Productivity and Efficiency Analysis

6) Multiple outputs and bad outputs

6e) StoNED with multiple outputs

Timo Kuosmanen

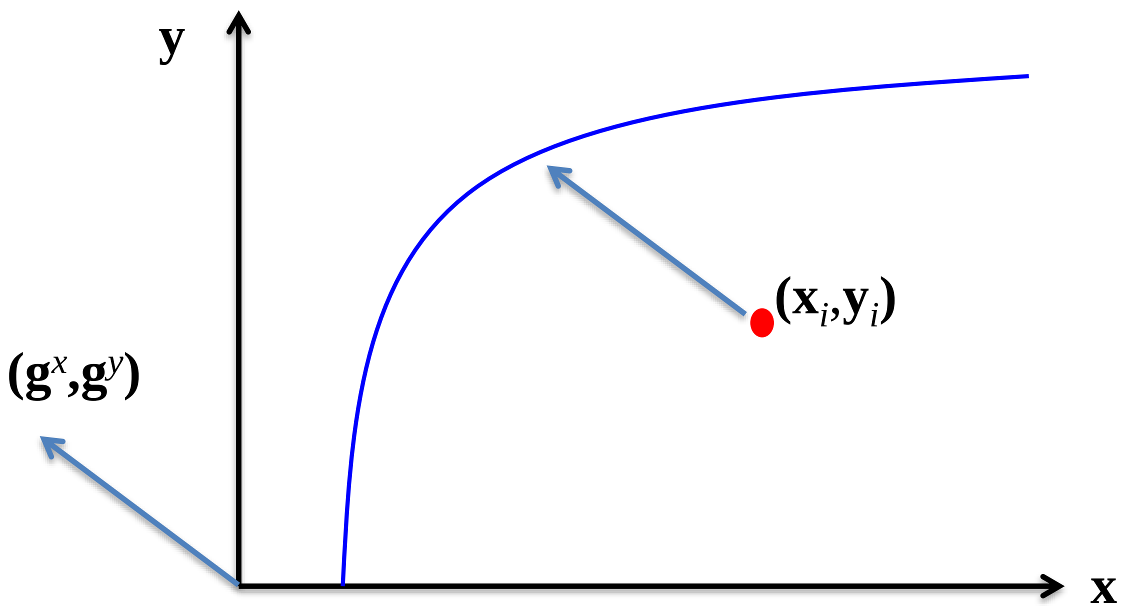
Aalto University School of Business

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

Motivation

- Modeling joint production as parallel processes ignores synergies of joint production
- DEA can handle synergies and bad outputs, but is sensitive to noise
- SFA extends to multiple outputs using parametric distance functions, but the parametrizations violate free disposability and/or convexity, and cannot handle specialized firms
- Need for better tools…

Directional distance function: illustration



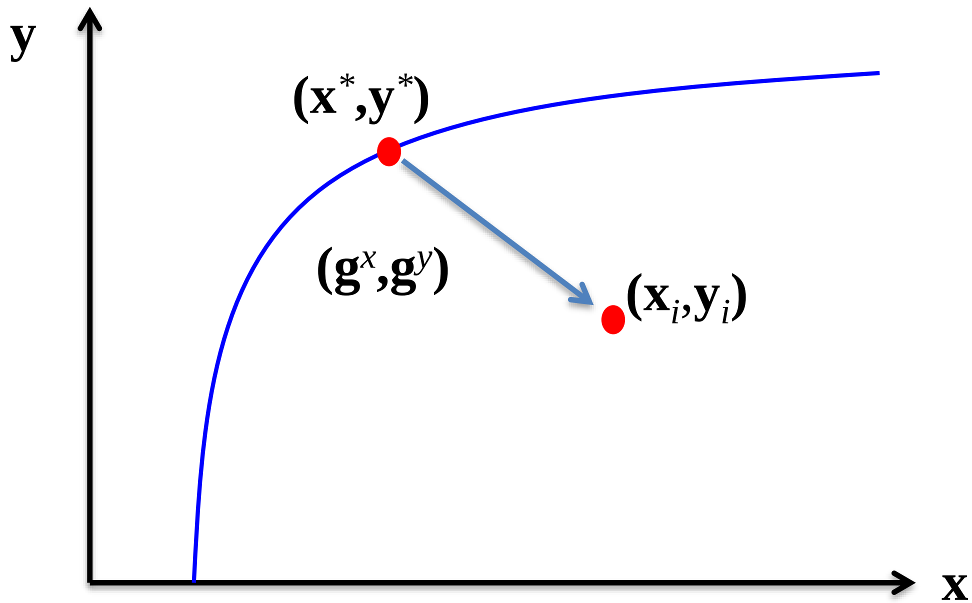
Directional data generating process (DGP):

Assume the observed data ($\mathbf{x}_i$, $\mathbf{y}_i$) are perturbed in direction ($\mathbf{g}^x$, $\mathbf{g}^y$) such that

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

Target points ($\mathbf{x}_i^*$, $\mathbf{y}_i^*$):

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



Kuosmanen & Johnson (2017) *EJOR*

Then DDF equals composite error term

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

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

Aalto University
School of Business

Kuosmanen & Johnson (2017) *EJOR*

Convex regression of DDF

Regression equation:

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

where

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

Note: the arbitrary choice of y_1 as the dependent variable does not affect results in any way. Any other output or input could be used.

Convex regression with multiple outputs

Regression equation:

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

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

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

Kuosmanen & Johnson (2017) *EJOR*

Convex regression with multiple outputs

Regression equation:

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

Convex nonparametric least squares (CNLS) estimator

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

subject to

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

Kuosmanen & Johnson (2017) *EJOR*

Convex regression with multiple outputs

Equivalent CNLS formulation

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

subject to

$$\boldsymbol{\gamma}'_i \mathbf{y}_i = \alpha_i + \boldsymbol{\beta}'_i \mathbf{x}_i - \varepsilon_i^{CNLS} \; \forall i = 1, \ldots, n$$

$$\alpha_i + \boldsymbol{\beta}'_i \mathbf{x}_i - \boldsymbol{\gamma}'_i \mathbf{y} \leq \alpha_h + \boldsymbol{\beta}'_i \mathbf{x}_i - \boldsymbol{\gamma}'_h \mathbf{y}_i \; \forall h, i = 1, \ldots, n$$

$$\boldsymbol{\gamma}'_{ii} \mathbf{g}^y + \boldsymbol{\beta}'_{ii} \mathbf{g}^x = 1 \; \forall i = 1, \ldots, n$$

$$\boldsymbol{\beta}_i \geq \mathbf{0} \; \forall i = 1, \ldots, n$$

$$\boldsymbol{\gamma}_i \geq \mathbf{0} \; \forall i = 1, \ldots, n$$

Kuosmanen, Johnson & Saastamoinen (2015) book ch.

Application to electricity distribution firms revisited

- Regulation periods 4 (2016-2019), and 5 (2020-2023)

Inputs:

Variable input:

x = Controllable operational expenditure (COPEX, €)

Fixed input:

K = Capital stock (replacement value, €)

Outputs:

Desirable outputs $\mathbf{y}$:

y_1 = Energy supply (GWh, weighted by voltage)

y_2 = Network length (km)

y_3 = Number of use points

Undesirable output:

b = Outages (hedonic damage cost, €)

Contextual variables:

z = Connection points / Use points

Application to electricity distribution firms revisited

Step 1: CNLS estimation of the DDF

Step 2: Kernel density estimation of the CNLS residuals

Step 3: Directional shifting of the DDF to the frontier

Step 4: Computing shadow prices of the frontier

Step 5: Excel spreadsheet for computing efficient level of COPEX x, given K, $\mathbf{y}$, b, and z.

Conditional yardstick competition

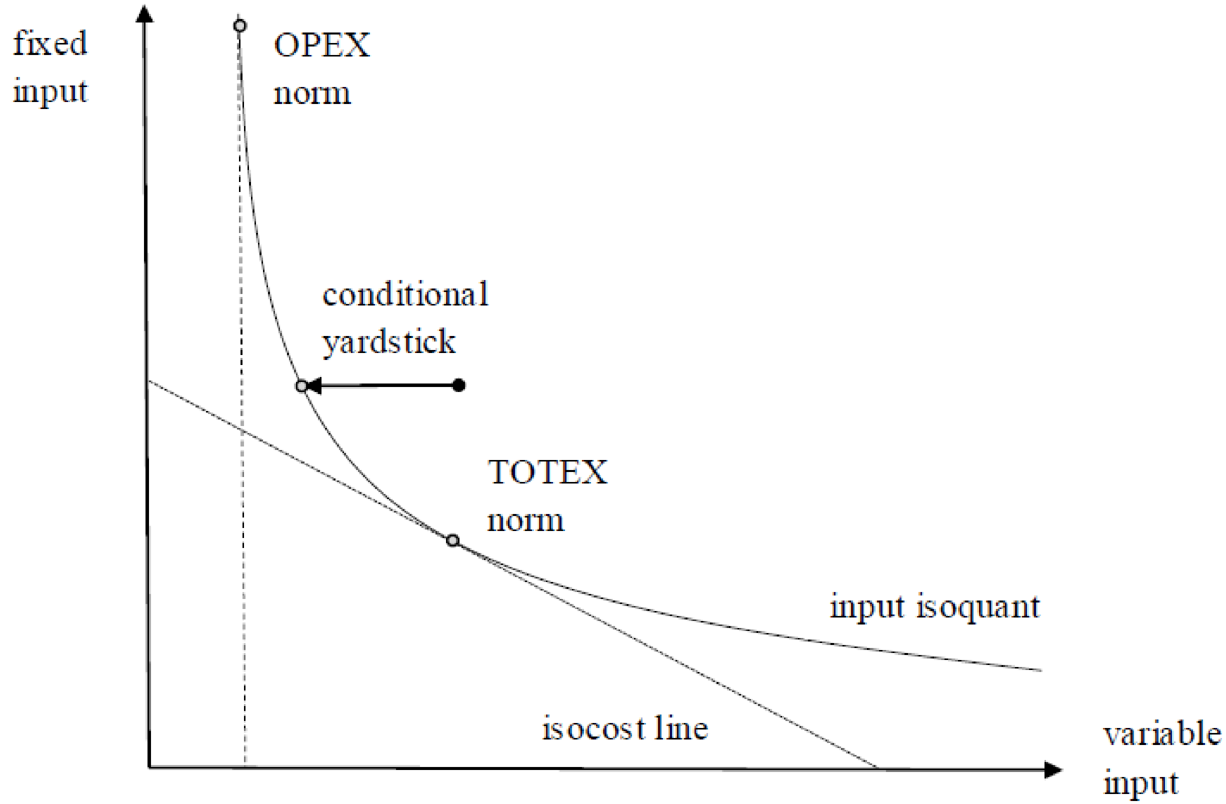


Figure 1: Illustration of the OPEX and TOTEX norms and the conditional yardstick.

Further research

- Extending StoNED to the radial input and output distance functions
- First attempt by Schaefer and Clermont (2018)
  - Sensitive to the choice of the output variable as the dependent variable on the LHS of the equation.
  - This problem is solvable, but the solution remains to be published.

Next lesson

7) Productivity analysis