y_{rj} + \omega_o \leq 0, (\forall j)$$
$$v_i \leq w_i, (\forall i)$$
$$u_r, v_i \geq 0, (\forall r, i), \ \omega_o : \text{ free}$$

to compute scale elasticity for $\text{DMU}_o$ (where $*$ represents optimal value). Following Baumol *et al.* (1988), he computed DSE at $(\boldsymbol{w}_o, \boldsymbol{y}_o)$ as

$$\rho_c (= \text{DSE}) = \gamma^* / (\sum_{r=1}^{s} u_r^* y_{ro}),$$

and shows the equivalence of IRS with DSE $> 1$, CRS with DSE $= 0$ and DRS with DSE $< 1$.

Note that under the assumption of unique optimal solution, the scale elasticity in primal form ($\rho_p$) in the BCC-O model and scale elasticity in dual form ($\rho_c$) in VRS Cost model are same when $\phi^* = 1$ and $v_o^* = \omega_o^* / (\omega_o^* - \gamma^*)$.[6]

model is primarily based on a number of simplifying assumptions: 1) its underlying factor-based technology set (input set $L(\boldsymbol{y})$) is *convex*, 2) input prices are exogenously given, and 3) input prices are known with full certainty by the DMUs. However, in many real-life applications, prices are not exogenous, but vary according to the actions by the DMUs (Chamberlin, 1933 and Robinson, 1933). Also, DMUs often face *ex ante* price uncertainty when making production decisions (Sandmo, 1971 and McCall, 1967). Looking at these realities leads us to suspect the harmless character of the *convexity* postulate for $L(y)$. See Cherchye *et al.* (2000a) and Kuosmanen (2003) for the details.

[6]Both primal and dual measures of scale elasticity are based on maintained hypothesis of *convex* structure of production technology ($L(\boldsymbol{y})$). Even though *convexity* axiom assumes away some economically important technological features such as *indivisible* production activities, economies of scale (= IRS), and economies of specialization (= diseconomies of scope), which all, in fact, result from *concavities* in production functions as pointed out by Farrell (1959), this postulate has rarely been exposed to empirical tests in a DEA tradition (a noteable exception is Briec *et al.*, 2004). Barring a few authors like Thrall (1999), recent literature favoring dropping of *convexity* axiom include, among others, Deprins *et al.* (1984), Scarf (1981a,b, 1986, 1994), Tulkens (1993), Tulkens and Vanden Eeckaut (1995), Bouhnik *et al.* (2001), Tone and Sahoo (2003), and Kuosmanen (2003); whereas the literature on some exciting economic analysis arising from violation of *convexity* include Yang and Ng (1993), Yang (1994), Yang and Rice (1994), Borland and Yang (1995) and Shi and Yang (1995). The only argument favoring *convexity* postulate is that there is possible reduction of small sample errors, which, but, comes at the cost of possible specification error, which is likely to be negligible in large samples. In fact, looking at the DEA-related economic literature (Afriat, 1972; Hanoch and Rothschild,

Otherwise,

$$\frac{\rho_c}{\rho_p} = \frac{1 - \frac{\omega_o^*}{\omega_o^* - \gamma^*}}{1 - \frac{1}{\phi^*} v_o^*}.$$

However, the details of the duality relationship between $\rho_p$ and $\rho_c$ can be found in Cooper *et al.* (1996) and Sueyoshi (1999, pp.1603-1604).

#### 2.3.3 An Alternative Measure of Scale Elasticity

This cost model however suffers from two problems: 1) scale elasticity in dual form $\rho_c$ is no different from its primal counterpart, i.e., scale elasticity in production $\rho_p$, thus giving the illusion that RTS and economies of scale are the one and same, and 2) this cost model declares a cost inefficient DMU as efficient one.

Concerning the first problem, it is to be noted that in the above production-cost relationship it has been implicitly maintained that in the special case of given input factor prices, the cost structure is entirely determined from the underlying production technology where IRS implies economies of scale. However, as the input market is typically imperfect in the real world, these two concepts can no longer be the same. A description concerning the conceptual differences between these two concepts lies beyond the scope of this study. However, the interested readers can refer to our earlier studies, e.g., Sahoo *et al.* (1999) and Tone and Sahoo (2003) where both the concepts have been critically reviewed by highlighting the fact that they have distinctive causative factors that do not permit them to be used interchangeably.

As regards the second problem, Tone (2002) has recently shown that if any two DMUs (A and B, say) have same amount of inputs and outputs, i.e., $\boldsymbol{x}_A = \boldsymbol{x}_B$ and $\boldsymbol{y}_A = \boldsymbol{y}_B$, and the unit input price vector of DMU A is twice that of DMU B, i.e., $\boldsymbol{w}_A = 2\boldsymbol{w}_B$, then both the DMUs exhibit the same cost efficiencies. This finding is 'strange' because they have achieved the same cost efficiency irrespective of their cost differential.[7] This second problem, which is due to the very definition of the 'classical' cost efficiency model

1972; and Varian, 1984) on production analysis, it appears that *convexity* properties are motivated from the perspective of economic objectives (e.g., cost minimization), but not as an inherent feature of the production technology. See Cherchye *et al.* (2000b) for details who provide empirical as well as theoretical arguments, less in favor of, but, mostly against the convexity postulate in DEA environment.

[7]See Tone (2002), Tone and Sahoo (2004a,b) for the detailed explanations.

accounting only for the direction (gradient) of unit price vector ($\boldsymbol{w}$) but not for its magnitude, are embedded in the *convex* structure of the supposed factor-based technology set $T$ as defined by

$$T = \{(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x} \geq X\boldsymbol{\lambda}, \boldsymbol{y} \leq Y\boldsymbol{\lambda}, \boldsymbol{e}\boldsymbol{\lambda} = 1, \boldsymbol{\lambda} \geq \boldsymbol{0}\},$$

where $T$ is defined only by using technical factors $X$ and $Y$, but has no concern with the unit input price $\boldsymbol{w}$.

Let us define an another cost-based technology set $T_c$[8] as

$$T_c = \{(\bar{\boldsymbol{x}}, \boldsymbol{y}) | \bar{\boldsymbol{x}} \geq \bar{X}\boldsymbol{\lambda}, \boldsymbol{y} \leq Y\boldsymbol{\lambda}, \boldsymbol{e}\boldsymbol{\lambda} = 1, \boldsymbol{\lambda} \geq \boldsymbol{0}\},$$

where $\bar{X} = (\bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_n)$ with $\bar{\boldsymbol{x}}_j = (w_{1j}x_{1j}, \ldots, w_{mj}x_{mj})^T$.

Here, we assume that the matrices $X$, $W$ and hence $\bar{X}$ are all positive. Also we assume that the elements of $\bar{x}_{ij} = (w_{ij}x_{ij})$ $(\forall(i,j))$ are denominated in homogeneous units, e.g., dollar, cent or pound so that adding up the elements of $\bar{x}_{ij}$ has a meaning.

The new cost efficiency $\bar{\gamma}^*$[9] is defined as

$$\bar{\gamma}^* = \boldsymbol{e}\bar{\boldsymbol{x}}_o^* / \boldsymbol{e}\bar{\boldsymbol{x}}_o,$$

where $\bar{\boldsymbol{x}}_o^*$ is the optimal solution of the LP given below.

$$[\text{NCOST}] \qquad \boldsymbol{e}\bar{\boldsymbol{x}}^* = \min \ \boldsymbol{e}\bar{\boldsymbol{x}}$$

---

[8] This technology structure was developed while looking at the compelling arguments for dropping of *convexity* axiom, though it is regarded as a standard regularity property without having a valid justification. To cite one important argument for admitting *indivisibilities* as inherent feature of $T$ (but which was assumed away through convexification of it), Scarf (1994) maintains that if technology involves serious indivisibilities, then it is impossible to detect optimal equilibrium of the firm facing CRS. If production really does obey CRS, there is no economic justification for the existence of large firms, i.e., there is nothing to be gained by organizing economic activity in large, durable and complex units such as assembly lines, bridges, transportation and communication networks, giant presses, and complex manufacturing plants, which are available in specific discrete sizes, and whose economic usefulness manifests itself only when the scale of operation is large. However, our cost-based technology set $T_c$ approximates reality in terms of not requiring *convexity* assumption for the factor-based technology set $L(\boldsymbol{y})$. But, note that our convexification of $T_c$ is again an assumption.

[9] Unlike the classical cost efficiency estimate $\gamma^*$, this new cost efficiency estimate $\bar{\gamma}^*$ does always satisfy the property of 'monotonicity' with respect to input and input price, i.e., if $\boldsymbol{x}_A = k\boldsymbol{x}_B (k > 0)$, $\boldsymbol{y}_A = \boldsymbol{y}_B$ and $\boldsymbol{w}_A = \boldsymbol{w}_B$, then we have $\bar{\gamma}^*_A = (1/k)\bar{\gamma}^*_B$, and if $\boldsymbol{x}_A = \boldsymbol{x}_B$, $\boldsymbol{y}_A = \boldsymbol{y}_B$ and $\boldsymbol{w}_A = k\boldsymbol{w}_B (k > 0)$, then we have $\bar{\gamma}^*_A = (1/k)\bar{\gamma}^*_B$

$$\begin{aligned} \text{subject to} \quad & \bar{\boldsymbol{x}} \geq \bar{X}\boldsymbol{\lambda} \\ & \boldsymbol{y}_o \leq Y\boldsymbol{\lambda} \\ & \boldsymbol{e}\boldsymbol{\lambda} = 1 \\ & \boldsymbol{\lambda} \geq \boldsymbol{0}. \end{aligned}$$

The new cost efficiency is evaluated by the program [NCOST].[10] The constraint includes $m$ inequalities, since $\bar{\boldsymbol{x}}$ is an $m$-vector. Considering the objective function form $\boldsymbol{e}\bar{\boldsymbol{x}}$ and the input constraints in [NCOST], the aggregation of these $m$ constraints into one yields the following new program [NCOST-1]:

$$\begin{aligned} \text{[NCOST-1]} \quad & \min \ \boldsymbol{e}\bar{\boldsymbol{x}} \\ \text{subject to} \quad & \boldsymbol{e}\bar{\boldsymbol{x}} \geq \boldsymbol{e}\bar{X}\boldsymbol{\lambda} \\ & \boldsymbol{y}_o \leq Y\boldsymbol{\lambda} \\ & \boldsymbol{e}\boldsymbol{\lambda} = 1 \\ & \boldsymbol{\lambda} \geq \boldsymbol{0}. \end{aligned}$$

This program is simpler than the former in that it has only one aggregated constraint on the input part.

This aggregated model presents a correspondence between cost (input) and production (outputs). Let us denote $\boldsymbol{e}\bar{\boldsymbol{x}}_j$ by $\bar{w}_j$, i.e.,

$$\bar{w}_j = \sum_{i=1}^{m} x_{ij} w_{ij}. \ (j = 1, \ldots, n)$$

$\bar{w}_j$ is the input cost for the DMU$_j$ for producing the output vector $\boldsymbol{y}_j$. Using this notation and notifying the expressions in [NCOST-1], the new aggregated scheme reduces to the following LP:

$$\text{[NCOST-2]} \qquad \min \sum_{j=1}^{n} \bar{w}_j \lambda_j$$

[10] As regards the evaluation of cost efficiency, *convex* cost-based technology set $T_c$ proves to be superior over the *convex* factor-based technology set $T$ on two counts. First, cost efficiency based on $T_c$, as just discussed in the earlier footnote, does always satisfy *monotonicity* property with respect to input and input price. Second, this model does not necessarily require the assumption of *convexity* to hold for factor-based technology set $L(\boldsymbol{y})$, since *convexity* assumption for $L(\boldsymbol{y})$, though harmless, is not at all required for analysis of cost efficiency. And, more importantly, convex $L(y)$ excludes some interesting economic features such as *indivisibilities*, scale, and scope.

$$\begin{aligned} \text{subject to} \quad & \boldsymbol{y}_o \leq Y\boldsymbol{\lambda} \\ & \boldsymbol{e\lambda} = 1 \\ & \boldsymbol{\lambda} \geq \boldsymbol{0}. \end{aligned}$$

To compute scale elasticity, we consider the dual LP for [NCOST-2] to serve the purpose.

$$\begin{aligned} \text{[NCOST(Dual)]} \quad & \max \sum_{r=1}^{s} u_r y_{ro} + \delta_o \\ \text{subject to} \quad & \sum_{r=1}^{s} u_r y_{rj} + \delta \leq \bar{w}_j \quad (j = 1, \ldots, n) \\ & u_r \geq 0 \ (\forall j), \ \ \delta : \text{free}. \end{aligned}$$

We have now scale elasticity[11] at $(\bar{w}_o, \boldsymbol{y}_o)$ as

$$\rho_c = \frac{1}{1 - \delta^*/\bar{w}_o}.$$

RTS are *increasing* if $\delta^* > 0$ $(\rho_c < 1)$, *constant* if $\delta^* = 0$ $(\rho_c = 1)$, and *decreasing* if $\delta^* < 0$ $(\rho_c > 1)$.

If there are multiple optima in $\delta^*$, then let its sup (inf) be $\bar{\delta}^*$ $(\underline{\delta}^*)$. Then RTS are characterized as *increasing* if $\underline{\delta}^* > 0$ $(\underline{\rho_c} > 1)$, *constant* if $\bar{\delta}^* \geq 0 \geq \underline{\delta}^*$ $(\underline{\rho_c} \leq 1 \leq \bar{\rho}_c)$, and *decreasing* if $\bar{\delta}^* < 0$ $(\bar{\rho}_c < 1)$.

To note that the method discussed above for characterizing RTS holds true for efficient DMU. However, if a DMU is found inefficient, then project it onto the efficient frontier, and then solve the above LP to compute $\rho_c$[12].

[11]This measure of RTS is very different from the standard approach discussed in Aly *et al.* (1990). This is the case because the former is derived from the cost-based technology set whereas the latter is from the factor-based technology set.

[12]A detailed method of doing so is extensively discussed in Tone and Sahoo (2004b). One application of the use of this new method in the area of Indian life insurance is found in Tone and Sahoo (2004a).

# 3 Parametric Models

## 3.1 Scale Elasticity in Translog Production Function

Following Griliches and Ringstad (1971) and Christensen *et al.* (1973), the technology characterized by translog production function is represented by

$$\begin{aligned}\ln y_j &= \alpha_o + \sum_{i=1}^{m} \alpha_i \ln x_{ij} + \frac{1}{2}\sum_{i=1}^{m}\sum_{i'=1}^{m} \alpha_{ii'}(\ln x_{ij})(\ln x_{i'j}) \\ &\quad + \delta_t t + \delta_{tt} t^2 + \sum_{i=1}^{m} \delta_{it}(\ln x_{ij})t + v_j,\end{aligned}$$

where $v_j$ is assumed to follow a normal distribution with 0 mean and $\sigma_v^2$ variance, i.e., $v_j \sim N(0, \sigma_v^2)$. The local scale elasticity of firm $j$ in production $\rho_{jp}$ is computed as:

$$\rho_{jp} = \sum_{i=1}^{m} \frac{\partial \ln y_j}{\partial \ln x_{ij}} = \sum_{i=1}^{m}\Big[\alpha_i + \frac{1}{2}\sum_{i'=1}^{m} \alpha_{ii'} \ln x_{i'j} + \frac{1}{2}\sum_{i'=1}^{m} \alpha_{i'i} \ln x_{i'j} + \sum_{i=1}^{m} \delta_{it}.t\Big].$$

## 3.2 Scale Elasticity in Translog Cost Function

The following translog cost function:

$$\begin{aligned}\ln C_j(\boldsymbol{w}, y) &= \alpha_o + \sum_{i=1}^{m} \alpha_i \ln w_{ij} + \beta_y \ln y_j + \frac{1}{2}\sum_{i=1}^{m}\sum_{i'=1}^{m} \alpha_{ii'}(\ln w_{ij})(\ln w_{i'j}) \\ &\quad + \frac{1}{2}\beta_{yy}(\ln y_j)^2 + \sum_{i=1}^{m} \gamma_{iy}(\ln w_{ij})(\ln y_j) + \delta_t t + \delta_{tt} t^2 \\ &\quad + \sum_{i=1}^{m} \delta_{it}(\ln w_{ij})t + \delta_{yt}(\ln y_j)t + v_j,\end{aligned}$$

can be employed to estimate the cost-output relationship with the following restrictions:

$$\sum_{i=1}^{m} \alpha_i = 1,\ \sum_{i'=1}^{m} \alpha_{ii'} = 0\ (i = 1, \ldots, m), \alpha_{ii'} = \alpha_{i'i}\ (\forall i, i'),\ \sum_{i=1}^{m} \gamma_{iy} = 0\ \sum_{i=1}^{m} \delta_{it} = 0$$

to ensure linear homogeneity in input prices, where $v_j \sim N(0, \sigma_v^2)$. The local scale elasticity in cost of firm $j$, $\rho_{jc}$, is then computed by

$$\begin{aligned}\rho_{jc} &= 1\bigg/\frac{\partial \ln C_j(\boldsymbol{w}, y)}{\partial \ln y_j} \\ &= 1\bigg/\left[\beta_y + \beta_{yy}\ln y_j + \sum_{i=1}^{m}\gamma_{iy}\ln w_{ij} + \delta_{yt}.t\right].\end{aligned}$$

To note here that in the multiple input and multiple output environment, under standard regularity conditions, at the cost minimizing input vector $\boldsymbol{x}^*(\boldsymbol{w}, \boldsymbol{y})$, the scale elasticity estimates in both primal and dual environments ($\rho_p$ and $\rho_c$) are the same, i.e.,

$$\rho_{jp} \equiv -\sum_{i}^{m} x_{ij}^* \frac{\partial\psi}{\partial x_{ij}}\bigg/\sum_{r}^{s} y_{rj}\frac{\partial\psi}{\partial y_{rj}} = C_j(\boldsymbol{w}, \boldsymbol{y})\bigg/\sum_{r=1}^{s} y_{rj}\frac{\partial C_j(\boldsymbol{w}, \boldsymbol{y})}{\partial y_{rj}} \equiv \rho_{jc}.$$

See Baumol *et al.* (1988, pp. 63-64) for its proof. Further, it is shown there that any differentiable cost function, whatever the number of outputs involved, and whether or not it is derived from a homogeneous production process, has a local degree of homogeneity, which is reciprocal of the homogeneity parameter of the production process.

## 3.3 Scale Elasticity in Stochastic Frontier Translog Cost Function

Following Stevenson (1980) and Kumbhakar and Lovell (2000), in the VRS environment the following stochastic version of translog cost function:

$$\begin{aligned}\ln C_j(\boldsymbol{w}, y) &= \alpha_o + \sum_{i=1}^{m}\alpha_i \ln w_{ij} + \beta_y \ln y_j + \frac{1}{2}\sum_{i=1}^{m}\sum_{i'=1}^{m}\alpha_{ii'}(\ln w_{ij})(\ln w_{i'j}) \\ &\quad + \frac{1}{2}\beta_{yy}(\ln y_j)^2 + \sum_{i=1}^{m}\gamma_{iy}(\ln w_{ij})(\ln y_j) + \delta_t t + \delta_{tt}t^2 \\ &\quad + \sum_{i=1}^{m}\delta_{it}(\ln w_{ij})t + \delta_{yt}(\ln y_j)t + v_j + u_j,\end{aligned}$$

can be employed to estimate the cost-output relationship with the following restrictions:

$$\sum_{i=1}^{m}\alpha_i = 1,\ \sum_{i'=1}^{m}\alpha_{ii'} = 0\ (i = 1, \ldots, m), \alpha_{ii'} = \alpha_{i'i}\ (\forall i, i'),\ \sum_{i=1}^{m}\gamma_i = 0,\ \sum_{i=1}^{m}\delta_{it} = 0$$

to ensure linear homogeneity in input prices. Here, $v_j \sim N(0, \sigma_v^2)$ and the inefficiency term $u_j$ is assumed to follow a half-normal distribution or a truncated normal distribution.

The local scale elasticity of firm $j$ in cost environment, $\rho_{jc}$, is then computed by

$$\begin{aligned} \rho_{jc} &= 1\bigg/\frac{\partial \ln C_j(\boldsymbol{w}, y)}{\partial \ln y_j} \\ &= 1\bigg/\left[\beta_y + \beta_{yy} \ln y_j + \sum_{i=1}^{m} \gamma_{iy} \ln w_{ij} + \delta_{yt} t\right]. \end{aligned}$$

In parametric literature, the *ad hoc* functional forms chosen for the production technology without knowing its underlying relationship between inputs and outputs, often, cast doubt on the results. To cite one study here, for all functional forms Hasenkamp (1976) considered, his findings suggest economies of scale, where as for flexible functional forms, his results reveal economies of specialization (i.e., violation of *convexity* assumption). In spite of large empirical evidence (not reported here) against the convexity axiom for factor-based technology, the parametric approach imposes restrictions on the parameters of production and cost functions to bring *convexity* assumption in for the underlying production technology.

# 4 Empirical Results

## 4.1 The Data

For the empirical illustration concerning the comparison between [NCOST] and translog estimates of scale elasticity, we have used the panel data of 18 electric power companies in the U.S. over a period spanning from 1992 to 1999. These companies have several functions such as generation, transmission, distribution, and so on. However, our study uses only generation data because we focus on the productive performance of these companies. The US data set is obtained from the 'FERC FORM1' disclosed by the Federal Energy Regulatory Commission (FERC).

Concerning the selection of data on inputs and output, we have considered three inputs: Capital ($x_1$), Fuel ($x_2$), and Labor ($x_3$), and one output: electric power generated (GWh) ($y$). Capital is taken as the total installed generating capacity (MW). Labor is considered the total number of employees. Here, because of unavailability of data, we have not taken outsourcing into consideration. Fuel is taken as the total consumed fuel not only in fossil power plant but also in nuclear power plant. The amount of fuel consumption for gas, coal and petroleum is converted to British Thermal Unit (BTU). Compared to the fossil fuel, the amount of consumption of nuclear fuel is difficult to assess in which case it is calculated backward from generated electric power from nuclear power plants assuming thermal efficiency of 35% and then is converted into BTU.

As regards the price data, unit price of capital ($w_1$) is taken as the ratio of total capital cost, $c(x_1)$, (which is the total sum of Maintenance and Depreciation expenses in generation division) to capital input ($x_1$). Unit price of fuel ($w_2$) is taken as the ratio of total fuel cost, $c(x_2)$, (which is the total sum of fossil, nuclear and other fuel expenses) to fuel input $x_2$. Unit price of labor, i.e., wage rate ($w_3$) is computed as the ratio of total wages and salaries, $c(x_3)$, to total number of employees ($x_3$). All input cost data are realized through Producer Prices Index (PPI) of U.S.

In the spirit of earlier studies of Boussofiane *et al.* (1991), Ray and Kim (1995) and Sueyoshi (1997, 1999), each year's company's annual performance is considered here a distinct DMU. So, we have in total 144 (=18x8) DMUs in our sample period. Prior to formal modeling of production process, it is common in empirical work in the DEA literature (Färe *et al.*, 1987, Grosskopf and Valdmanis, 1987, and Rangan *et al.*, 1988) to present descriptive statistics on the input-output data, which serves to provide some intution on the plausibility of the derivative DEA-efficiency coefficients. In a similar context, Besley (1989) and Hammond (1981) have proposed the evaluation of efficiency "*ex ante*" and "*ex post*." Thus, the efficiency predictions in this section are termed "*ex ante*" in the sense that they are derived from the descriptive statistics on the data. Analogously, the DEA efficiency scores can be interpreted as "*ex post*" predictions of efficiency. Table 1 containts means, maxima, minima and standatrd deviations of input-output data based on the full panel data set comprising 18 companies for eight years.

Table 1: Descriptive Statistics Input and Output Data

From the extreme values in the data on single input (aggerated cost) and single output (electric power generated), our crude *ex ante* prediction is that DMU 2 in 1997 and DMU 5 in 1999 might be among the best practice companies with minimal cost and maximal output respectively. As regards the curvature of the cost frontier, the scatter plot of cost *vis-a-vis* output is shown below in Figure 1, and the cost frontier is then drawn, keeping in mind the fact that the cost-based technology set is *convex*.

Figure 1: Cost Frontier of US Electric Companies

We find in this figure eight companies over various years: 2(1997), 5(1999), 8(1998, 1999), 13(1999), 16(1999), 17(1997) and 18(1996), characterized, respectively, by points A, H, F, G, D, C, E and B, exhibiting efficient performance, and the companies in the remaining years of operation under inefficient performance behavior. Concerning the returns to scale behavior of the efficient units, we find here DMUs A and B operating under IRS, DMU C under CRS, and DMUs D, E, F, G and H under DRS. And as regards the RTS behavior of inefficient units, our *ex ante* prediction is that a large number of units appear to be operating under DRS.

## 4.2 RTS Results in DEA Models

We now turn to discuss the quantitative information on input-oriented RTS behaviour of the US electric companies in both primal and dual environments. We only analyze here the scale elasticity estimates of these 18 companies obtained in both BCC and [NCOST] DEA models.[13]

Concerning the scale elasticity behavior in production environment, BCC estimates reveal more units operating under IRS (74 units under IRS, 34 units under CRS and 36 units under DRS). However, as regards the evaluation of

[13] The elasticity estimates for each of the 18 companies over a period of eight years are all reported in Appendix A. The methods used for the computation of scale elasticity are BCC-I and [NCOST-I]. Both BCC-I and [NCOST-I] estimates on scale elasticity are the arithmetic average of lower (inf.) and upper (sup.) bounds of elasticity, thus taking care of the problem of the case of alternate optima.

scale elasticity in cost environment,[14] [NCOST] results favors more units operating under DRS (48 units under IRS, 31 units under CRS and 65 units under DRS). This apparent contradiction coming out from these two sets of estimates can be explained from the fact that the BCC estimates are derived from a *convex* factor-based technology set where as the [NCOST] estimates are from *convex* cost-based technology set.

However, in the spirit of Färe *et al.* (1987), an *ad hoc* procedure has been adopted here to limit the impact of noise on the estimated level of scale elasticity. Here a more stable picture of performance can be extracted by performing separate envelopments on the successive eight cross sections, and derive the mean-year scale elasticity score of these 18 companies over a period of eight years. Spearman's rank correlation coefficient was used to establish whether noise in outcomes made any unexpected or abrupt change in scale elasticity rankings of the companies year-on-year *vis-a-vis* mean efficiencies of these units over eight years. High value of rank correlation coefficient was taken to represent stable scale elasticity scores, which reflect underlying levels of performance. The rank correlation estimates are exhibited below in Table 2.

Table 2: Spearman's Rank Correlation

The earlier [NCOST]-based scale elasticity estimates, based on the panel data, of these 18 companies are found to be significantly correlated with their cross-section based elasticity estimates for all the eight years. This finding of significanr rank correlation suggests that these scale elasticity estimates are stable, and can be taken as the basis of acceptable targets. The cross-section based mean-year scale elasticity estimates of these companies for each of these years are found to be, respectively, 1.083, 1.631, 1.237, 2.275, 1.520, 1.024, 1.108 and 1.477. This finding is in broad agreement with our panel data based mean-year elasticity estimates (3.326, 4.498, 3.831, 2.895, 2.330, 2.157,

[14]Note here that, classical dual [COST] estimates of scale elasticity are not reported because we have already demonstrated the superiority of [NCOST] estimates over the classical [COST] estimates in our forthcoming paper (Tone and Sahoo, 2004b).

3.117 and 3.498).[15] However, the magnitude of the cross-section based mean-year elasticity estimates are lower compared to those in our panel-based case. This result is not surprising because of more exreme scale elasticity values in the latter case, otherwise, the result would have been reverse because the creation of envelops for successive cross-sections limits the number of firms to 18 each, which increases the chance of inefficient units to be come efficient in our DEA evaluation process.

## 4.3 RTS Results in Econometric Models

We have considered here both translog cost function and stochastic frontier translog cost function models for the computation of scale elasticity.

### 4.3.1 Translog Cost Function Model

For the computation of scale elasticity,[16], we have considered first the non-frontier and non-stochastic case where we employed translog cost function (TCF), without and with time trend. We have formulated the following structure given below where the third input price ($w_3$) is standardized.

$$\begin{aligned}
\ln \frac{C_j}{w_{3j}}(\boldsymbol{w}, \boldsymbol{y}) &= \alpha_o + \alpha_1 \ln \frac{w_{1j}}{w_{3j}} + \alpha_2 \ln \frac{w_{2j}}{w_{3j}} + \beta_y \ln y_j + \frac{1}{2}\beta_{yy}(\ln y_j)^2 \\
&\quad + \frac{1}{2}\alpha_{11}\Big(\ln \frac{w_{1j}}{w_{3j}}\Big)^2 + \frac{1}{2}\alpha_{22}\Big(\ln \frac{w_{2j}}{w_{3j}}\Big)^2 + \alpha_{12}\Big(\ln \frac{w_{1j}}{w_{3j}}\Big)\Big(\ln \frac{w_{2j}}{w_{3j}}\Big) \\
&\quad + \gamma_{1y}\Big(\ln \frac{w_{1j}}{w_{3j}}\Big)(\ln y_j) + \gamma_{2y}\Big(\ln \frac{w_{2j}}{w_{3j}}\Big)(\ln y_j) + \delta_t t + \delta_{tt} t^2 \\
&\quad + \delta_{1t}\Big(\ln \frac{w_{1j}}{w_{3j}}\Big)t + \delta_{2t}\Big(\ln \frac{w_{2j}}{w_{3j}}\Big)t + \delta_{yt}(\ln y_j)t + v_j
\end{aligned}$$

[15] It is to be noted here that the magnitude of each of these mean estimates, being all greater than 1 does not mean that Electric industry operates under IRS for all these years. In fact, these mean values are greatly distorted because of a few extreme scale elasticity values for some units (units 2 and 6 in cross-section based data, and units 2, 6, 11 and 15 in panel based data). These estreme values are due to the fact that DEA evaluates RTS of inefficient unit by upwardly projecting them onto the efficient frontier. Otherwise, the trend is clearly in favor of DRS in most of the cases.

[16] Before scale elasticity calculations are made, we applied the Likelihood Ratio (LR) test to verify whether CRS assumption is justified for the industry, and we find that it is no longer valid, i.e., industry exhibits VRS.

The methods of estimation used are Ordinary Least Squares and Auto Regression. The estimates of coefficients along with their standard errors are all reported in Table 3.

Table 3: Estimates of Translog Cost Function

Regression is run for three times: first one with no time trend using OLS method (TCF1), second one with time trend using OLS (TCF2), and third one with time trend using Auto regression. As is seen in this table, the consideration of time trend appears to have no noticeable impact on the estimates of translog cost coefficients and values of $R^2$. A graphical illustration for the comparison of scale elasticity estimates obtained from the above three regressions is shown below in Figure 2.

Figure 2: Scale Elasticity Behavior in Translog Cost Function Models

We observe here that the average elasticity estimates of US electric companies exhibit IRS for all the eight years.[17] Barring the first case (TCF1) where we have not considered the time trend, average trend scale elasticity behavior of these companies in the other two models are same.

However, the problem with the TCF model is that it does not take into account the possibility that a firm's performance may be affected by factors entirely outside its control such as poor machine performance, bad weather, input supply breakdown, etc., and by factors under its control labeled 'inefficiency.' In effect, the single term inefficiency mixed with the effects of exogeneous shocks, measurement error and inefficiency is subject to questions. So there appears the stochastic frontier, which was motivated by the idea that deviations from the cost frontier might not be entirely under the control of the firms being studied. So the error term in the stochastic frontier

[17] The detailed estimates of scale elasticity of these companies over all the years are shown in Appendix B.

is composed of two parts: one part, $v_j$ (systematic), which is unrestricted, permits the random variation of the frontier across firms and captures the effect of measurement error, other statistical noise and randdom shocks outside the control of firm, and the other part, $u_j$ (one-sided error term), captures the effect of inefficiency relative to the stochastic frontier.

### 4.3.2 Stochastic Frontier Translog Cost Function

We have employed here the stochastic and frontier version of the same translog cost function where we assumed two distributions for the inefficiency term: one with half-normal distribution (STCF1) and other with truncated normal distribution (STCF2). Now, for computing scale elasticity in VRS environment, we have formulated the following structure given below where the third input price ($w_3$) is standardized.

$$\begin{aligned}\ln \frac{C_j}{w_{3j}}(\boldsymbol{w}, \boldsymbol{y}) &= \alpha_o + \alpha_1 \ln \frac{w_{1j}}{w_{3j}} + \alpha_2 \ln \frac{w_{2j}}{w_{3j}} + \beta_y \ln y_j + \frac{1}{2}\beta_{yy}(\ln y_j)^2 \\ &+\frac{1}{2}\alpha_{11}\left(\ln \frac{w_{1j}}{w_{3j}}\right)^2 + \frac{1}{2}\alpha_{22}\left(\ln \frac{w_{2j}}{w_{3j}}\right)^2 + \alpha_{12}\left(\ln \frac{w_{1j}}{w_{3j}}\right)\left(\ln \frac{w_{2j}}{w_{3j}}\right) \\ &+\gamma_{1y}\left(\ln \frac{w_{1j}}{w_{3j}}\right)(\ln y_j) + \gamma_{2y}\left(\ln \frac{w_{2j}}{w_{3j}}\right)(\ln y_j) + \delta_t t + \delta_{tt} t^2 \\ &+\delta_{1t}\left(\ln \frac{w_{1j}}{w_{3j}}\right)t + \delta_{2t}\left(\ln \frac{w_{2j}}{w_{3j}}\right)t + \delta_{yt}(\ln y_j)t + v_j + u_j\end{aligned}$$

The results on the coefficients along with their standard errors in each case are all exhibited in Table 4.

Table 4: Estimates of Stochastic Frontier Translog Cost Functions

Concerning average scale elasticity behavior of the US electric companies,[18] we find also in Figure 3 the trend scale elasticity behavior in each of the two

[18] Like our non-stochastic case, to justify the validity of computation of scale elasticity, we have too applied here LR-test to examine whether industry exhibits CRS. We find in each of two cases that CRS can no longer be the maintained assumption as LR-test stastistic, $\chi^2$, is found to be statistically significant.

cases more or less the same.[19] However, there appears to be a striking difference between the trend scale elasticity behavior revealed through TCF and STCF estimates where the TCF trend completely dominates its stochastic counterpart. This result is not surprising because the main factor contributing this difference is the statistical 'noise' factor. Thus, neglecting this factor, will lead, at least in this case, to a too greater picture of scale elasticity estimates in TCF setting.

## 4.4 RTS Results: A Comparison

Let us now turn to compare the scale elasticity estimates obtained in both TCF and STCF models with those in our new method. To start with, let us examine the distribution of returns to scale between DEA and translog methods, which is exhibited in Table 5.

Table 5: Distribution of RTS in DEA and Translog Models

We find here that, as expected from our *ex ante* prediction, [NCOST] reports more units operation under DRS (48 units under IRS, 31 units under CRS, and 65 units under DRS). In TCF setting, both OLS and auto regression methods yield the diametrically opposite information on RTS possibilities, favoring more units opearting under IRS. The statistical *t*-test results with 5% confidence interval suggest that 69-78 companies operate under IRS, 40-55 units under CRS and 20-26 units under DRS. However, the RTS possibilities in STCF environment do not go in line with those in TCF setting, favoring more units of operation under CRS. The statistical *t*-test results with 5% confience interval indicate that in the case of inefficiency term following half-normal distribution, 29 units are found operating under IRS, 83 units under CRS and the remaining 32 units under DRS, whereas in the case of the truncated normal distribution, 47 units are found operating under IRS, 59 units under CRS and the remaining 38 units under DRS. This divergent information concerning the distribution of RTS in STCF setting is largely

[19]The stochastic scale elasticity estimates of each of these 18 companies over eight years are exhibited in Appendix C.

attributed to the functional forms for the inefficiency distribution. Since, the RTS behavior in STFC setting is largely dependent upon the distribution of the inefficiency term, the question remains to verifying the validity of the assumption of half-normal and truncated normal distribution. However, the apparent contradiction in revealing RTS information between [NCOST] and STFC is clearly evident in that in the former DRS is favored for more units (as per our *a priori ex ante* prediction), and in the latter, IRS and CRS are favored for majority of the companies.

Concerning the investigation of optimal scale of operations, our new model [NCOST] offers different policy prescription in that it suggests the lower level of operation as against the very high level of such operation yielded in both TCF and STCF models. This divergent information on RTS behavior of these units yielded from these two methods can further be reconfirmed from the $\chi^2$ statistic results. Table 6 [A through F] gives a comparative picture of estimates of scale elasticity of [NCOST] *vis-a-vis* with those in BCC, TCF and STCF methods.

Table 6: Comparison of Scale Elasticity Estimates

We find $\chi^2$ statistics in all the cases significant, which suggest that scale elasticity estimates obtained in [NCOST] model are significantly different from those in BCC, TCF and STCF models. The difference in results between BCC and [NCOST] models is due to the fact that both are not dual to each other. However, the difference in results obtained from translog cost model and [NCOST] model can be explained through by comparing the theoretical setups of their models. The translog sets of estimates are based on the *convex* factor-based technology set whereas [NCOST] set of estimates are on *convex* cost-based technology set.

Now, it is worth comparing our findings with those in some of early empirical studies. We find that the findings of these studies are unable to persuasively establish whether scale economies are present. In fact, the findings are contradictory to each other. To cite a few, among others, Kaserman and Mayo (1991) found the existence of economies of vertical integration in the privately owned U.S. electric power companies in 1981, but division-specific economies of scales were absent, which go in line with the findings of our stochastic translog method. Gilsdorf (1994) and Hayashi *et al.* (1997), on the other hand, found the existence of division-specific scale economies, supporting our [NCOST] findings. We believe that the findings of our [NCOST] model that precisely uncovers the real differences in input factor prices, are realistic, otherwise, policy makers in most of the states in the U.S. would not have brought changes in appropriate policy in terms of introducing competition into generation division.

## 4.5 RTS in [NCOST]: Is It Rational?

Before proceeding further, let us first intuitively demonstrate the rationality of the empirical evaluation of scale elasticity estimates[20] in our new method. We see in Figure 1 that cost-based technology frontier is piecewise-linear comprising only eight efficient DMUs (A: 2(1997), B: 18(1996), C: 16(1999), D: 13(1999), E: 17(1997), F: 8(1999), G: 8(1998) and H: 5(1999)), and seven efficient facets ($\overline{AB}$, $\overline{BC}$, $\overline{CD}$, $\overline{DE}$, $\overline{EF}$, $\overline{FG}$ and $\overline{GH}$). We clearly see here that the first two facets ($\overline{AB}$ and $\overline{BC}$) are characterized by IRS, and DRS prevails on the remaining facets. The cost, output and the computational procedure for the calculation of scale elasticity are all exhibited in Table 7.

Table 7: Empirical Evaluation of Scale Elasticity

Now let us compute the scale elasticity of an inefficient unit such as 16(1997) (also indicated by point **I** in Figure 1). This DMU's input efficiency score in our [NCOST] model is 0.410602, and its peer DMUs are D [= 13(1999)] and E [=17(1997)]. So the projected cost and output values of this DMU are, respectively, 753544325.196 and 44442.017. The average cost, marginal cost and scale elasticity of DMU I [16(1997)] are computed as follows:

[20]Note that [NCOST] estimates of elasticity presented earlier are based on one output and three distinct input costs. This was ideally done for better comparison because translog cost model deals with three inputs. Since elasticity calculation based on three input costs and one output cannot be exhibited in graph, we presented here in Table 7 the same with help of one aggregated input cost and one output.

$$\text{AC} = C/Y = 753544325.196/44442.017 = 16955.673,$$
$$\text{MC} = \frac{1}{\text{slope of facet } \overline{DE}} = dC/dY = 18748.719,$$
$$\text{Scale Elasticity} = \text{AC}/\text{MC} = 16955.673/18748.719 = 0.904364.$$

The scale elasticity score of 0.904364 indicates that DMU I operates under DRS. The fundamental difference between our new [NCOST] method and translog cost method for the estimation of cost frontier is that in the former the structure of the frontier is assumed to be *convex* whereas in the latter, the technology structure, depending upon data, may be either *convex* or *non-convex*. We intend to verify this in the light of three new artificial data sets: one in which cost-based technology set is convex, and in the other two, it is not.

## 4.6 Verifying RTS in New Data Sets

We now examine this relationship in three new atrificial data sets, each consisting of 144 firms. The scatter plots of output *vis-a-vis* cost corresponding to Data Sets I, II and III are exhibited, respectively, in Figures: 3, 4 and 5. We find here that the observed cost-based technology set in Figure 3, corresponding to Data Set I, is *convex* whereas corresponding to Data Sets, II and III in Figure 4 and Figure 5, they are not. However, in all the cases, the hypothesis of *convex* structure for the underlying cost-based technology is maintained, as can be seen from both the figures where the cost frontiers are made of thick piecewise linear lines.

Figure 3: Cost Frontier (Data Set I)

Figure 4: Cost Frontier (Data Set II)

Figure 5: Cost Frontier (Data Set III)

We now report the summary of scale elasticity estimates for each firm. The computational procedure is just the same as the one we explained while demonstrating the rationality of empirical evaluation of scale elasticity in Table 7. A simple glance at Figures: 3, 4 and 5 reveals that the cost frontiers, corresponding to Data Sets: I, II and III, are, respectively, made of three (A, B and C). four (A, B, C and D) and three (A, B, C) efficient firms. The distribution of RTS in [NCOST] and TCF settings is exhibited in Table 8.

Table 8: Distribution of RTS in Data Sets: I, II and III

We find here that in case of Data Set I where observed cost-based technology set is convex, as expected, both the methods favor more number of units operating under IRS. [NCOST] reports 68 units opearting under IRS, 33 units under CRS, and the remaining 43 units under DRS, and with 5% level of confidence, our $t$-test results in TCF setting report 78 units under IRS, 48 units under CRS and 18 units under DRS. Now, coming to Data Set II where observed cost-based technology is not convex, both the methods are in broad agrement to our *a priori* expectation that most of the units operate under IRS. [NCOST] here reveals 134 units under IRS, 7 units under CRS and 3 units under DRS as opposed to all the 144 units (with 5% level of confidence interval) exhibiting IRS in translog method. However, in Data Set III where observed cost-based technology is also not convex, our *a priori* expectation was that more number of units appeared to be operating under DRS. But, both the methods, here, yield diametrically opposite conclusion concerning RTS possibilities. We find [NCOST] reporting 39 units operating under IRS, 16 units under CRS, and 89 units under DRS, whereas translog method yields 103, 41 and 0 units operating, respectively, under IRS, CRS and DRS.

The potential problem arises in the choice of these two methods when the observed cost-output relationship violates the requirement of *convex* structure of the cost-based technology set. The translog cost model has the well reputation of being flexible in approximating arbitrary production technologies. So the comparison between these two sets of estimates leads us to conclude that both methods broadly yield same information on returns to scale possibilities in the case of data where the observed cost-based technology structure is *convex*, and in case of data like Data Set II and III, where this *convex* structure is violated, the inference on RTS possibilities drawn from these two methods are not clear cut.

The relevant point to bear in mind here is that the difference between these two sets of scale elasticity estimates is largely attributed to the difference in the theoretical setups in these two models. In the traditional production economics literature, it is typically assumed that every DMU is endowed with a particular input price vector. For a particular factor-based *convex* technology set, every DMU's efficient cost frontier is *convex* in the output given its input price vector. In this setup, it is easy to see that if one plots the least cost values against output for various DMUs that are characterized by different output and input price vector combinations, the resulting cost frontier need not be *convex* in output. To summarize, while every input price vector-specific cost frontier is *convex* in output, the cost frontier made up arbitrary points from these cost frontiers need not be *convex* when DMUs do not have control over input prices.[21]

However, the setup in our [NCOST] model assumes that DMUs not only control over on the mix and quantities of inputs used but also exercise any control over input prices[22] This is what one expects an ideal measure of scale elasticity to assign some value to each of the aforementioned aspects of

[21] The choice of preferred input prices is one important aspect of production planning process that is considered very important in determining international competitiveness, but is clearly overlooked in this conventional model. Since, this aspect does not lend itself well to direct measurement in this traditional model, many companies miss opportunities to bolster their performance by not locating themselves in places where some of important inputs are cheaply available. Another aspect, often overlooked in this model, is input cost savings due to bulk-buying at preferential lower prices.

[22] Using several different aspects of production planning process, this model indeed imputes a multifactor perspective in its scale elasticity estimates to track overall performance. The scale elasticity estimate obtained in this model often gives managers a convenient scorecard to answer question: 'How is a company really doing?' This model has theoretical advantage over traditional counterpart not only studying but also influencing scale elasticity behavior because use of elasticity estimates by companies will make a real difference in their subsequent decisions and priorities to improve performance.

production planning. Increasingly, companies are too discovering the competitive power of undertaking each of these aspects - or the dangers of not doing so. Now, we face a difficult trade off in the use of these two models when the observed cost-based technology set is not *convex*. The choice of using translog model, which has property of being flexible, is associated with imposition of *ad hoc* functional form without knowing the underlying relationship between inputs and outputs, which cast doubts on the results obtainable by this method. Furthermore, the imposition of normality restrictions on the parameters of translog production/cost function brings in the *convexity* postulate, which is definitely not a normal feature of factor-based production technology, as has been evident from empirical literature.[23]

In contrast, the choice of using [NCOST] model is associated with an imposition *convexity* postulate for cost-based technology,[24] but assuming away the same assumption for the factor-based technology set, which excludes important economic phenomena such as *indivisible* production activities, economies of scale and economies of specialization. So the question remains to be seen, which model to use for empirical valuation of scale elasticity, is largely dependent on the magnitude of relative cost of use of each model. In the light both empirical and theoretical arguments that favor dropping of convexity assumption for factor-based technology set in DEA, we believe, our proposed [NCOST] model does receive the favor.

# 5 Concluding Remarks

Investigation of scale elasticity for obtaining optimal scale of operations has significant bearings while recommending policy for restructuring any sector. So due care is warranted in this regard to ensure that the scale elasticity estimates should not be suffered from major shortcomings. We find in the parametric literature that the translog cost function has been very popular to estimating scale elasticity. We show here that first, the scale elasticity es-

[23]In fact, the empirical evidence suggests that non-convexities are surprisingly common feature of the underlying factor-based production technologies.

[24]A solid conceptual explanation for *convexity* postulate for cost-based technology set is of course required for establishing it to be a regular assumption in production economics literature. This excercise can be taken as a future research study, which is mentioned in the concluding remarks of this paper.

timates obtained from this precisely depend upon the functional forms of the inefficiency distribution, and second, these estimates are subject to question when the *convexity* postulate for the underlying factor-based technology set is not realistic. Similarly, the scale elasticity estimates obtained from traditional non-parametric method (Classical cost efficiency by Sueyoshi, 1997) can be argued to be misleading based on the grounds that first, this model is based on the *convex* factor-based technology set that rules out some economically important technological features such as *indivisibilities*, economies of scale and economies of specialization, and second, cost efficiency obtained from this model does not always satisfy the *monotonicity* property with respect to not only input but also input price. We suggest a new nonparametric method for the estimation of scale elasticity while pursuing further elasticity studies based on the premise that the elasticity estimates based on other methods can be illusory.

This study points to avenues for future research in two directions:

- Justifying further the *convex* cost-based technology set by providing a conceptual framework, which is essential to help readers better understand the situations under which our proposed new scale elasticity estimates will prove to be superior over the those in classical [COST] model.[25] Such a conceptual framework can be made by carefully describing the difference between our setup with that of classical model as well as describing the data generating processes for inputs, input price vectors and efficiency random variables, which, in conjunction with a *convex* factor-based technology set, can endogenously generate a *convex* cost-based technology set.

- Developing in terms of formulating a nonlinear DEA model by incorporating the relationship between input price and quantity as cost has a linkage with a production change (e.g., a bulk purchase).

[25]We have already demonstrated the superiority of the former over the latter in terms of exhibiting both *monotonicity* property, and not requiring the factor-based technology set to be convex.

# References


[1] Afriat, S. (1972) "Efficiency Estimation of Production Functions," *International Economic Review,* 13, 568-598.

[2] Aly, H. Y., R. Grabowski, C. Pasurka and N. Rangan. (1990) "Technical, Scale, and Allocative Efficiencies in U.S. Banking: An Empirical Investigation," *Review of Economics and Statistics,* 72, 211-218.

[3] Banker, R. D., A. Charnes and W. W. Cooper. (1984) "Models for the Estimation of Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science,* 30, 1078-1092.

[4] Banker, R. D. and R. M. Thrall. (1992) "Estimation of Returns to Scale Using Data Envelopment Analysis," *European Journal of Operational Research,* 62, 74-84.

[5] Banker, R. D., I. Bardhan and W. W. Cooper. (1996a) "A Note on Returns to Scale in DEA," *European Journal of Operational Research,* 88, 583-585.

[6] Banker, R. D., H. Chang and W. W. Cooper. (1996b) "Equivalence and Implementation of Alternative Methods for determining Returns to Scale in Data Envelopment Analysis," *European Journal of Operational Research,* 89, 473-481.

[7] Baumol, W. J., J. C. Panzar and R. D. Willig. (1988) *Contestable Markets and The Theory of Industry Structure,* New York: Harcourt Brace Jovanovich.

[8] Besley, J. (1989) "Ex ante Evaluation of Health Cares and the Provision for Ill-health," *Economic Journal,* 99, 132-146.

[9] Borland, J. and X. Yang. (1995) "Specialization, Product Development, Evolution of the Institution of the Firm, and Economic Growth," *Journal of Evolutionary Economics,* 5, 19-42.

[10] Bouhnik, S., B. Golany, U. Passy, S. T. Hackman and D. A. Vlatsa. (2001) "Lower Bound Restrictions on Intensities in Data Envelopment Analysis," *Journal of Productivity Analysis,* 16, 241-261.

[11] Boussofiane, A., R. G. Dyson and E. Thanassoulis. (1991) "Applied Data Envelopment Analysis," *European Journal of Operational Research,* 52, 1-15.

[12] Briec, W., K. Kerstens and P. Vanden Eeckaut (2004) "Non-convex Technologies and Cost Functions: Definitions, Duality and Nonparametric Tests of Convexity," *Journal of Economics,* 81, 155-192.

[13] Byrnes, P. and V. Valdmans. (1994) "Analyzing Technical and Allocative Efficiency of Hospitals." In: Charnes, Cooper, Lewin and Seiford (ed). *Data Envelopment Analysis: Theory, Methodology and Applications,* Norwell Mass: Kluwer Academic Publishers.

[14] Chamberlin, E. H. (1933) *The Theory of Monopolistic Competition: A Reorientation of the Theory of Value,* Cambridge: Harvard University Press.

[15] Chambers, R. G. (1988) *Applied Production Analysis: A Dual Approach,* Cambridge UK: Cambridge University Press.

[16] Charnes, A., W. W. Cooper and E. Rhodes. (1978) "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research,* 2: 429-444.

[17] Cherchye, L., T. Kuosmanen and T. Post. (2000a) "What is the Economic Meaning of FDH? A Reply to Thrall," *Journal of Productivity Analysis,* 13, 259-263.

[18] Cherchye, L., T. Kuosmanen and T. Post. (2000b) "Why Convexify? An Assessment of Convexity Axioms in DEA," Helsinki School of Economics and Business Administration, *Working Paper W-270.*

[19] Christensen, L. R. and W. H. Green. (1976) "Economies of Scale in Electric Power Generation," *Journal of Political Economy,* 84, 655-676.

[20] Christensen, L. R., D. W. Jorgenson and L. J. Lau. (1973) "Transcendental Logarithmic Production Frontiers," *Review of Economics and Statistics,* 55, 28-45.

[21] Coelli, T., D. S. P. Rao and G. E. Battese. (1998) *An Introduction to Efficiency and Productivity Analysis,* Boston: Kluwer Academic Publishers.

[22] Cooper, W. W., R. G. Thompson and R. M. Thrall. (1996) "Introduction: Extensions and New Developments in DEA," *Annals of Operations Research,* 66, 3-46.

[23] Cooper, W. W., L. M. Seiford and K. Tone. (1999). *Data Envelopment Analysis - A Comprehensive Text with Models, Applications, References and DEA-Solver Software,* Norwell Mass: Kluwer Academic Publishers.

[24] Deprins, D., L. Simar and H. Tulkens. (1984) "Measuring Labor Efficiency in Post Offices." In: Marchand, Pestieau and Tulkens (ed.): *The Performance of Public Enterprises: Concepts and Measurements,* Amsterdam: North-Holland.

[25] Färe, R., S. Grosskopf and C. A. K. Lovell. (1985) *The Measurement of Efficiency of Production,* Boston: Kluwer Nijhoff.

[26] Färe, R., S. Grosskopf and C. A. K. Lovell. (1986) "Scale Economies and Duality," *Journal of Economics,* 46, 175-182.

[27] Färe, R., S. Grosskopf and J. Logan. (1987) "The Comparative Efficiency of Western Coalfired Steam-electric Generating Plants: 1977-1979," *Engineering Costs and Production Economics,* 11, 21-30.

[28] Färe, R., S. Grosskopf and C. A. K. Lovell. (1988) "Scale Elasticity and Scale Efficiency," *Journal of Institutional and Theoretical Economics,* 144, 721-729.

[29] Färe, R., S. Grosskopf and C. A. K. Lovell. (1994) *Production Frontiers,* Cambridge: Cambridge University Press.

[30] Farrell, M. J. (1957) "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society* (Series A), 120, 253-290.

[31] Farrell, M. J. (1959) "Convexity Assumptions in Theory of Competitive Markets," *Journal of Political Economy,* 67, 377-391.

[32] Førsund, F. R. (1996) "On the Calculation of the Scale Elasticity in DEA Models," *Journal of Productivity Analysis,* 7, 283-302.

[33] Førsund, F. R. and L. Hjalmarsson. (2004) "Are All Scales Optimal in DEA? Theory and Empirical Evidence," *Journal of Productivity Analysis,* 21, 25-48

[34] Frisch, R. (1965) *Theory of Production,* Dordrecht: D. Reidel Publishing Company.

[35] Fukuyama, H. (2001) "Returns to Scale and Scale Elasticity in Farrell, Russell and Additive Models," *Journal of Productivity Analysis,* 16, 225-239.

[36] Gilsdorf, K. (1994) "Vertical Integration Efficiencies and Electric Utilities: A Cost Complementarity Perspective," *Quarterly Review of Economics and Finance,* 34, 261-282.

[37] Griliches, Z. and V. Ringstad. (1971) *Economies of Scale and the Form of Production Function,* Amsterdam: North-Holland Publishing Company.

[38] Grosskopf, S. and V. Valdmanis. (1987) "Measuring Hospital Performance: A Non-parametric Approach," *Journal of Health Economics,* 6, 89-107.

[39] Hammond, P. (1981) "Ex ante and Ex post Welfare Optimality Under Uncertainty," *Economica,* 48, 235-250.

[40] Hanoch, G. (1970) "Homotheticity in Joint Production," *Journal of Economic Theory,* 2, 423-426.

[41] Hanoch, G. and M. Rothschild. (1972) "Testing Assumptions in Production Theory: A Nonparametric Approach," *Journal of Political Economy,* 80, 256-275.

[42] Hasencamp, G. (1976) "A Study of Multiple-Output Production Functions," *Journal of Econometrics,* 4, 253-262.

[43] Hayashi, P. M., J. Y. Goo and W. M. C. Chamberlain. (1997) "Vertical Economies: The Case of U.S. Electric Utility Industry," *Southern Economic Journal,* 63, 710-725.

[44] Kaserman, D. L. and J. W. Mayo. "The Measurement of Vertical Economies and The Efficient Structure of the Electric Utility Industry," *Journal of Industrial Economics,* 34, 483-502.

[45] Kumbhakar, S. C., and C. A. K. Lovell. (2000) *Stochastic Frontier Analysis,* Cambridge: Cambridge University Press.

[46] Kuosmanen, T. (2003) "Duality Theory of Non-convex Technologies," *Journal of Productivity Analysis,* 20, 273-304.

[47] Löthgren, M. and Tambour, M. (1996) "Alternative Approaches to Estimate Returns to Scale in DEA-Models," *Stockholm School of Economics Working Paper No. 90.*

[48] McCall, J. J. (1967) Competitive Production for Constant Risk Utility Functions," *Review of Economic Studies,* 34, 417-420.

[49] Panzar, J. C. and R.D. Willig. (1977) "Economies of Scale in Multi-output Production," *Quarterly Journal of Economics,* XLI, 481-493.

[50] Rangan, N., R. Grabowski, H. Aly, and C. Pasurka. (1988) "The Technical Efficiency of US Banks," *Economics Letters,* 28, 169-176.

[51] Ray, S. C. (1999) "Measuring Scale Efficiency from a Translog Production Function," *Journal of Productivity Analysis,* 11, 183-194.

[52] Ray, S. C. and H. L. Kim. (1995) "Cost Efficiency in the U.S. Steel Industry: A Nonparametric Analysis Using Data Envelopment Analysis," *European Journal of Operational Research,* 80, 654-671.

[53] Robinson, J. (1933) *The Economics of Imperfect Competition,* London: Macmillan.

[54] Sahoo, B. K., P. K. J. Mohapatra and M. L. Trivedi. (1999) "A Comparative Application of Data Envelopment Analysis and Frontier Translog Production Function for Estimating Returns to Scale and Efficiencies," *International Journal of Systems Science,* 30, 379-394.

[55] Sandmo, A. (1971) "On the Theory of the Competitive Firm under Price Uncertainty," *American Economic Review,* 61, 65-73.

[56] Scarf, H. E. (1981a) "Production Sets with Indivisibilities Part I: Generalities," *Econometrica,* 49, 1-32.

[57] Scarf, H. E. (1981b) "Production Sets with Indivisibilities Part II: The Case of Two Activities," *Econometrica,* 49, 395-423.

[58] Scarf, H. E. (1986) "Neighborhood Systems for Production Sets with Indivisibilities," *Econometrica,* 54, 507-432.

[59] Scarf, H. E. (1994) "The Allocation of Resources in the Presence of Indivisibilities," *Journal of Economic Perspectives,* 8, 111-128.

[60] Shephard, R. W. (1970) *Theory of Cost and Production Functions,* Princeton: Princeton University Press.

[61] Shi, H. and X. Yang. (1995) "A New Theory of Industrilization," *Journal of Comparative Economics,* 20, 171-189.

[62] Starrett, D. A. (1977) "Measuring Returns to Scale in the Aggregate, and the Scale Effect of Public Goods," *Econometrica,* 45, 1439-1455.

[63] Stevenson, R. E. (1980) "Likelihood Functions for Generalized Stochastic Frontier Estimation," *Journal of Econometrics,* 13, 57-66.

[64] Sueyoshi, T. (1997) "Measuring Efficiencies and Returns to Scale of Nippon Telegraph & Telephone in Production and Cost Analyses," *Management Science,* 43, 779-796.

[65] Sueyoshi, T. (1999) "DEA Duality on Returns to Scale (RTS) in Production and Cost Analyses: An Occurrence of Multiple Solutions and Differences Between Production-based and Cost-based RTS Estimates," *Management Science,* 45, 1593-1608.

[66] Thrall, R. M. (1999) "What is the Economic Meaning of FDH?," *Journal of Productivity Analysis,* 11, 243-250.

[67] Tone, K. (1996) "A Simple Characterization of Returns to Scale in DEA," *Journal of Operations Research Society of Japan*, 39, 604-613.

[68] Tone, K. (2002) "A Strange Case of Cost and Allocative Efficiencies in DEA," *Journal of the Operational Research Society*, 53, 1225-1231.

[69] Tone, K. and B. K. Sahoo. (2003) "Scale, Indivisibilities and Production Function in Data Envelopment Analysis," *International Journal of Production Economics*, 84, 165-192.

[70] Tone, K. and B. K. Sahoo. (2004) "Degree of Scale Economies and Congestion: A Unified DEA Approach," *European Journal of Operational Research*, 158, 755-772.

[71] Tone, K. and B. K. Sahoo. (2004a) "Cost Efficiency and Returns to Scale in Life Insurance Corporation of India Using Data Envelopment Analysis," *Socio-Economic Planning Sciences* (forthcoming).

[72] Tone, K. and B. K. Sahoo. (2004b) "Re-examining Scale Elasticity in DEA," *Annals of Operations Research* (forthcoming).

[73] Tulkens, H. (1993) "On FDH Analysis: Some Methodological Issues and Applications to Retail Banking, Courts and Urban Transit," *Journal of Productivity Analysis,* 4, 183-210.

[74] Tulkens, H. and P. Vanden Eeckaut. (1995) "Non-parametric Efficiency, Progress and Regress Measures for Panel Data: Methodological Aspects," *European Journal of Operational Research,* 80, 474-499.

[75] Varian, H. R. (1984) "The Nonparametric Approach to Production Analysis," *Econometrica,* 52, 279-297.

[76] Varian, H. R. (1992) *Microeconomic Analysis,* New York: Norton.

[77] Yang, X. (1994) "Endogeneous vs. Exogeneous Comparative Advantage and Economics of Specialization vs. Economies of Scale," *Journal of Economics,* 60,29-54.

[78] Yang, X. and Y. K. Ng. (1993) *Specialization and Economic Organization*. Amsterdam: Elsevier Science.

[79] Yang, X. and R. Rice. (1994) "An Equilibrium Model Endogenizing the Emergence of a Dual Structure Between the Urban and Rural Sectors," *Journal of Urban Economics,* 35, 346-368.