Productivity and Efficiency Analysis

5) Contextual variables

5c) Contextual variables in DEA

Timo Kuosmanen

Aalto University School of Business

https://people.aalto.fi/timo.kuosmanen

Contextual variables as nondiscretionary inputs

$$y_i = f(\mathbf{x}_i, \mathbf{z}_i) - u_i + v_i$$

This is the classic approach in DEA (Banker and Morey, 1986).

But are contextual variables **z** substitutable with inputs **x**?

Do contextual variables satisfy the maintained axioms such as free disposability and convexity?

2-stage DEA

In DEA application, it is common to estimate the effect of *z*-variables in two stages:

1) Estimate the DEA efficiency (*Eff*) using inputs $\mathbf{x}$ and outputs $\mathbf{y}$.

2) Regress the estimated *Eff* scores on z-variables using OLS or Tobit:

$$Eff_i = \boldsymbol{\delta}'\mathbf{z}_i + w_i$$

Simar & Wilson (2007)

- Critique of 2-DEA: what is the model?
- Probabilistic data generating process (DGP) with **z** variables, assuming away noise.
- In the 2nd stage regression, truncated Tobit regression (rather than censored) is a consistent estimator of coefficients $\boldsymbol{\delta}$.
- Conventional statistical inferences fail: apply bootstrap.

Banker & Natarajan (2008)

- Semi-nonparametric model

$$\ln y_i = \ln f(\mathbf{x}_i) + \boldsymbol{\delta}'\mathbf{z}_i - u_i + v_i$$

where the noise term v is double-truncated:

$$-V^M \leq v \leq V^M.$$

- Estimate DEA efficiency (*Eff*) using data of $\mathbf{x}, y$
- Second stage OLS regression of *Eff* on $\mathbf{z}$ is statistically consistent.
- Monte Carlo simulations to assess the OLS regression in the 2nd stage

Johnson & Kuosmanen (2012)

- Follow Banker and Natarajan (2008), but relax some of their assumptions.
- Formally show dependence of the bias in the 2nd stage OLS regression on the finite sample bias of DEA in the 1st stage.
  - Analogous to the omitted variable bias in two-stage SFA
- Propose one-stage DEA estimator (1-DEA) based on convex regression.
- Replicate Monte Carlo simulations by Banker and Natarajan (2008)

1-DEA problem

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\delta},\hat{\phi}} \quad \sum_{i=1}^{n} \varepsilon_i^2$$

$$\begin{aligned}
\text{s.t.} \quad & \ln y_i = \ln \hat{\phi}_i + \mathbf{z}_i \boldsymbol{\delta} + \varepsilon_i \quad \text{for all } i = 1, \ldots, n \\
& \hat{\phi}_i = \alpha_i + \mathbf{x}_i \boldsymbol{\beta}_i \quad \text{for all } i = 1, \ldots, n \\
& \hat{\phi}_i \leqslant \alpha_h + \mathbf{x}_i \boldsymbol{\beta}_h \quad \text{for all } h, i = 1, \ldots, n \\
& \boldsymbol{\beta}_i \geqslant \mathbf{0} \quad \text{for all } i = 1, \ldots, n \\
& \varepsilon_i \leqslant V^M \quad \text{for all } i = 1, \ldots, n
\end{aligned}$$

Note: output-oriented DEA VRS is the special case where $V^M = 0$ and $\boldsymbol{\delta} = \mathbf{0}$

Johnson & Kuosmanen (2012): baseline scenario

Table 1
Performance (by RMSD) of alternative methods

Estimation Method	$\rho = -0.8$	$\rho = 0.0$	$\rho = 0.8$
OLS	774	117	753
2-DEA	74.0	18.9	57.6
1-DEA, $V^M = 0$	32.3	20.3	32.0
1-DEA, $V^M = 1.04$	29.5	15.2	30.5
1-DEA, $V^M \to \infty$	29.1	17.9	30.1

Consider first the base case of Banker and Natarajan where the parameter values are set as $\delta = -0.2, \sigma_u = 0.15$, $\sigma_v = 0.04$, $n = 100$.

Next lesson

5d) Semi-nonparametric approach