Productivity and Efficiency Analysis

5) Contextual variables

b) SFA modeling of z-variables

Timo Kuosmanen

Aalto University School of Business

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

Contextual variables in the inefficiency term

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

This is the most standard approach in the parametric SFA literature.

2-stage estimation strategy

It might be tempting to estimate the effect of z-variables in two stages:

1) Estimate the frontier model by SFA

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i - u_i + v_i$$

2) Regress the estimated *u* on z-variables:

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

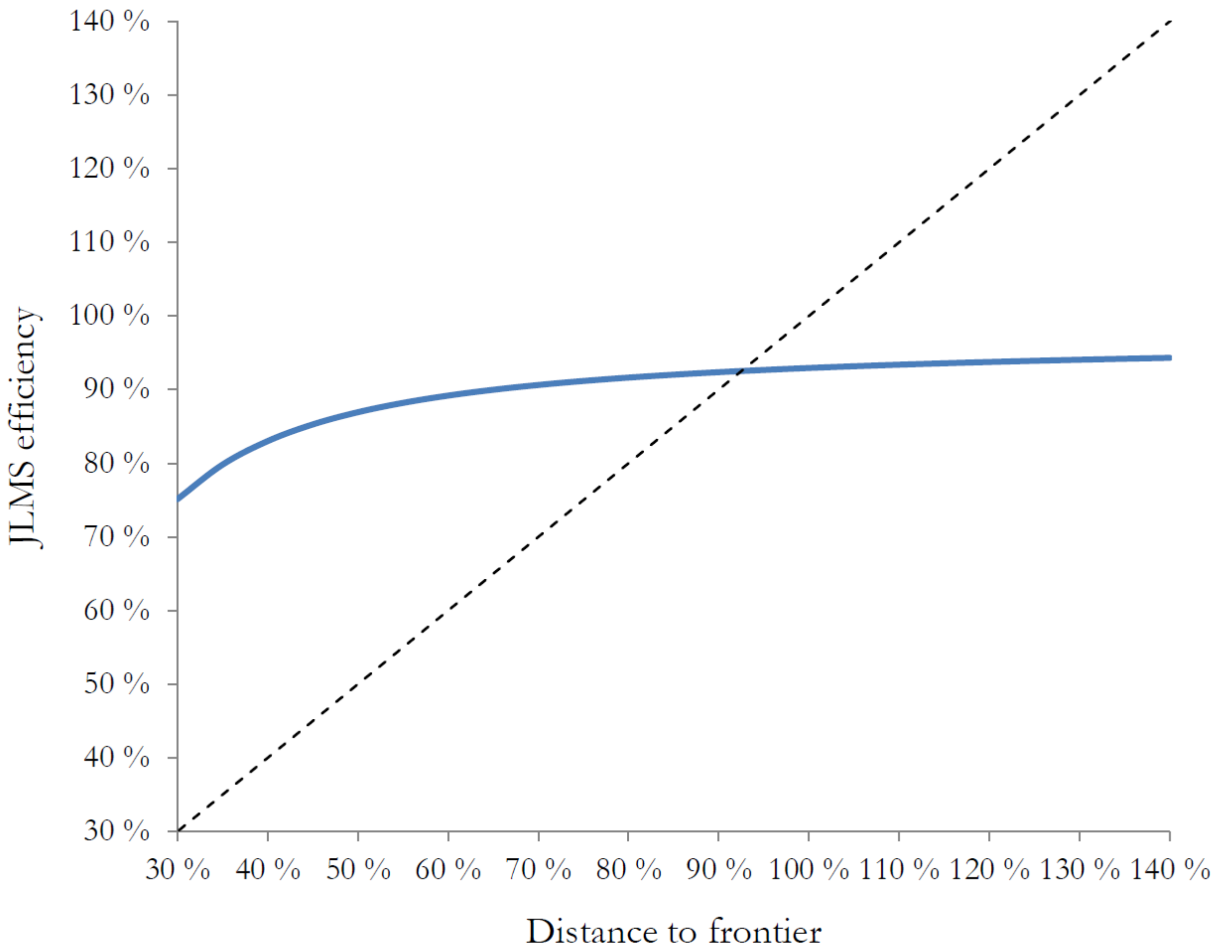
2-stage approach is biased

Two sources of bias (Wang & Schmidt, 2002):

1) If $\mathbf{z}$ correlates with $\mathbf{x}$, then the SFA estimates in stage 1 are subject to omitted variable bias (endogeneity problem).
2) Even if $\mathbf{z}$ and $\mathbf{x}$ are uncorrelated, the 2$^{\text{nd}}$ stage regression of u on $\mathbf{z}$ is biased because of the "shrinkage" property of the JLMS estimator.

SFA efficiency

Illustration of the “shrinkage” of the JLMS transformation:



1-stage ML estimation

Two alternative parametrizations:

1) Truncated normal model of inefficiency (Kumbhakar et al., 1991): $N^{+}(\boldsymbol{\delta}'\mathbf{z}_i, \sigma_u)$

2) Models with scaling property (Reifschneider and Stevenson, 1991): $u_i = u^{*} \cdot \exp(\boldsymbol{\delta}'\mathbf{z}_i)$,

where u^{*} is a random inefficiency term that does not depend on $\mathbf{z}$.

Scaling property (Wang & Schmidt, 2002)

Scaling property is appealing because

1) The shape of the inefficiency distribution is the same across all firms.
2) The coefficients $\boldsymbol{\delta}$ have an intuitive interpretation as the marginal effects of $\mathbf{z}$ on inefficiency u
3) Coefficients $\boldsymbol{\delta}$ can be estimated without specifying the distribution of u^*.

$$E(y_i | \mathbf{x}_i, \mathbf{z}_i) = \boldsymbol{\beta}'\mathbf{x}_i - (\boldsymbol{\delta}'\mathbf{z}_i)\mu^*$$

Truncated N^+ model

Parameters **δ** of the truncated normal model are more difficult to interpret, and may lead to misinterpretations.

Example: Méon and Weill (2010) apply the truncated N model to estimate the effect of corruption on economic growth of countries.

They find that corruption *improves* efficiency, serving as "grease on the wheels".

Truncated N^+ model

- Saastamoinen & Kuosmanen (2014) re-examine the "grease on the wheels" hypothesis, using OLS regression, SFA and StoNED:
- OLS and CNLS residuals indicated that corruption increases the variance of the composite error term (heteroscedasticity)
- However, corruption has significant negative effect on the expected growth (grit on the wheels)
- SFA formulation used does not allow the signs of the mean and variance effects differ. The estimated coefficient captures the heteroscedasticity effect.

Truncated N⁺ model

The mean and variance of the truncated normal inefficiency term *u* both depend on **z**s:

$$E[u|\text{truncation}] = \boldsymbol{\theta}'\mathbf{z}_i + \sigma_w \lambda(\omega)$$
$$\text{Var}[u|\text{truncation}] = \sigma_w^2[1 - \upsilon(\omega)]$$

where

$$\omega = (-\boldsymbol{\theta}'\mathbf{z}_i)/\sigma_w$$
$$\lambda(\omega) = \phi(\omega)/[1 - \Phi(\omega)]$$
$$\upsilon(\omega) = \lambda(\omega)[\lambda(\omega) - \omega]$$

Next lesson

5c) Contextual variables in DEA