Productivity and Efficiency Analysis

5) Contextual variables

5d) Semi-nonparametric approach

Timo Kuosmanen

Aalto University School of Business

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

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:

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

- We next relax the truncation of the noise term and develop a more efficient one-stage estimator and statistical inferences.

Johnson & Kuosmanen (2011)

- Introduce the one-stage StoNED estimator
- Show statistical properties of the estimator
- Monte Carlo simulations

StoNEZD problem

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\delta},\hat{\phi}} \sum_{i=1}^{n} (\ln y_i - \ln \hat{\phi}_i - \mathbf{z}_i'\boldsymbol{\delta})^2$$

$s.t.$

$$\hat{\phi}_i = \alpha_i + \mathbf{x}_i\boldsymbol{\beta}_i \ \forall i = 1, \ldots, n$$

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

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

Note: equivalent to 1-DEA where V^M approaches to infinity.

Johnson & Kuosmanen (2011)

Theorem 1 *If the data-generating process satisfies the maintained assumptions stated in Sect. 2, the StoNEZD-estimator for the coefficients of the contextual variables* ($\hat{\boldsymbol{\delta}}^S$) *is statistically unbiased:*

$$E(\hat{\boldsymbol{\delta}}^S) = \boldsymbol{\delta}. \tag{4}$$

asymptotically normally distributed according to:

$$\hat{\boldsymbol{\delta}}^S \sim {}_a N(\boldsymbol{\delta}, (\sigma_v^2 + \sigma_u^2)(\mathbf{Z}'\mathbf{Z})^{-1}). \tag{5}$$

Johnson & Kuosmanen (2011)

Theorem 3 *If the conditions stated in Theorem 2 are satisfied and the skewness of the inefficiency terms* u_i *is finite such that* $E(|u_i - \mu|^3) = \gamma < \infty$, *then the StoNEZD-estimator for the coefficients of the contextual variables* ($\hat{\boldsymbol{\delta}}^S$) *converges to the true* $\boldsymbol{\delta}$ *at the standard parametric rate on the order of* $n^{-1/2}$. *Specifically, there exist a positive constant* C *such that for all* n,

$$|\hat{\boldsymbol{\delta}}^S - \boldsymbol{\delta}| \leq \frac{C\gamma}{\sqrt{n}(\sigma_v + \tilde{\sigma}_u)^3}\mathbf{1}. \tag{6}$$

Johnson & Kuosmanen (2011)

Theorem 3 *If the conditions stated in Theorem 2 are satisfied and the skewness of the inefficiency terms* u_i *is finite such that* $E(|u_i - \mu|^3) = \gamma < \infty$, *then the StoNEZD-estimator for the coefficients of the contextual variables* ($\hat{\boldsymbol{\delta}}^S$) *converges to the true* $\boldsymbol{\delta}$ *at the standard parametric rate on the order of* $n^{-1/2}$. *Specifically, there exist a positive constant* C *such that for all* n,

$$|\hat{\boldsymbol{\delta}}^S - \boldsymbol{\delta}| \leq \frac{C\gamma}{\sqrt{n}(\sigma_v + \tilde{\sigma}_u)^3}\mathbf{1}. \tag{6}$$

Johnson & Kuosmanen (2011)

Theorem 4 *If the conditions stated in Theorem 2 are satisfied, then the StoNEZD-estimator for the coefficients of the contextual variables* ($\hat{\boldsymbol{\delta}}^S$) *is asymptotically efficient. That is,*

$$\mathrm{AsyVar}(\hat{\boldsymbol{\delta}}^S) \leq \mathrm{AsyVar}(\hat{\boldsymbol{\delta}}) \tag{7}$$

for any other consistent, asymptotically normally distributed estimator $\hat{\boldsymbol{\delta}}$.

Monte Carlo simulations

Table 3 Performance (by RMSD) of alternative methods in estimating the effect of contextual variable (δ) varying the number of firms (n) and the correlation of x and z (ρ). Triple signal scenario

Scenario	Correlation (ρ)	Estimation method	
		Two-stage DEA	StoNEZD
$n = 50$	−0.8	71.5	13.8
	−0.4	28.9	9.60
	0.0	12.6	8.96
	0.4	15.7	9.73
	0.8	44.6	15.0
$n = 100$	−0.8	66.7	9.73
	−0.4	23.3	6.81
	0.0	7.66	6.81
	0.4	15.6	6.95
	0.8	46.9	9.28
$n = 200$	−0.8	67.9	5.78
	−0.4	20.6	3.77
	0.0	4.28	3.46
	0.4	11.1	3.77
	0.8	39.1	5.95

Case: $\delta = -0.6$, $\sigma_u = 0.15$, $\sigma_v = 0.04$, $M = 100$

Statistical inferences in StoNEZD

A simple trick to compute standard errors for $\hat{\boldsymbol{\delta}}^{StoNEZD}$ is to run OLS regression where the contextual variables $\mathbf{Z}$ are regressors and the dependent variable is the difference between the natural log of observed output subtracting the natual log of the input aggregation plus 1, specifically $\ln y_i - \ln(\hat{\phi}_i + 1) = \hat{\boldsymbol{\delta}}'\mathbf{z}_i + \hat{\varepsilon}_i^{CNLS}$. This OLS regression will yield the same coefficients $\hat{\boldsymbol{\delta}}'^{StoNEZD}$ that were obtained as the optimal solution to problem (7.46),[25] but also return the standard errors and other standard diagnostic statistics such as t-ratios, p-values, and confidence intervals.

[25] Note that this two-stage regression procedure is not subject to the problems of the 2-DEA procedure because we do control for the effects of the contextual variables in the first stage CNLS

Statistical inferences in StoNEZD: Finnish electricity distribution firms

Table 3
Parameter estimates of the z variable (proportion of underground cables in the total length of network).

Parameter	Estimate
δ coefficient	0.3600
Standard error	0.0581
t-statistic	6.1942
p-value	0.0000
95% lower limit	0.2443
95% upper limit	0.4752
Partial R^2	0.3060

Next lesson

6) Multiple outputs