Aalto University
School of Economics
Helsinki, Finland

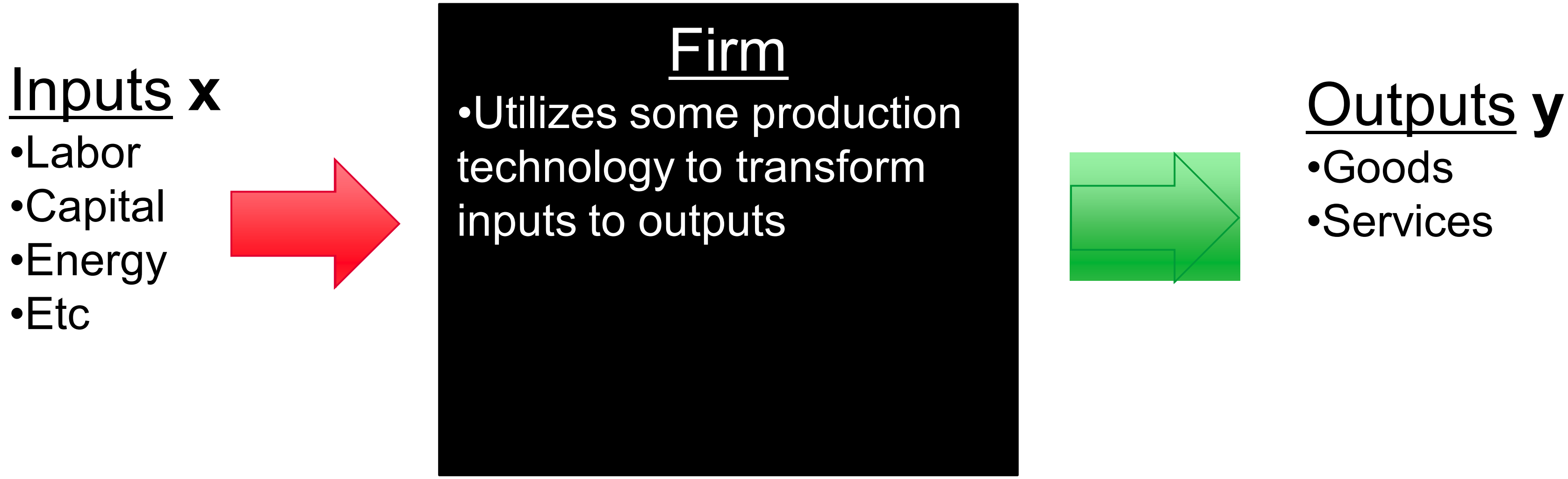
How Operational Conditions and Practices Affect Productive Performance?

Efficient Semi-parametric One-Stage Estimators

Andrew Johnson (Texas A&M) and Timo Kuosmanen (Aalto University)

North American Productivity Workshop 2010
Rice University, Houston, Texas
2-5 June 2010

Conceptual setting



Conceptual setting

<u>Inputs</u> **x**
- Labor
- Capital
- Energy
- Etc

<u>Firm</u>
- Utilizes some production technology to transform inputs to outputs

Managerial practices
- incentive schemes
- ownership

<u>Outputs</u> **y**
- Goods
- Services

Operational conditions
- **competition**
- **regulation**

Conceptual setting

Inputs x
- Labor
- Capital
- Energy
- Etc

Firm
- Utilizes some production technology to transform inputs to outputs

Managerial practices
- incentive schemes
- ownership

Outputs y
- Goods
- Services

Contextual variables z

Operational conditions
- **competition**
- **regulation**

Semiparametric model

$$y = f(\mathbf{x}) \cdot e^{\mathbf{z}\boldsymbol{\delta} - u + v}$$

where

- *f* is a nonparametric, monotononic increasing and concave production function
- **δ** is a parameter vector representing marginal effects of contextual variables on performance
- *u* is a random non-negative inefficiency term
- *v* is a random symmetric noise term

Classification of approaches

	Parametric frontier	Nonparametric frontier
No noise: $v = 0$	**PP** Aigner and Chu (1968)	**DEA** Farrell (1957) Charnes et al. (1978)
Noise assumed		
ML estimation	**SFA – ML** Aigner et al. (1977), Meeusen & Vanden Broeck (1977), Battese & Corra (1977)	**BM** Banker & Maindiratta (1992), Sarath & Maindiratta (1997)
Least-squares	**SFA – MOLS** Schmidt (1976), Aigner et al. (1977)	**StoNED** Kuosmanen (2006), Kuosmanen & Kortelainen (2007)

StoNEZD estimator

(Stochastic Non-smooth Envelopment of Z-variables Data)

Strategy: Estimate *f* and **δ jointly** from the equation

$$\ln y = \ln f(\mathbf{x}) + \mathbf{z}\boldsymbol{\delta} + \varepsilon$$

using Convex Nonparametric Least Squares (CNLS)

- Kuosmanen (2008) *Econometric Journal*
- Kuosmanen & Johnson (2010) *Operations Research*

Firm-specific inefficiencies *u* can be estimated from residuals *e*:

- In cross-section: conditional expectation $E[u|e]$ (Jondrow et al. 1982)
- In panel data: fixed or random effects

StoNEZD estimator

Solve the least squares problem

$$\min_{f,\boldsymbol{\beta},\varepsilon} \sum_{i=1}^{n} \varepsilon_i^2$$

$$s.t.$$

$$\ln y_i = \ln f(\mathbf{x}_i) + \mathbf{z}_i \boldsymbol{\delta} + \varepsilon_i \quad \forall i$$

f is monotonic increasing and concave

StoNEZD estimator

Equivalent CNLS formulation:

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

$$s.t.$$

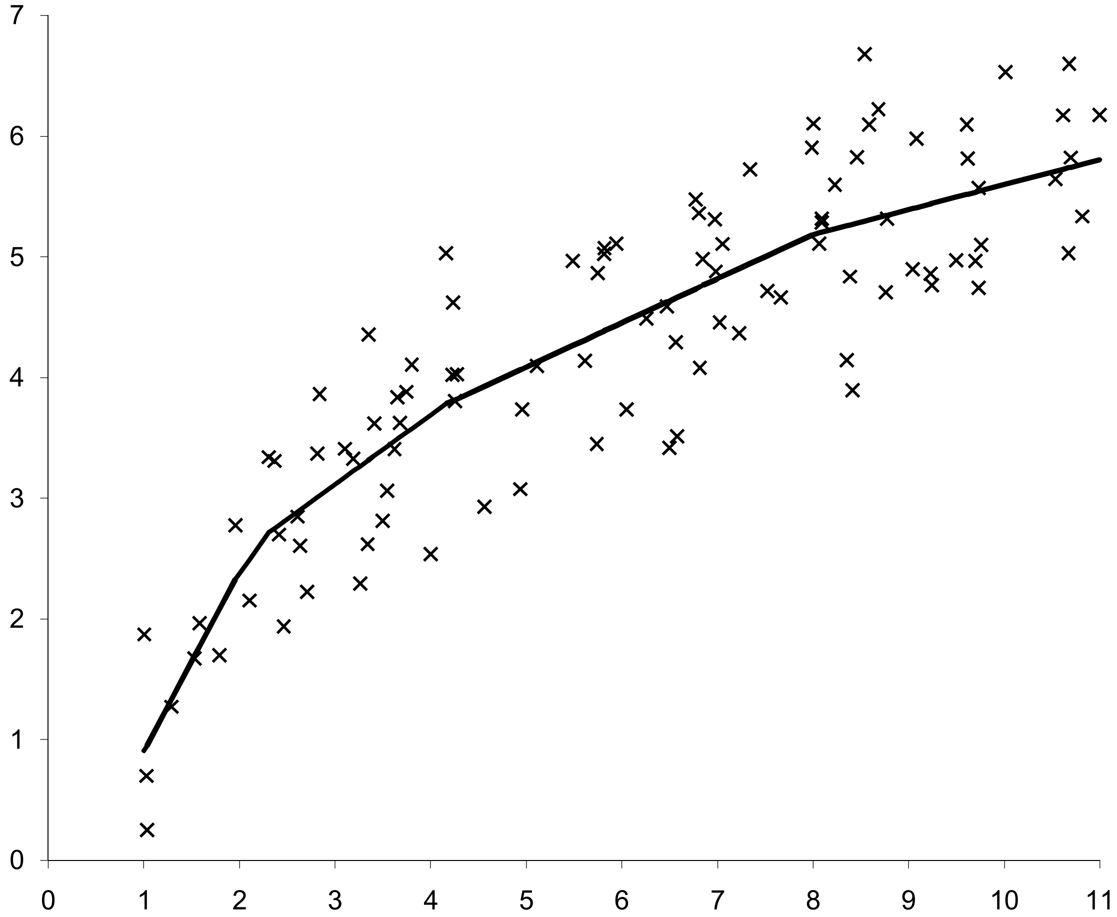
$$\ln y_i = \ln \phi_i + \mathbf{z}_i \boldsymbol{\delta} + \varepsilon_i \quad \forall i$$ Regression equation

$$\phi_i = \alpha_i + \mathbf{x}_i \boldsymbol{\beta}_i \quad \forall i$$ Fitted $E(y|\mathbf{x})$

$$\alpha_i + \mathbf{x}_i \boldsymbol{\beta}_i \le \alpha_h + \mathbf{x}_i \boldsymbol{\beta}_h \quad \forall i,h$$ Concavity (Afriat)

$$\boldsymbol{\beta}_i \ge \mathbf{0} \quad \forall i$$ Monotonicity

Illustration



Statistical properties

The coefficients of the contextual variables **δ** satisfy

$$\hat{\boldsymbol{\delta}}^{StoNEZD} = [(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'](\ln \mathbf{y} - \ln \hat{\mathbf{y}}^{StoNEZD})$$

Two-stage interpretation: If we regress ($\ln \mathbf{y}$ - $\ln \mathbf{y}^{StoNEZD}$) on contextual variables **Z** by OLS, we obtain the same coefficients $\boldsymbol{\delta}^{StoNEZD}$

- We also get the standard errors and t-statistics

Statistical properties

We prove that $\hat{\boldsymbol{\delta}}^{StoNEZD} = [(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'](\ln \mathbf{y} - \ln \hat{\mathbf{y}}^{StoNEZD})$

is

- **Consistent**

Statistical properties

We prove that $\hat{\boldsymbol{\delta}}^{StoNEZD} = [(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'](\ln \mathbf{y} - \ln \hat{\mathbf{y}}^{StoNEZD})$

is

- **Consistent**
- **Asymptotically efficient**

Statistical properties

We prove that $\hat{\boldsymbol{\delta}}^{StoNEZD} = [(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'](\ln \mathbf{y} - \ln \hat{\mathbf{y}}^{StoNEZD})$

is

- **Consistent**
- **Asymptotically efficient**
- **Asymptotically normal**

Statistical properties

We prove that $\hat{\boldsymbol{\delta}}^{StoNEZD} = [(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'](\ln \mathbf{y} - \ln \hat{\mathbf{y}}^{StoNEZD})$

is

- **Consistent**
- **Asymptotically efficient**
- **Asymptotically normal**
- **Converges at standard parametric rate $n^{-1/2}$**

Statistical properties

We prove that $\hat{\boldsymbol{\delta}}^{StoNEZD} = [(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'](\ln \mathbf{y} - \ln \hat{\mathbf{y}}^{StoNEZD})$

is

- Consistent
- Asymptotically efficient
- Asymptotically normal
- Converges at standard parametric rate $n^{-1/2}$

$\Rightarrow$ Conventional methods of statistical inference justified (*t*-tests, confidence intervals, etc.)

How about heteroskedasticity?

Contextual variables **z** can also affect the variance of *u* (or *v*) (Wang 2002, Alvarez et al. 2006), causing heteroskedasticity

In the case of heteroskedasticity, the least-squares estimator is unbiased and consistent, but inefficient.

The main problem: standard errors are incorrect.

Standard econometric solutions:

- Use robust standard errors (White's heteroskedasticity consistent estimator)
- Estimate heteroskedasticity, and apply generalized least squares (GLS) estimator

Monte Carlo simulations

Replicate Banker and Natarajan (2008, *OR*) simulations:

$$y_i = (x_i^3 - 12x_i^2 + 48x_i - 37) \cdot \exp(\delta z_i + v_i - u_i)$$

Baseline scenario:
δ = -0.2
σ_u = 0.15
σ_v = 0.04
n = 200

How correlation (ρ) between x and z variables affects precision?

Monte Carlo simulations

Banker and Natarajan (2008, *OR*)

Root Mean Squared Deviation (RMSD) (%) for *δ* at different levels of correlation (*ρ*) between x and z variables

Estimator	ρ=-0.8	ρ=-0.2	**ρ= 0**	ρ=+0.2	ρ=+0.8
OLS: y on z	750	217	**93.0**	214	744
SFA cubic 1-stage MLE	15.7	9.5	**9.3**	9.5	15.4
SFA CD 1-stage MLE	51.7	19.4	**19.3**	20.5	52.4
SFA translog 1-st. MLE	43.9	36.9	**32.8**	32.2	40.7
2-DEA w. OLS	83.0	16.3	**10.8**	11.4	36.5
2-DEA w. ML	85.8	18.0	**11.1**	10.5	41.7
StoNEZD	17.4	10.5	**10.3**	10.8	15.3

Monte Carlo simulations

Replicate Banker and Natarajan (2008, *OR*) simulations:

$$y_i = (x_i^3 - 12x_i^2 + 48x_i - 37) \cdot \exp(\delta z_i + v_i - u_i)$$

Baseline scenario:	alternative scenarios
δ = -0.2	(-0.4, -0.6) larger signal
σ_u = 0.15	
σ_v = 0.04	(0, 0.09) no noise vs. heavy noise
n = 200	(50, 100) small sample

Estimation of unexplained inefficiency u_i also considered

Monte Carlo simulations

Alternative scenarios

Root Mean Squared Deviation (RMSD) (%) for δ at different levels of correlation (ρ) between x and z variables

Estimator	ρ=-0.8	ρ=-0.2	**ρ= 0**	ρ=+0.2	ρ=+0.8
OLS: y on z	379	116	**64.1**	112	383
2-DEA	78.0	17.2	**11.3**	11.9	51.1
StoNEZD	12.8	7.95	**9.71**	9.38	15.0

Increased signal: δ = -0.4, σ_u = 0.15, σ_v = 0.04, n = 100.

Conclusions

- Joint estimation of nonparametric frontier and effects of contextual variables possible in the *StoNEZD* method
- Both stochastic noise and correlation between **x** and **z** variables taken into account
- Estimator of **z** variables is consistent, asy. efficient, asy. normal, and converges at standard parametric rate:
  - Conventional standard errors, t-tests, p-values ,… are valid
- StoNEZD performs almost equally well as the correctly specified parametric MLE in MC simulations

Thank you for your attention!

Working paper available at SSRN:
http://ssrn.com/abstract=1485733

Internet: http://www.nomepre.net/stoned/
E-mail: timo.kuosmanen@aalto.fi