Estimation and Inference in Nonparametric Frontier Models: Recent Developments and Perspectives

Léopold Simar

Université Catholique de Louvain
Belgium
leopold.simar@uclouvain.be

Paul W. Wilson

Clemson University
USA
pww@clemson.edu

Full text available at: http://dx.doi.org/10.1561/0800000020
Editorial Scope

Foundations and Trends® in Econometrics will publish survey and tutorial articles in the following topics:

- Econometric Models:
  - Identification
  - Model Choice and Specification Analysis
  - Non-linear Regression Models
- Simultaneous Equation Models
- Estimation Frameworks
- Biased Estimation
- Computational Problems
- Microeconometrics
- Treatment Modeling
- Discrete Choice Modeling
- Models for Count Data
- Duration Models
- Limited Dependent Variables
- Panel Data
- Time Series Analysis:
  - Dynamic Specification
  - Inference and Causality
  - Continuous Time Stochastic Models
  - Modeling Non-linear Time Series
  - Unit Roots
  - Cointegration
- Latent Variable Models
- Qualitative Response Models
- Hypothesis Testing
- Econometric Theory:
  - Interactions-based Models
  - Duration Models
- Financial Econometrics
- Measurement Error in Survey Data
- Productivity Measurement and Analysis
- Semiparametric and Nonparametric Estimation
- Bootstrap Methods
- Nonstationary Time Series
- Robust Estimation

Information for Librarians

Foundations and Trends® in Econometrics, 2011, Volume 5, 4 issues. ISSN paper version 1551-3076. ISSN online version 1551-3084. Also available as a combined paper and online subscription.
Estimation and Inference in Nonparametric Frontier Models: Recent Developments and Perspectives

Léopold Simar\(^1\) and Paul W. Wilson\(^2\)

\(^1\) Institut de Statistique, Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, Voie du Roman Pays 20, B 1348 Louvain-la-Neuve, Belgium, leopold.simar@uclouvain.be

\(^2\) Department of Economics and School of Computing, 228 Sirrine Hall, Clemson University, Clemson, South Carolina, 29634–1309, USA, pww@clemson.edu

Abstract

Nonparametric estimators are widely used to estimate the productive efficiency of firms and other organizations, but often without any attempt to make statistical inference. Recent work has provided statistical properties of these estimators as well as methods for making statistical inference, and a link between frontier estimation and extreme value theory has been established. New estimators that avoid many of the problems inherent with traditional efficiency estimators have also been developed; these new estimators are robust with respect to outliers and avoid the well-known curse of dimensionality. Statistical properties, including asymptotic distributions, of the new estimators...
have been uncovered. Finally, several approaches exist for introducing environmental variables into production models; both two-stage approaches, in which estimated efficiencies are regressed on environmental variables, and conditional efficiency measures, as well as the underlying assumptions required for either approach, are examined.
## Contents

1 Nonparametric Statistical Models of Production: Combining Economics and Statistics 1
  1.1 Economic Considerations 1
  1.2 Statistical Considerations 9

2 The Nonparametric Envelopment Estimators 15
  2.1 The Statistical Model 15
  2.2 The FDH Estimator 17
  2.3 The DEA Estimators 23
  2.4 An Alternative Probabilistic Formulation of the DGP 26
  2.5 Properties of FDH and DEA Estimators 32

3 Bootstrap Inference Using DEA and FDH Estimators 45
  3.1 General Idea 45
  3.2 Confidence Intervals for Efficiency of a Particular Point 59

4 Robust Order-\(m\) Estimators 67
  4.1 The Need for Robustness 67
  4.2 Definitions and Basic Ideas 69
  4.3 Nonparametric Order-\(m\) Estimators 85
5 Robust Order-α Estimators 95
   5.1 Definition and Basic Ideas 95
   5.2 Nonparametric Order-α Estimators 104
   5.3 Statistical Properties 107
   5.4 Empirical Examples 111

6 Outlier Detection 115

7 Explaining Inefficiency 123
   7.1 Introducing Environmental Variables 123
   7.2 Two-Stage Regression Approach 125
   7.3 Conditional Efficiency Measures 133

8 Unanswered Questions, Promising Ideas 141

Acknowledgments 147

References 149
1

Nonparametric Statistical Models of Production:
Combining Economics and Statistics

The economic theory underlying analysis of efficiency in production dates at least to the work of Koopmans (1951), Debreu (1951), and Farrell (1957). Farrell made the first attempt to estimate efficiency from a set of observed production units, but the statistical properties of his estimator were only considered much later.

The discussion that follows introduces basic concepts and notation; Economic Considerations introduces an economic framework, to which Statistical Considerations adds a statistical paradigm.

1.1 Economic Considerations

Producers transform inputs into outputs; for example, in manufacturing, inputs typically include labor, capital, energy, materials, and perhaps other things, while outputs are the products produced. There may be one, several, perhaps many different products that are produced. Of course, production is constrained by what is possible or feasible. Let \( x \in \mathbb{R}^p_+ \) and \( y \in \mathbb{R}^q_+ \) denote vectors of input and output quantities, respectively, and let

\[
P = \{ (x, y) \mid x \text{ can produce } y \} \tag{1.1}
\]
denote the set of feasible combinations of inputs and outputs, i.e., the production set. Any output quantities $y$ can be produced using input quantities $x$ if and only if $(x, y) \in \mathcal{P}$. However, the points in $\mathcal{P}$ are not equally desirable.

The following three assumptions regarding $\mathcal{P}$ are standard in microeconomic theory of the firm; see, for example, Shephard (1970) and Färe (1988).

**Assumption 1.1.** $\mathcal{P}$ is closed.

**Assumption 1.2.** All production requires use of some inputs: $(x, y) \notin \mathcal{P}$ if $x = 0$ and $y \geq 0$, $y \neq 0$.

**Assumption 1.3.** Both inputs and outputs are freely disposable: if $(x, y) \in \mathcal{P}$, then for any $(x', y')$ such that $x' \geq x$ and $y' \leq y$, $(x', y') \in \mathcal{P}$.

Assumption 1.1 ensures that the boundary of $\mathcal{P}$ is included in $\mathcal{P}$. Assumption 1.2 means that there are no “free lunches.” The free disposability assumption is sometimes called strong disposability and is equivalent to an assumption of monotonicity of the technology. This property also characterizes the technical possibility of wasting resources (i.e., the possibility of producing less with more resources).

For purposes of efficiency measurement, the upper boundary of $\mathcal{P}$ is relevant. The efficient subset of points in $\mathcal{P}$ is the upper boundary (frontier) of $\mathcal{P}$, i.e., the locus of optimal production plans (e.g., minimal achievable input level for a given output, or maximal achievable output given the level of the inputs). The upper boundary of $\mathcal{P}$,

$$\mathcal{P}^\partial = \{(x, y) \in \mathcal{P} \mid (\gamma^{-1} x, \gamma y) \notin \mathcal{P} \forall \gamma \in (1, \infty)\}$$  \hspace{1cm} (1.2)

is sometimes referred to as the technology or the production frontier, and is given by the intersection of $\mathcal{P}$ and the closure of its complement.

---

1 Throughout, inequalities involving vectors are assumed to hold element by element; e.g., $a \leq b$ denotes $a_j \leq b_j$ for each $j = 1, \ldots, k$, where $k$ is the length of $a$ and $b$. 
Firms that are technically inefficient operate at points in the interior of \( \mathcal{P} \), while those that are technically efficient operate somewhere along the technology defined by \( \mathcal{P}^\partial \).

Various features of the production set \( \mathcal{P} \) and its frontier \( \mathcal{P}^\partial \) are often of interest to applied researchers. One such feature is \textit{returns to scale}. Strictly speaking, returns to scale is a feature of the frontier, \( \mathcal{P}^\partial \), but it is common to ascribe such features to the set \( \mathcal{P} \). There are several possibilities.

**Definition 1.1.** The frontier \( \mathcal{P}^\partial \) displays globally constant returns to scale (CRS) if and only if \((\alpha x, \alpha y) \in \mathcal{P} \land (x, y) \in \mathcal{P} \) and \( \alpha \in [0, \infty) \).

**Definition 1.2.** The frontier \( \mathcal{P}^\partial \) displays globally variable returns to scale (VRS) if and only if for any \((x, y) \in \mathcal{P} \), there exist constants \( a(x, y) \in \mathbb{R}_{+}^1, b(x, y) \in \mathbb{R}_{++}^1 \) such that \( a(x, y) \leq b(x, y) \) and \((\alpha x, \alpha y) \in \mathcal{P} \land \alpha \in [a(x, y), b(x, y)] \).

**Definition 1.3.** The frontier \( \mathcal{P}^\partial \) displays globally nonincreasing returns to scale (NIRS) if and only if \((\alpha x, \alpha y) \in \mathcal{P} \land (x, y) \in \mathcal{P} \) and \( \alpha \in [0, 1] \).

Note that Definition 1.2 encompasses Definitions 1.1 and 1.3; i.e., CRS and NIRS are special cases of VRS. In the same way, Definition 1.3 encompasses Definition 1.1 in that CRS is a special case of NIRS. In other words, assuming either CRS or NIRS is more restrictive than assuming VRS; assuming CRS is more restrictive than assuming NIRS.

The production set \( \mathcal{P} \) can also be described by its sections or level sets. For instance, the input requirement set for some \( y \in \mathbb{R}_+^q \) is given by

\[
\mathcal{X}(y) = \{x \in \mathbb{R}_+^p \mid (x, y) \in \mathcal{P}\},
\]

i.e., the set of all input vectors \( x \) that can produce output vector \( y \). The boundary of this set, i.e., the (input-oriented) efficiency boundary \( \mathcal{X}^\partial(y) \), is defined for a given \( y \in \mathbb{R}_+^q \) by

\[
\mathcal{X}^\partial(y) = \{x \mid x \in \mathcal{X}(y), \theta x \notin \mathcal{X}(y), \forall \theta \in (0, 1)\}.
\]
Alternatively, the output feasibility set for some $x \in \mathbb{R}^p_+$ is defined by

$$\mathcal{Y}(x) = \{ y \in \mathbb{R}^q_+ \mid (x, y) \in \mathcal{P} \},$$

(1.5)

which gives the set of all output vectors $y$ than can be produced with given input quantities $x$. The (output-oriented) efficiency boundary $\mathcal{Y}^\partial(x)$ is defined, for a given $x \in \mathbb{R}^p_+$, as

$$\mathcal{Y}^\partial(x) = \{ y \mid y \in \mathcal{Y}(x), \lambda y \notin \mathcal{Y}(x), \forall \lambda > 1 \}.$$  

(1.6)

Then the production set $\mathcal{P}$ corresponds to the union of all sets $X(y)$ over all $y \in \mathbb{R}^q_+$, or to the union of all sets $\mathcal{Y}(x)$ over all $x \in \mathbb{R}^p_+$.

The Debreu–Farrell input measure of technical efficiency for a given point $(x, y) \in \mathbb{R}^{p+q}_+$ is given by

$$\theta(x, y \mid \mathcal{P}) = \inf \{ \theta \mid \theta x \in \mathcal{X}(y) \} = \inf \{ \theta \mid (\theta x, y) \in \mathcal{P} \}. $$

(1.7)

Note that this measure is defined for some points in $\mathbb{R}^{p+q}_+$ not necessarily in $\mathcal{P}$ (i.e., points for which a solution exists in (1.7)). Given an output level $y$, and an input mix (a direction) given by the vector $x$, the corresponding efficient level of input is given by

$$x^\partial(y) = \theta(x, y \mid \mathcal{P}) x,$$

(1.8)

which is the projection of $(x, y)$ onto the efficient boundary $\mathcal{P}^\partial$, along the ray $x$ and orthogonal to the vector $y$.

Figure 1.1 illustrates a point $(x_0, y_0) \in \mathcal{P}$ for $p = q = 1$. The level set $x^\partial(y)$ defined in (1.8) contains just one point in Figure 1.1 in terms of the labels on the horizontal axis, $\theta(x_0, y_0 \mid \mathcal{P}) = x^\partial(y_0)/x_0 < 1$.

In general, for $(x, y) \in \mathcal{P}$, $\theta(x, y \mid \mathcal{P})$ gives the feasible proportionate reduction of inputs that a unit located at $(x, y)$ could undertake to become technically efficient. By construction, for all $(x, y) \in \mathcal{P}$, $\theta(x, y \mid \mathcal{P}) \in (0, 1]$; $(x, y)$ is technically efficient if and only if $\theta(x, y \mid \mathcal{P}) = 1$. This measure is the reciprocal of the Shephard (1970) input distance function.

Similarly, in the output direction, the Debreu–Farrell output measure of technical efficiency is given by

$$\lambda(x, y \mid \mathcal{P}) = \sup \{ \lambda \mid \lambda y \in \mathcal{Y}(x) \} = \sup \{ \lambda \mid (x, \lambda y) \in \mathcal{P} \} $$

(1.9)
1.1 Economic Considerations

for \((x, y) \in \mathbb{R}^{p+q}\). Analogous to the input-oriented case described above, \(\lambda(x, y \mid \mathcal{P})\) gives the feasible proportionate increase in outputs for a unit operating at \((x, y) \in \mathcal{P}\) that would achieve technical efficiency. By construction, for all \((x, y) \in \mathcal{P}\), \(\lambda(x, y \mid \mathcal{P}) \in [1, \infty)\) and \((x, y)\) is technically efficient if and only if \(\lambda(x, y \mid \mathcal{P}) = 1\).

The output efficiency measure \(\lambda(x, y \mid \mathcal{P})\) is the reciprocal of the output distance function. The efficient level of output, for the input level \(x\) and for the direction of the output vector determined by \(y\), is given by

\[
y^\beta(x) = \lambda(x, y \mid \mathcal{P})y.
\] (1.10)

Figure 1.2 illustrates the same point \((x_0, y_0) \in \mathcal{P}\) shown in Figure 1.1. Here, the set \(y^\beta(x_0)\) defined by (1.10) also contains a single point, and in terms of the labels on the vertical axis in Figure 1.2, \(\lambda(x_0, y_0 \mid \mathcal{P}) = y^\beta(x_0)/y_0 > 1\).

Efficiency can be measured in other directions, although care should be taken to avoid having efficiency measures depend on units of measurement for inputs or outputs. For example, a hyperbolic measure of efficiency is given by

\[
\gamma(x, y \mid \mathcal{P}) = \sup \{\gamma \mid (\gamma^{-1}x, \gamma y) \in \mathcal{P}\}
\] (1.11)
for \((x, y) \in \mathbb{R}_+^{p+q}\). This hyperbolic measure of efficiency gives the simultaneous proportionate, feasible reduction in input levels and the proportionate, feasible increase in output levels for a unit operating at \((x, y) \in \mathcal{P}\) that would result in technical efficiency, and is the reciprocal of the hyperbolic graph measure of efficiency defined by [Färe et al. (1985)].

In terms of the illustration in Figure 1.3, the firm operating at the point \((x_0, y_0) \in \mathcal{P}\) can become technically efficient by moving along the curved (hyperbolic) path from \((x_0, y_0)\) to \((x_0^\gamma(x_0, y_0), y_0^\gamma(x_0, y_0))\). By construction, \(\gamma(x, y | \mathcal{P}) \in [1, \infty)\) for all \((x, y) \in \mathcal{P}\); in addition, \((x, y)\) is technically efficient if and only if \(\gamma(x, y | \mathcal{P}) = 1\). For \((x, y)\) in the interior of \(\mathcal{P}\), the corresponding hyperbolic-efficient levels of inputs are outputs that are given by

\[
\left(x_0^\gamma(x, y), y_0^\gamma(x, y)\right) = \left(\gamma(x, y | \mathcal{P})^{-1}x, \gamma(x, y | \mathcal{P})y\right).
\]

Alternatively, Chambers et al. (1996) introduced the directional efficiency measure defined by

\[
\delta(x, y | u, v, \mathcal{P}) = \sup\{\delta \mid (x - \delta u, y + \delta v) \in \mathcal{P}\},
\]

where \(u\) and \(v\) are direction vectors with \(u \in \mathbb{R}_+^{p}\), \(v \in \mathbb{R}_+^{q}\), and \([u^\top \ v^\top] \neq 0\). This distance function projects a point \((x, y)\) onto the
1.1 Economic Considerations

Fig. 1.3 Hyperbolic efficiency measure.

frontier $\mathcal{P}^\partial$ in the direction $(-u,v)$, with

$$\left(x_\partial^\partial(x,y \mid u,v), y_\partial^\partial(x,y \mid u,v)\right) = (x - \delta(x,y \mid u,v,\mathcal{P})u, y + \delta(x,y \mid u,v,\mathcal{P})v)$$

(1.14)

giving the directionally-efficient (in the direction $(u,v)$) levels of inputs and outputs.

The directional distance function is illustrated in Figure 1.4. Setting $u = x_0, v = y_0$, the firm operating at $(x_0,y_0)$ becomes technically efficient when it moves to $(x_\partial^\partial(x_0,y_0), y_\partial^\partial(x_0,y_0))$ defined by (1.14) (the notation indicating dependence on direction vectors is suppressed here and in Figure 1.4 to conserve space). By construction, $\delta(x,y \mid u,v,\mathcal{P}) \in [0,\infty)$ for all $(x,y) \in \mathcal{P}$; a point $(x,y) \in \mathcal{P}$ is technically efficient if and only if $\delta(x,y \mid u,v,\mathcal{P}) = 0$.

Färe et al. (2008, p. 534) state that the directional distance function is independent of units of measurement in the sense that

$$\delta(\alpha_x \circ x, \alpha_y \circ y \mid \alpha_x \circ u, \alpha_y \circ v, \mathcal{P}) = \delta(x,y \mid u,v,\mathcal{P}),$$

(1.15)

where $\alpha_x \in \mathbb{R}^{p}_{++}, \alpha_y \in \mathbb{R}^{q}_{++}$, and $\circ$ denotes the Hadamard product. However, while (1.15) is true, it also indicates that if units of

---

2 The Hadamard product of two arrays $A = [a_{ij}]$ and $B = [b_{ij}]$ with the same dimensions is given by the array $C = [c_{ij}]$ having the same dimensions as $A$ and $B$ where $c_{ij} = a_{ij}b_{ij}$;
measurement for inputs or outputs are changed, the corresponding
direction vector must be rescaled to avoid changing the value of the
directional distance function. Instead of being homogeneous of degree
zero with respect to inputs and outputs, the directional distance func-
tion is only homogeneous of degree zero with respect to inputs, outputs,
and direction vectors.

This feature of the directional distance function makes the range of
reasonable choices for the direction vectors less broad than has been
suggested in the literature. For example, Färe et al. (2008, p. 533)
note that the direction vectors should be specified in the same units
as the inputs and outputs, but then go on to suggest choosing \( u = 1 \),
\( v = 1 \) or to optimize \( u \) and \( v \) to minimize distance to the (estimated)
frontier. But, if one specifies \( u = 1 \), \( v = 1 \), and then changes the units
of measurement, this will require re-scaling also \( u \) and \( v \) so that they no
longer equal unity in order to avoid changing the value of the distance
function. Hence the choice of \((1,1)\) for \((u,v)\) is arbitrary, and therefore
rather meaningless. Moreover, if the direction vectors are optimized to

\[ e.g., \text{ see } \] Marcus and Kahn (1959), Marcus and Thompson (1963), and Johnson (1974a, 1974b).
1.2 Statistical Considerations

minimize distance to the estimated frontier, then the results will be sensitive to the units of measurement that are used.

It is easy to show that if \( u = x \) and \( v = 0 \), then

\[
\delta(x, y | u = x, v = 0, \mathcal{P}) = 1 - \theta(x, y | \mathcal{P}).
\] (1.16)

Similarly, if \( u = 0 \) and \( v = y \), then

\[
\delta(x, y | u = 0, v = y, \mathcal{P}) = \lambda(x, y | \mathcal{P}) - 1.
\] (1.17)

A common choice, when \( x \in \mathbb{R}^p_{++} \) and \( y \in \mathbb{R}^q_{++} \), is to set \( u = x \) and \( v = y \). One can also set the direction vectors equal to the sample means of inputs and outputs in order to use a common direction for all observations. For additional properties of the directional distance function, see Chambers et al. (1996).

Both the hyperbolic and the directional measures are measures of technical efficiency, as are the input- and output-oriented measures discussed above. Technical efficiency refers to what is possible, but as suggested earlier, not everything that is possible is desirable. Firms may want to maximize profits, which requires considering the prices of inputs and outputs in addition to their quantities. In the case of government provision of goods and services, or in regulated industries, producers’ goals may be cost minimization or perhaps revenue maximization.

A variety of assumptions on \( \mathcal{P} \) are found in the literature (e.g., free disposability, convexity, etc.; see Shephard, 1970 for examples). The assumptions about \( \mathcal{P} \) determine the appropriate estimator that should be used to estimate \( \mathcal{P}^\partial, \theta(x, y | \mathcal{P}), \lambda(x, y | \mathcal{P}), \gamma(x, y), \) or \( \delta(x, y) \). This issue will be discussed next.

1.2 Statistical Considerations

In real-world research problems, the attainable set \( \mathcal{P} \), as well as \( \mathcal{X}(y), \mathcal{X}^\partial(y), \mathcal{Y}(x), \) and \( \mathcal{Y}^\partial(x) \) are unknown to the analyst. Consequently, the efficiency scores \( \theta(x, y | \mathcal{P}), \lambda(x, y | \mathcal{P}), \gamma(x, y | \mathcal{P}), \) and \( \delta(x, y | u, v, \mathcal{P}) \) corresponding to a particular unit operating at \((x, y) \in \mathcal{P}\) are also unknown.
In ordinary settings, the only information available to the researcher is a sample

\[ S_n = \{(X_i, Y_i), \ i = 1, \ldots, n\} \quad (1.18) \]

of observations on input and output levels for a set of production units engaged in the activity of interest. The statistical paradigm raises the following question that must be answered: what can be learned by observing \( S_n \)? In other words, how can the information in \( S_n \) be used to estimate \( \theta(x, y | \mathcal{P}), \lambda(x, y | \mathcal{P}), \gamma(x, y | \mathcal{P}), \delta(x, y | u, v, \mathcal{P}) \), or other things of interest?

Answering these questions involves much more than reading the data in \( S_n \) into a computer program and pushing some buttons on the keyboard to solve some linear programs. A relevant question is, “what is learned from an estimate of \( \theta(x, y | \mathcal{P}), \lambda(x, y | \mathcal{P}) \), or other numbers computed from \( S_n \)?” The answer is clear and certain: almost nothing. One might learn, for example, that unit A uses less input quantities while producing greater output quantities than unit B, but little else can be learned from estimates of the efficiency measures introduced above without doing some additional work.

Before anything can be learned about \( \theta(x, y | \mathcal{P}), \lambda(x, y | \mathcal{P}), \gamma(x, y | \mathcal{P}), \delta(x, y | u, v, \mathcal{P}) \), or by extension about \( \mathcal{P} \) and its various features, one must use methods of statistical analysis to understand the properties of whatever estimators have been used to obtain estimates of the things of interest. This raises the following questions: Is the estimator consistent? Is the estimator biased? If the estimator is biased, does the bias disappear as the sample size tends toward infinity? If the estimator is biased, can the bias be corrected, and at what cost; i.e., does correcting the bias introduce too much noise? Can confidence intervals for the values of interest be estimated, and if so, how? How might one test interesting hypotheses about the production process? Notions of statistical consistency, etc. are discussed below.

\(^3\) Following standard notation, random variables are denoted by upper-case letters, and realizations of random variables and other nonstochastic quantities by lower-case letters.

\(^4\) Note that an estimator is a random variable, while an estimate is a realization of an estimator (random variable). An estimator can take perhaps infinitely many values with different probabilities, while an estimate is merely a known, nonrandom value.
Before these questions can be answered, a statistical model must be defined; without a statistical model, one cannot know what is estimated. Statistical models consist of two parts: (i) a probability model, which in the present case includes assumptions on the production set $\mathcal{P}$ and the distribution of input and output vectors $(x, y)$ over $\mathcal{P}$; and (ii) a sampling model describing how data are obtained from the probability model. The statistical model provides a theoretical description of the mechanism that yields the data in the sample $S_n$, and is sometimes called the data-generating process (DGP). In typical research settings, the task is to use the data in $S_n$ to learn something about the features of the DGP.

In cases where a group of productive units are observed at the same point in time, i.e., where cross-sectional data are observed, it is convenient and often reasonable to assume the sampling process involves independent draws from the probability distribution defined in the DGP’s probability model. With regard to the probability model, one must attempt reasonable assumptions. Of course, there are trade-offs here; the assumptions on the probability model must be strong enough to permit estimation using estimators that have useful properties, and to allow those properties to be deduced, yet not so strong as to impose conditions on the DGP that do not reflect reality. The goal should be, in all cases, to make minimal, flexible assumptions in order to let the data reveal as much as possible about the underlying DGP, as opposed to making strong, untested assumptions that might influence the results of estimation and inference in perhaps large and misleading ways. The assumptions defining the statistical model are of crucial importance, since any inference that might be made will typically be valid only if the assumptions are in fact true.

The above considerations apply equally to parametric as well as nonparametric approaches to estimation and inference. One can imagine a spectrum of estimation approaches, ranging from fully parametric (most restrictive) to fully nonparametric (least restrictive). Fully parametric estimation strategies necessarily involve stronger assumptions on the probability model, which is completely specified in terms of a specific probability distribution function, structural equations, etc. Semi-parametric strategies are less restrictive; in these approaches, some (but
not all) features of the probability model are left unspecified (for example, in a regression setting one might specify parametric forms for some, but not all, of the moments of a distribution function in the probability model). Fully nonparametric approaches assume no parametric forms for any features of the probability model. Instead, only (relatively) mild assumptions on broad features of the probability distribution are made, usually involving assumptions of various types of continuity, degrees of smoothness, etc.

With fully nonparametric approaches to efficiency estimation, no specific analytical function describing the frontier is assumed. In addition, possibly restrictive assumptions on the stochastic part of the model, describing the probabilistic behavior of the observations in the sample with respect to the efficient boundary of $\mathcal{P}$, are also avoided. There is, however, a cost for this flexibility; in particular, all observed input–output pairs $(X_i, Y_i)$ are assumed to be technically attainable; observations $(X_i, Y_i)$ on input, output vectors are assumed to be drawn randomly and independently from a population of firms whose input–output vectors are distributed on the attainable set $\mathcal{P}$ according to some unknown probability law described by a probability density function $f(x, y)$ or the corresponding distribution function $F(x, y) = \Pr(X \leq x, Y \leq y)$, with

$$\Pr(X_i, Y_i) \in \mathcal{P}) = 1. \quad (1.19)$$

By contrast, fully parametric approaches to efficiency estimation developed by Aigner et al. (1977), Meeusen and van den Broeck (1977), Battese and Corra (1977), Jondrow et al. (1982), and others allow some observations to lie outside the production set $\mathcal{P}$ by incorporating a (two-sided) stochastic term reflecting measurement error or other noise in addition to a (one-sided) stochastic term reflecting inefficiency. Introduction of the stochastic noise term, however, incurs a cost: some parametric structure is required for such models to be identified, which in turn requires assumptions that may or may not be supported by data. In addition, such models typically allow for only a single response variable, i.e., a single output variable in a production framework; researchers typically work in a cost framework when there are multiple outputs, but this in turn requires data on input-prices.
1.2 Statistical Considerations

To the extent that the fully parametric approach allows for measurement error, it only does so for the response variable, and not for any of the explanatory variables. Perhaps most problematic, writing the model in a regression framework introduces issues of causality and exogeneity that do not arise in the fully nonparametric approach, which more closely resembles an exercise in density estimation.

The fully parametric approaches are often called stochastic frontier analysis, while the fully nonparametric approaches are frequently called deterministic frontier analysis. This terminology is unfortunate since it is misleading and has created a good bit of confusion in the literature. In both approaches, there is only one frontier, and it is fixed, not stochastic. In both approaches, the location of the frontier is unknown, and this is what necessitates estimation and gives rise to uncertainty. In both approaches, the distance from a given observation \((X_i, Y_i)\) to the frontier (in any direction) is unknown, and must be estimated.

The most popular nonparametric efficiency estimators are based on the idea of estimating the attainable set \(\mathcal{P}\) by the smallest set \(\hat{\mathcal{P}}\) within some class of sets that envelop the observed data. Depending on assumptions made on \(\mathcal{P}\), this idea leads to the Free Disposal Hull (FDH) estimator of Deprins et al. (1984), which relies only on an assumption of free disposability, and the Data Envelopment Analysis (DEA) estimators which incorporate additional assumptions. Farrell (1957) was the first to use a DEA estimator in an empirical application, but the idea remained obscure until it was popularized by Charnes et al. (1978) and Banker et al. (1984). Charnes et al. estimated \(\mathcal{P}\) by the convex cone of the FDH estimator of \(\mathcal{P}\), thus imposing an assumption of constant returns to scale, while Banker et al. used the convex hull of the FDH estimator of \(\mathcal{P}\), thereby allowing for VRS.

The primary advantage of nonparametric models and estimators lies in their great flexibility (as opposed to parametric, deterministic frontier models). In addition, the nonparametric estimators are easy to compute, and today most of their statistical properties are well-established. As will be discussed below, inference is available using bootstrap methods.

The main drawbacks of the fully nonparametric DEA and FDH estimators is that they are very sensitive to outliers and extreme values,
Nonparametric Statistical Models of Production

and that noisy data are not allowed. Fortunately, robust alternatives to DEA and FDH estimators are available for use in fully nonparametric models; these alternative approaches will be described later. Also, as discussed below in the last section of this survey, “stochastic” versions of DEA and FDH estimators are the object of current research.

It should be noted that allowing for noise in frontier models presents difficult problems even in a fully parametric framework where one can rely on the assumed parametric structure. In fully parametric models where the DGP involves a one-sided error process reflecting inefficiency and a two-sided error process reflecting statistical noise, numerical identification of the statistical model’s features is sometimes highly problematic even with large (but finite) samples; see Ritter and Simar (1997) for examples.

Apart from the issue of numerical identification, fully parametric frontier models that incorporate a noise term present other difficulties. Efficiency estimates in these models are based on residual terms that are unidentified. Researchers instead base efficiency estimates on an expectation, conditional on a composite residual; estimating an expected inefficiency is rather different from estimating actual inefficiency. An additional problem arises from the fact that, even if the fully parametric, stochastic frontier model is correctly specified, there is typically a nontrivial probability of drawing samples with the “wrong” skewness (e.g., when estimating cost functions, one would expect composite residuals with right-skewness, but it is certainly possible to draw finite samples with left-skewness — the probability of doing so depends on the sample size and the mean of the composite errors). Since there are apparently no published studies, and also apparently no working papers in circulation, where researchers report composite residuals with the “wrong” skewness when fully parametric, stochastic frontier models are estimated, it appears that estimates are sometimes, perhaps often, conditioned (i) on either drawing observations until the desired skewness is obtained or (ii) on model specifications that result in the desired skewness. This raises formidable questions for inference; see Simar and Wilson (2010) for discussion.

Full text available at: http://dx.doi.org/10.1561/0800000020


References


References


References


Full text available at: http://dx.doi.org/10.1561/0800000020
References


References


