Tutorial - Stochastic DEA, Wuhan 2016.

O. B. Olesen

version 22.5.2016
Three different directions for extending DEA:

- In Olesen and Peteresen (2016) the notion of Stochastic DEA is at focus: Three different directions for extending DEA are reviewed:

  1. The first direction extends DEA to be able to handle estimated deviations from frontier practice as random deviations, e.g. Banker (1993) Maximum likelihood, consistency and DEA: A statistical foundation

  2. The second direction extends DEA to be able to handle random noise in the form of either measurement errors or specification errors. An example of 1) and 2) is Banker and Maindiratta (1992) Maximum likelihood estimation of monotone and concave production frontiers

  3. The third direction extends DEA to be able to regard or conceive the PPS as a random PPS, based on the random variation in data. An example of 2) and 3) is Cooper et al. (1988) Chance constrained programming formulations for stochastic characterizations of efficiency and dominance in DEA
Motivation (1):

- DEA was originally developed within the *Management Science (MS)* framework,
- **Distributional characteristics** of the deviation of inefficient DMUs from the best practice frontier **not considered**
- Without any specification of random inefficiency and/or noise,
  - i.e. without consideration for i) measurement errors, ii) sample noise and iii) specification errors.
- Any given observed set of DMUs was **NOT** seen as the result of some sampling process from a larger population, hence **sampling error was not "an issue"**
Motivation (1):

- When papers on Stochastic DEA began to appear, they took off in two very different directions.
- One approach, initiated by Banker (93) included statistical axioms defining a statistical model and a sampling process into the DEA framework.
- If the analyst accepts these rather restrictive axioms then DEA provides a consistent but biased estimator of the true frontier.
- Korostev et al. (95) and Kneip & Simar (98) proved consistency more generally and provided results on the rate of convergence.
- Models for approximating sample distributions using bootstrapping procedures were first developed in Simar & Wilson (98)
Another approach, initiated by Land et al. (1993), Olesen et al. (1995) and Cooper et al. (1998) focussed on specifying a random reference technology. Observed input and output data are replaced by DMU-specific distributions. Consequently, the generation of the DMUs, and the inefficiency were not seen as random draws from a common density on the input output space. The performance "within" the DMUs was related to the DMU-specific distributions. The theory of chance constraints was used to formalize the models.
Two Different Approaches To Stochastic DEA (1)

- Seen from the *Management Science* (MS) *framework*, DEA focuses on measuring deviations of inefficient DMUs from the best practice frontier,
  - No specific statistical model (no Data Generating Process (DGP)).
  - Stochastic DEA related to the MS framework replaces the observed input output observations with DMU specific distributions.
  - Focus on incorporating a random technology including measurement error (and possibly random inefficiency) into the specification of the reference technology.

- The *statistical* (or econometric) *framework* insists on an axiomatic statistical approach:
  - Data is regarded as a sample from a large population.
  - A statistical model including a specification of a DGP.
  - Axioms specify how the inefficiency deviations from a true frontier are generated.
  - Insists on a consistent estimator to allow for an estimation of the impact from sampling noise on the estimated efficiency scores (e.g. using bootstrapping).
Banker (93) approach (1), The DGP in Banker (93):

- We assume that we have a random sample of $n$ firms, given by the data $(Y_j, X_j) \in \mathbb{R}_+ \times \mathbb{R}^m_+, j \in N \equiv 1, \ldots, n$.
- Let the true $T = \{(x, y) \mid y \leq \phi(x)\}$ be a Production Possibility Set (PPS) that satisfies the maintained hypotheses:
  - i) $\phi(\cdot)$ satisfies monotonicity and concavity, ii) we maintain inclusion of observations, and iii) minimal extrapolation.
- We assume that data are generated according to a DGP that is very similar to the process behind "Deterministic Frontier Analysis".
- Based on the choice of inputs $x \in X \subset \mathbb{R}^m_+$ the output $y$ is generated randomly according to a DGP: $y = \phi(x) + u$,
- where $u$ is random, $u \leq 0$ with mean and standard deviation: $\bar{u}, \sigma_u$. $\phi(\cdot)$ is an unknown true frontier function and
- $Y^f_j = \phi(X_j), j \in N$ are the unknown frontier output values corresponding to the $n$ input vectors $X_j, j \in N$. 
For $k = 1, \cdots, n$, let us define

$$\hat{Y}_k^f = \widehat{\varphi} (X_k) = \max Y_k^f, \text{ s.t. } (X_k, Y_k^f) \in \hat{T}$$

$$\hat{T} = \left\{ (x, y) \mid \sum_{j \in N} \mu_j X_j \leq x, \sum_{j \in N} \mu_j Y_j \geq y, \sum_{j \in N} \mu_j = 1, \mu_j \geq 0, j \in N \right\}$$

with residuals

$$\hat{u}_k \equiv Y_k - \hat{Y}_k^f, k \in N$$
Banker (93) approach (3), DEA as Max Likelihood Estimation

PROPOSITION 2. If the probability density function $f(u)$ satisfies

1. The deviations from the frontier are iid distributed on a one-sided support. Hence, only negative residuals $u_k, k \in N$ are allowed in an output oriented model with one output and multiple inputs), and
2. The corresponding density function $f(u)$ is monotonically decreasing in the absolute size of the residuals.

and $X_k$ and $u_k$ are independently distributed, then the optimal solutions $\hat{Y}_k^f = \phi(X_k)$ solving (**) for $k \in N$, and $\hat{u}_k \equiv Y_k - \hat{Y}_k^f, k \in N$ solve the MLE problem:

$$\max_{f(\bullet), \phi(\bullet)} \prod_{j \in N} f(u_j)$$

s.t. $\phi(\bullet)$ satisfies monotonicity and concavity

$f(u) = 0$ for $u > 0$
Banker (93) approach (4), The consistency result

- Banker (93) shows that DEA (with one output) provides a consistent estimator of the best practice frontier as a piecewise linear monotonically increasing and concave production function $\varphi(\bullet)$, if one is willing to accept these two additional axioms.
- Hence, $\hat{T}$ is a consistent estimator and the distribution of the estimated residuals $\hat{u}_k, k \in N$ will of course converge to the distribution of the true residuals as sample size increases.
- Any finite sample of $n$ data points (generated from a Data Generating Process (DGP)) will provide a biased (why?) estimator and only asymptotically will $\hat{T}$ approach the true $T$.
- Imagine that we draw a lot of different samples of size $n$ (identical set of inputs). These samples would provide us with an empirical estimator of the distributions of the estimated residuals $\hat{u}_k, k \in N$.
- These sampling distributions or distributions of the estimated "residuals" $u_j$ are functions of the sample size.
- The sampling distributions allow for inference (e.g. confidence regions) of the estimated residuals (additive inefficiency).
Sample noise, n=10, inefficiency st.dev. 0.3, true $0.5x^{0.5}$:
Consistency, $n \in \{10, 20, \ldots, 90\}$, true prod. func $0.5x^{0.5}$:
This approach rests on the assumptions:
The data set is regarded as a sample from a large population, and we assume:

1. The deviations from the frontier are iid distributed on a one-sided support. Hence, only negative residuals $u_k, k \in N$ are allowed in an output oriented model with one output and multiple inputs), and

2. The corresponding density function is monotonically decreasing in the absolute size of the residuals.

The DEA estimator can provide inference if one is willing to accept that these axioms are reflecting reality.

Are we willing to accept these assumptions?
How restrictive are these assumption?
First two directions, non-parametric SFA (BM92), (1)

- Let us introduce both random noise and random inefficiency \( \rightarrow \) a stochastic PPS \( T^{BM} \) (Banker and Maindiratta (92)) (BM92).
- We assume that we have a random sample of \( n \) firms, given by the data \((Y_j, X_j) \in \mathbb{R}_+ \times \mathbb{R}_+^m, j \in N = \{1, \cdots, n\}\).
- We assume that data are generated according to a DGP that is very similar to the process behind the Stochastic Frontier Analysis.
- Based on the choice of inputs \( x \in \mathbb{R}_+^m \) the output \( y \) is generated randomly according to a DGP: \( y = \phi(x) + u + v \), where \( u \) and \( v \) are random variables, \( u \leq 0 \) with means and standard deviations: \( \bar{u}, 0, \sigma_u, \sigma_v \).
- \( \phi(\bullet) \) is an unknown true frontier function and \( Y^f_j = \phi(X_j), j \in N \) are the unknown frontier output values corresponding to the \( n \) input vectors \( X_j, j \in N \).

\[
\min_{v_j \in \mathbb{R}, u_j \leq 0, j \in N} \sum_{j \in N} (v_j + u_j)^2, \text{ s.t. } Y_j + u_k \in T^{BM} = ???
\]
Motivation (3) Disentangling random noise and random efficiency, n=10, st.dev. 0.3, true prod. func. 0.5x^0.5:
First two directions, non-parametric SFA, (BM92) (2)

- The Stochastic Frontier (SF) is $\varphi(x) + v$.
- However, we do not choose a specific functional form for $\varphi(x)$.
- We maintain instead that $\varphi(\cdot)$ belongs to the family of production functions $F_1$ that satisfies monotonicity and concavity.
- We assume that the observed output from an arbitrary DMU $j$ is an independent draw from the distribution of the random variable $\tilde{Y} = \varphi(X_j) + u + v$.
- We assume that $u_j \leq 0, j \in N$ are iid random inefficiency and $v_j, j \in N$ are iid random noise.
- In addition $u_j$ and $v_j$ are independent of each other and of the inputs $X_j, j \in N$.
- Let us assume that random inefficiency and random noise are half normal $N_+ (\bar{u}, \sigma^2_u)$ and normal $N (0, \sigma^2_v)$, respectively.
- Hence, the true random PPS can be written using the SF $\varphi(x) + v$ as $T^{BM}$
First two directions, non-parametric SFA (3)

\[ T^{BM} = \{ (\tilde{y}, x) , x \in \mathbb{R}_+^m | \tilde{y} = \phi(x) + \nu - s^+, \nu \sim N(0, \sigma^2_\nu), s^+ \geq 0 \} \]

Let us define the following \( n \) auxiliary functions

\[
g_i \left( Y_1^f, \ldots, Y_n^f \right) = \max \left[ \sum_{j \in N} \lambda_{ij} Y_j^f | \sum_{j \in N} \lambda_{ij} X_j \leq X_i, \sum_{j \in N} \lambda_{ij} = 1, \lambda_{ij} \geq 0, j \in N \right], i \in N
\]

Estimates of \( Y_1^f, \ldots, Y_n^f \), follow from

\[
\min_{j \in N} \sum_{j \in N} (v_j + u_j)^2
\]

s.t. \( Y_j - \left( Y_j^f + v_j + u_j \right) = 0 \quad j \in N \)

\( Y_i^f - g_i \left( Y_1^f, \ldots, Y_n^f \right) \geq 0 \quad i \in N \)

\( g_i \left( Y_1^f, \ldots, Y_n^f \right) \) is defined above \( i \in N \)

\( Y_j^f \geq 0, u_j \leq 0, v_j \in \mathbb{R}, j \in N \)
But how do we determine the estimated PPS $\hat{T}^{BM}$?

First, notice that we only know the estimated $\varphi (x)$ at the observed input vectors: $\hat{Y}_j^f = \hat{\varphi} (X_j), j \in N$.

You can show that the tightest estimator of $\varphi (x)$ is the convex envelopment of $(X_j, \hat{\varphi} (X_j)), j \in N$.

Hence, in general we use

$$
\hat{T}^{BM} = \left\{ (\tilde{y}, x) : x \in \mathbb{R}_+^m | \tilde{y} = y + \nu, y = \sum_{j \in N} \mu_j \hat{Y}_j^f - s^+, \sum_{j \in N} \mu_j X_j \leq x, \right. \\
\left. \sum_{j \in N} \mu_j = 1, \mu_j \geq 0, \forall j, \nu \sim N \left(0, \sigma^2_v\right), s^+ \geq 0 \right\}
$$

$$
= \left\{ (\tilde{y}, x) : x \in \mathbb{R}_+^m | \tilde{y} = \hat{\varphi} (x) + \nu - s^+, \nu \sim N \left(0, \sigma^2_v\right), s^+ \geq 0 \right\}
$$

The estimators $Y_j^f, j \in N$ can be shown to be equivalent to the CNLS estimator, Kuosmanen et al. (2010) with $Y_j^f = \alpha_j + w_j^T X_j$. 
The estimators $Y_j^f, j \in N$ can be shown to be equivalent to the CNLS estimator, Kuosmanen et al. (2010) with $Y_j^f = \alpha_j + w_j^T X_j$

$$\min \sum_{j \in N} (v_j + u_j)^2$$

$$s.t. \quad Y_j - \left( \alpha_j + w_j^T X_j + v_j + u_j \right) = 0 \quad j \in N$$

$$(\alpha_i + w_i^T X_i) - \left( \alpha_j + w_j^T X_i \right) \leq 0 \quad i, j \in N, i \neq j$$

$$w_j \geq 0, j \in N, \alpha_j \in \mathbb{R}, j \in N$$

Various methods can be used to correct these estimates to reflect the fact, that the residuals are composed of both error and inefficiency.
The random PPS $\widehat{T}^{BM}$ is related to the random PPS in Cooper et al. (98).

We focus on the random PPS in Cooper et al. 98 with $m$ deterministic inputs and only one random output $\tilde{y}$.

We assume that we have observations from $n$ DMUs in the form of input vectors $X_1, \cdots, X_n$ and $n$ random output levels $\tilde{Y}_1, \cdots, \tilde{Y}_n$, where we have information on the distribution of $\tilde{Y}_j, j = 1, \cdots, n$.

The random PPS, here denoted $T^{Cooper}$ can be expressed as

\[
T^{Cooper} = \left\{ (\tilde{y}, x) : x \in \mathbb{R}_+^m \mid \tilde{y} = \sum_{j=1}^n \mu_j \tilde{Y}_j - s^+, x \geq \sum_{j=1}^n \mu_j X_j, \sum_{j=1}^n \mu_j = 1, \mu_j \geq 0, \forall j, s^+ \geq 0 \right\}
\]
DMU specific distributions - where is the frontier??:

One random output and one random input.
DMU specific distributions - where is the frontier??:

One random output and one deterministic input.
Note, the similarities between $T^{Cooper}$ and $\hat{T}^{BM}$

For comparison reasons regard $\tilde{Y}_j$ as a noise perturbed $Y_j^f$, i.e. $\tilde{Y}_j = Y_j^f + v_j$.

The input, mean-output combinations along the SF from $T^{Cooper}$ will not be piecewise linear

A convex combination of iid $N(0, \sigma_v^2)$ residuals will not have a constant variance.
Hence, we can either express a random PPS $T_1$ using the same approach as in $T^{Cooper}$, namely to accept convex combinations of unknown random frontier output vectors $Y^f_j e^{v_j}, j = 1, \ldots, n$:

$$T_1 = \left\{ (\tilde{y}, x) , x \in \mathbb{R}^m_+ \mid \tilde{y} = \sum_{j=1}^{n} (\mu_j Y^f_j + v_j) - s^+, x \geq \sum_{j=1}^{n} \mu_j X_j, \sum_{j=1}^{n} \mu_j = 1, \mu_j \geq 0, v_j \sim N\left(0, \sigma^2_v\right), \forall j, s^+ \geq 0 \right\}$$

or we can express a random PPS $T_2$ generated by noise perturbations of any feasible output given an input vector derived from the convex hull of the deterministic vectors $(X_j, Y^f_j), j = 1, \ldots, n$.

$$T_2 = \left\{ (\tilde{y}, x) , x \in \mathbb{R}^m_+ \mid \tilde{y} = y + v, y = \sum_{j=1}^{n} \mu_j Y^f_j - s^+, x \geq \sum_{j=1}^{n} \mu_j X_j, \sum_{j=1}^{n} \mu_j = 1, \lambda_j \geq 0, v \sim N\left(0, \sigma^2_v\right), \forall j, s^+ \geq 0 \right\}$$
Two different approaches have been adopted within the area of efficiency analysis, DEA and SFA.

The DEA models were developed within the management science tradition and SFA within the statistical or econometric tradition.

The main strength of DEA reflects that DEA is a non-parametric estimation of the production possibility set based upon fundamental axioms from production theory including convexity and monotonicity.

No functional form for the frontier or the distribution of inefficiency is assumed.

The outcome of an efficiency analysis based upon DEA is for these reasons easy to communicate to decision makers.

The identification of a reference set of DMUs to compare with is an important piece of information for DMUs termed inefficient.
Banker (93) establishes a statistical foundation for DEA by demonstrating that DEA (with one output) provides a consistent estimator of the frontier, if the deviations from the frontier are iid with a one sided support and a monotonic decreasing density.

Simar and Wilson (98) provide a foundation for bootstrapping of confidence intervals for DEA efficiency measures.

However, a common inefficiency distribution is often maintained to allow for the relatively simple homogenous bootstrap.

The statistical tradition insists upon an axiomatic foundation for a statistical model (e.g. a DGP).

Observed data are considered a sample from an underlying large population.

The underlying assumptions partly reflect what is necessary to get a consistent estimator of a true frontier.
We have argued that the Management Science approach for efficiency analysis provides an outcome that is meaningful as seen from a decision making point of view.

We have also argued that statistical inference based upon the homogeneous bootstrap approach is accompanied by a heroic assumption of a common distribution of inefficiency for all DMUs and that an aggregation of inputs and outputs may well be necessary to provide access to stable inference due to the curse of dimensionality.

As a consequence, a statistical framework seems to prioritize aggregation possibilities, e.g. linear combinations of inputs and outputs, which best summarize the information provided by all DMUs.

The statistical (or econometric) approach to estimating and testing technology characteristics typically investigates to what extent the general tendency in data questions a given hypothesis.

The individual data point is not at focus, because the set of all data are regarded as a particular sample from the DGP.
Chance Constrained DEA (CCDEA) and semi-parametric SFA may appear to be competing approaches within the area of Stochastic DEA.

Proponents of semi-parametric SFA may argue that the approach is superior compared to CCDEA, because it is consistent with basic axioms in production theory and has a firm statistical foundation including an underlying DGP.

By contrast, proponents of CCDEA may argue the superiority of this approach, because

1. DMU-specific distributions of noise and inefficiency are easily accommodated,
2. the case of multiple inputs and multiple outputs can be handled in a straightforward manner, and
3. recent work by Simar establishes a DGP for a multivariate, cross-sectional, nonparametric stochastic frontier model that may be applicable in the context of CCDEA.