DEA as Nonparametric Least Squares Regression & Stochastic Nonparametric Envelopment of Data (StoNED)

Timo Kuosmanen

MTT Agrifood Research Finland

→ *Helsinki School of Economics (HSE)*

Based on papers

1) "Data Envelopment Analysis as Nonparametric Least Squares Regression," with

Andrew Johnson

Texas A&M University

2) "Stochastic Nonparametric Envelopment of Data: Cross-sectional Frontier Estimation Subject to Shape Constraints," with

Mika Kortelainen

Aston Business School

Motivation

- DEA seen as fundamentally different from regression based methods (such as SFA)
- Despite better understanding of statistical foundation of DEA, major conceptual and operational barriers remain
- Vision for future: integrating competing paradigms to a unified approach to productive efficiency analysis

Two separate paradigms

- **Axiomatic approach**
  - DEA
  - FDH
  - Activity analysis

- **Estimation approach**
  - SFA
  - Parametric programming
  - COLS, MOLS

Two separate paradigms

- **Axiomatic approach (DEA)**
  - Axioms of production theory
    - free disposability, convexity, RTS
  - Multi-output production sets
  - Minimum extrapolation principle
  - Measurement of efficiency
  - Linear programming
  - Focus on the frontier

- Microeconomic theory & operations research

Two separate paradigms

- **Estimation approach (SFA)**
  - Toolbox of regression analysis
  - Probabilistic estimation and statistical inference
  - Single-output production function
    - Cobb-Douglas, translog
  - Focus on the unexplained residual
    - Decomposed to inefficiency and noise

- Econometrics & statistics

Challenge by Peter Schmidt

"I am very skeptical of non-statistical measurement exercises, certainly as they are now carried out and perhaps in any way in which they could be carried out.

…

I see no virtue whatever in a non-statistical approach to data."

Schmidt, P. (1985). "Frontier Production Functions." *Econometric Reviews* **4**(2): 289-328.

Response by Rajiv Banker

- Banker, R.D. (1993). "Maximum-Likelihood, Consistency and Data Envelopment Analysis - a Statistical Foundation." *Management Science* **39**(10): 1265-1273.

- DEA estimators are statistically consistent
- DEA provides maximum likelihood (ML) estimators for a broad class of inefficiency distributions
  – including the exponential and half-normal distributions considered by Schmidt (1985).

Statistical foundation of DEA

Subsequent research (mainly by Banker et al. and Simar et al.) has further enhanced the statistical underpinnings of DEA

- Consistency in the multi-output case
- Rates of convergence
- Nonparametric tests
  - e.g. Kolmogorov-Smirnov
- Bootstrapping

ML interpretation of DEA

- ML generally requires the inefficiency terms to be *identically and independently distributed* (i.i.d) according to some specific probability density function

- Incidental parameters problem: the usual asymptotic properties of ML estimators do not apply to DEA estimators!

=> ML interpretation of DEA is conceptually important, but has little practical use

Objectives of paper #1

- Recast DEA as nonparametric least squares regression
  - More direct link than the ML result by Banker

- Develop a new nonparametric least squares estimator (C^2NLS) that is *more efficient* than DEA

- Enhance the statistical foundation of DEA and lower barriers between the axiomatic and estimation paradigms

Classification

	Parametric	*Nonparametric*
Central Tendency	**OLS** Gauss, Legendre,… 1840s	**DEA** Farrell (1957) Charnes et al. (1978)
Deterministic frontier; Sign- Constraints	**PP** Aigner and Chu (1968)	
Deterministic frontier; 2-stage Estimation	**COLS** Richmond (1974) Greene (1980)	
Stochastic frontier; composite error term	**SFA** Aigner et al. (1977) Meeusen & Vanden Broeck (1977)	

Classification

	Parametric	
Central Tendency	**OLS** Gauss, Legendre,…	
Deterministic frontier; Sign- Constraints	**PP** Aigner and Chu (1968)	**DEA** Farrell (1957) Charnes et al. (1978)
Deterministic frontier; 2-stage Estimation	**COLS** Richmond (1974) Greene (1980)	**C^2NLS** Kuosmanen & Johnson (2008)
Stochastic frontier; composite error term	**SFA** Aigner et al. (1977) Meeusen & Vanden Broeck (1977)	**StoNED** Kuosmanen (2006) Kuosmanen & Kortelainen (2007)

Classification

	Parametric	*Non-parametric*
Central Tendency	**OLS** Gauss, Legendre,…	**CNLS** Hildreth (1954) Hanson & Pledger (1976)
Deterministic frontier; Sign- Constraints	**PP** Aigner and Chu (1968)	**DEA** Farrell (1957) Charnes et al. (1978)
Deterministic frontier; 2-stage Estimation	**COLS** Richmond (1974) Greene (1980)	**C^2NLS** Kuosmanen & Johnson (2008)
Stochastic frontier; composite error term	**SFA** Aigner et al. (1977) Meeusen & Vanden Broeck (1977)	**StoNED** Kuosmanen (2006) Kuosmanen & Kortelainen (2007)

Links established

	Parametric	*Non-parametric*	
Central Tendency	**OLS** Gauss, Legendre,…	**CNLS** Hildreth (1954) Hanson & Pledger (1976)	
Deterministic frontier; Sign- Constraints	**PP** Aigner and Chu (1968)	←	**DEA** Farrell (1957) Charnes et al. (1978)
Deterministic frontier; 2-stage Estimation	**COLS** Richmond (1974) Greene (1980)	←	**C^2NLS** Kuosmanen & Johnson (2008)
Stochastic frontier; composite error term	**SFA** Aigner et al. (1977) Meeusen & Vanden Broeck (1977)	←	**StoNED** Kuosmanen (2006) Kuosmanen & Kortelainen (2007)

OLS model & assumptions

$$y_i = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i \ \ \forall i = 1,...,n$$

A1. Linearity: There is a linear relationship between **y** and **X**.

A2. Full rank: columns of **X** are linearly independent and $n \geq K$.

A3. Exogeneity: $E(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \mathbf{0}$. Observations on **x** convey no information about the expected value of $\boldsymbol{\varepsilon}$.

A4. Spherical disturbances: $E(\boldsymbol{\varepsilon\varepsilon}' \mid \mathbf{X}) = \sigma^2\mathbf{I}$

- **A4a. Homoskedasticity:** $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$
- **A4b. Nonautocorrelation:** $\text{Cov}(\varepsilon_h, \varepsilon_j \mid \mathbf{X}) = 0$ for all i,h

OLS problem

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

$$s.t.$$

$$y_i = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i \;\; \forall i = 1,\ldots,n$$

Statistical properties of OLS

Finite sample properties

- OLS estimator is **unbiased**
- OLS estimator is **efficient**: it has the minimum variance among the linear unbiased estimators (Gauss-Markov theorem)

Asymptotic properties

- OLS is **consistent** for ”well-behaved” data
- OLS estimator is **asymptotically efficient**

Parametric programming (Aigner & Chu 1968, AER)

Quadratic

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

$$s.t.$$

$$y_i = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i \;\; \forall i = 1,\ldots,n$$

$$\varepsilon_i \leq 0 \;\; \forall i = 1,\ldots,n$$

Linear

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\varepsilon}} \sum_{l=1}^{n} -\varepsilon_i$$

$$s.t.$$

$$y_i = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i \;\; \forall i = 1,\ldots,n$$

$$\varepsilon_i \leq 0 \;\; \forall i = 1,\ldots,n$$

Parametric programming (Aigner & Chu 1968, AER)

Statistical foundation of PP (Schmidt 1976):

- the PP estimators are MLE if inefficiencies ε_i are i.i.d. and have
  - half-normal distribution (quadratic model)
  - exponential distribution (linear model)

Corrected OLS (COLS)

Simple 2-step approach

Step 1: Apply OLS to estimate the central tendency (average production function)

Step 2: Shift the estimated OLS curve upwards until all observations are enveloped

Note: COLS and PP estimators are not equivalent

Corrected OLS (COLS)

- The idea was first suggested by Winsten, in his comments to Farrell (1957)
- Richmond (1974) suggests this estimator, arguing that it is consistent
- Greene (1980) coins the name, and proves consistency formally

- Statistical inference assuming ε_i are i.i.d. gamma distributed (already considered by Afriat 1972)

Links established

	Parametric	*Non-parametric*
Central Tendency	**OLS** Gauss, Legendre,…	**CNLS** Hildreth (1954) Hanson & Pledger (1976)
Deterministic frontier; Sign- Constraints	**PP** Aigner and Chu (1968)	**DEA** Farrell (1957) Charnes et al. (1978)
Deterministic frontier; 2-stage Estimation	**COLS** Richmond (1974) Greene (1980)	**C^2NLS** Kuosmanen & Johnson (2008)
Stochastic frontier; composite error term	**SFA** Aigner et al. (1977) Meeusen & Vanden Broeck (1977)	**StoNED** Kuosmanen (2006) Kuosmanen and Kortelainen (2007)

Convex Nonparametric Least Squares (CNLS)

- Regression method that
  - builds upon the shape constraints
    - monotonicity, convexity
  - does not require prior assumptions about the functional form or the smoothness of the regression function

Origins of CNLS in Statistics

Hildreth, C. (1954): Point Estimates of Ordinates of Concave Functions. *Journal of the American Statistical Association* 49(267), 598-619.

Hanson, D.L., and G. Pledger (1976): Consistency in concave regression. *Annals of Statistics* 4(6), 1038-1050.

Dykstra, R.L. (1983): An algorithm for restricted least squares regression, *Journal of the American Statistical Association* 78, 837-842.

Nemirovskii, A.S., B.T. Polyak, and A.B. Tsybakov (1985) Rates of Convergence of Nonparametric Estimates of Maximum Likelihood Type, *Problems of Information Transmission* 21, 258-271.

Meyer, M.C. (1999) An Extension of the Mixed Primal-Dual Bases Algorithm to the Case of More Constraints than Dimensions, *J. Statistical Planning and Inference* 81,13-31.

Mammen, E., and C. Thomas-Agnan (1999): Smoothing splines and shape restrictions, *Scandinavian Journal of Statistics* 26, 239-252.

Groeneboom, P., G. Jongbloed, and J.A. Wellner (2001): Estimation of convex functions: characterizations and asymptotic theory, *Annals of Statistics* 29, 1653-1698.

Meyer, M.C. (2003) A Test for Linear vs. Convex Regression Function using Shape-Restricted Regression, *Biometrika* 90(1), 223–232.

Meyer, M.C. (2006) Consistency and Power in Tests with Shape-Restricted Alternatives, *Journal of Statistical Planning and Inference* 136, 3931-3947.

Original illustration by Hildreth

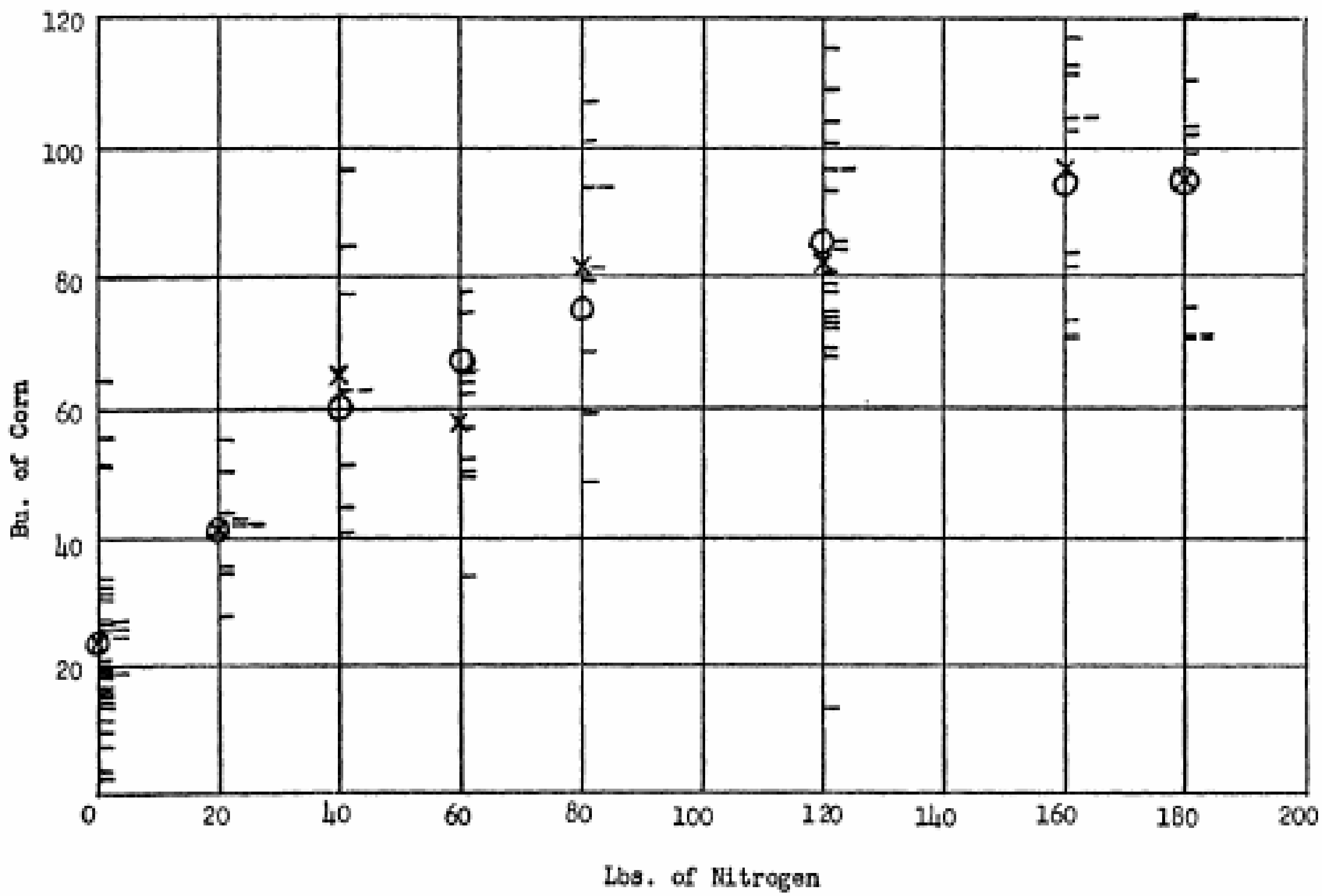


FIG. 1. Observations and Estimates.

— Observed yield (y_{st})

× Mean yield ($\bar{y}_s$)

○ Maximum likelihood estimate of expected yield ($\hat{\eta}_s$)

Modified illustration

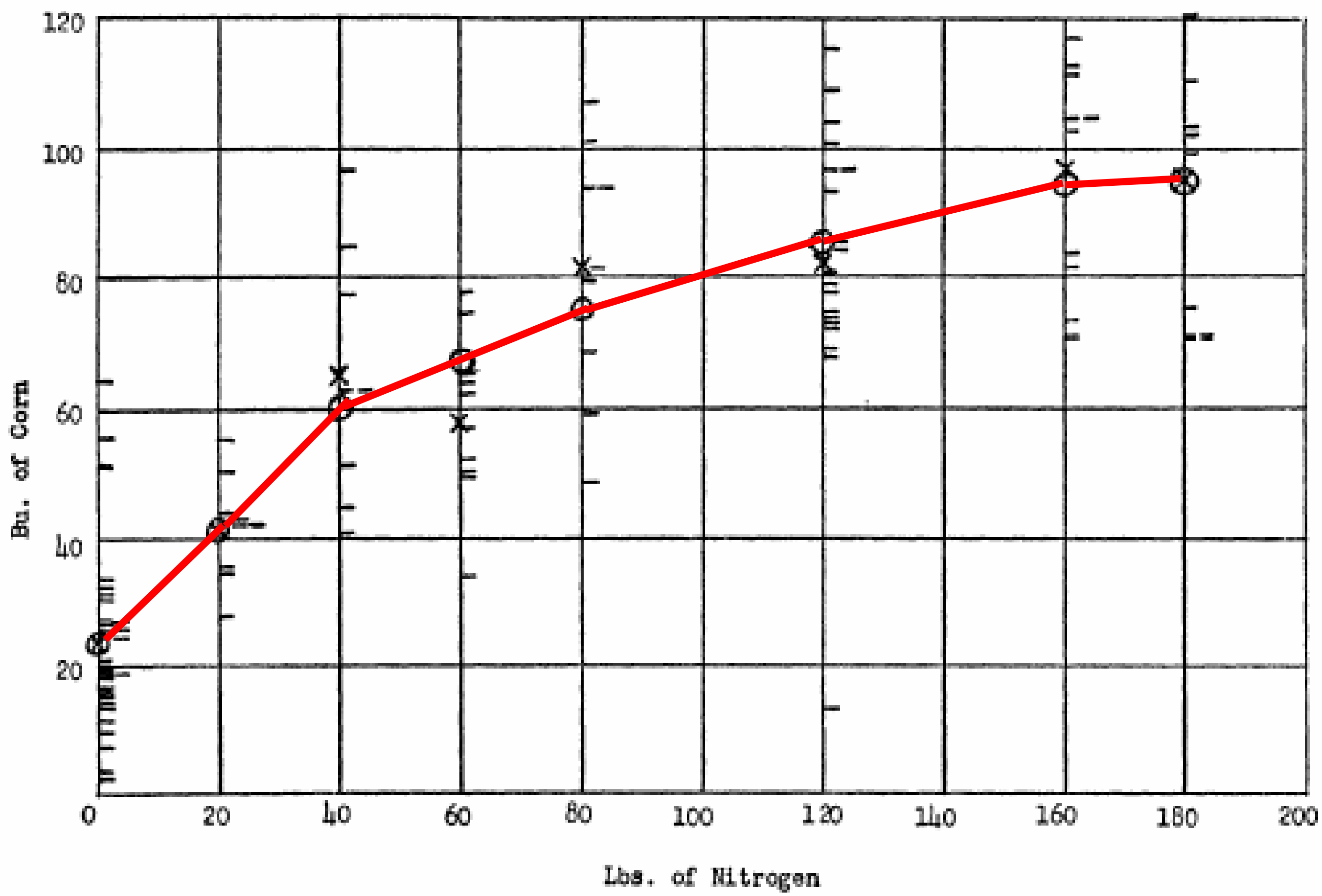


FIG. 1. Observations and Estimates.

— Observed yield (y_{nt})

× Mean yield ($\bar{y}_n$)

○ Maximum likelihood estimate of expected yield ($\hat{\eta}_n$)

CNLS model

Regression model: $y_i = f(\mathbf{x}_i) + \varepsilon_i,\ i = 1,...,n$

Assumptions:

- Function f belongs to the family of monotonic increasing and globally concave functions F_2.

- Errors $\boldsymbol{\varepsilon}$ are uncorrelated random variables with
  - $E(\boldsymbol{\varepsilon})=\mathbf{0}$ (exogeneity)
  - $E(\boldsymbol{\varepsilon\varepsilon}')= \sigma^2\mathbf{I}$ (homoskedasticity, no autocorrelation)

CNLS problem

$$\min_{f \in F_2} \sum_{I=1}^{n} \varepsilon_i^2$$

$$s.t.$$

$$y_i = f(\mathbf{x}_i) + \varepsilon_i$$

CNLS problem a single x case (Hanson & Pledger 1976)

Sort data in ascending order according to x (i.e., $x_1 < x_2 < \ldots < x_n$; assume away ties)

$$\min_{f \in F_2} \sum_{l=1}^{n} \varepsilon_i^2$$

$$s.t.$$

$$y_i = \hat{y}_i + \varepsilon_i \quad \forall i = 1,\ldots,n$$

$$\hat{y}_i \geq \hat{y}_{i-1} \quad \forall i = 2,\ldots,n$$

$$\frac{\hat{y}_i - \hat{y}_{i-1}}{x_i - x_{i-1}} \leq \frac{\hat{y}_{i-1} - \hat{y}_{i-2}}{x_{i-1} - x_{i-2}} \quad \forall i = 3,\ldots,n$$

CNLS problem with multiple x
(Kuosmanen 2008)

$$\min_{\alpha,\boldsymbol{\beta}} \sum_{I=1}^{n} \varepsilon_i^2$$

$s.t.$

$y_i = \alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_i + \varepsilon_i$ (regression equation)

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

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

Representation Theorem

Infinite dimensional problem

$$\min_{f \in F_2} \sum_{l=1}^{n} \varepsilon_i^2$$

$$s.t.$$

$$y_i = f(\mathbf{x}_i) + \varepsilon_i$$

=

Quadratic programming problem

$$\min_{\alpha,\boldsymbol{\beta}} \sum_{l=1}^{n} \varepsilon_i^2$$

$$s.t.$$

$$y_i = \alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_i + \varepsilon_i \quad \text{(regression equation)}$$

$$\boldsymbol{\beta}_i \geq \mathbf{0} \ \forall i = 1,...,n \quad \text{(monotonicity)}$$

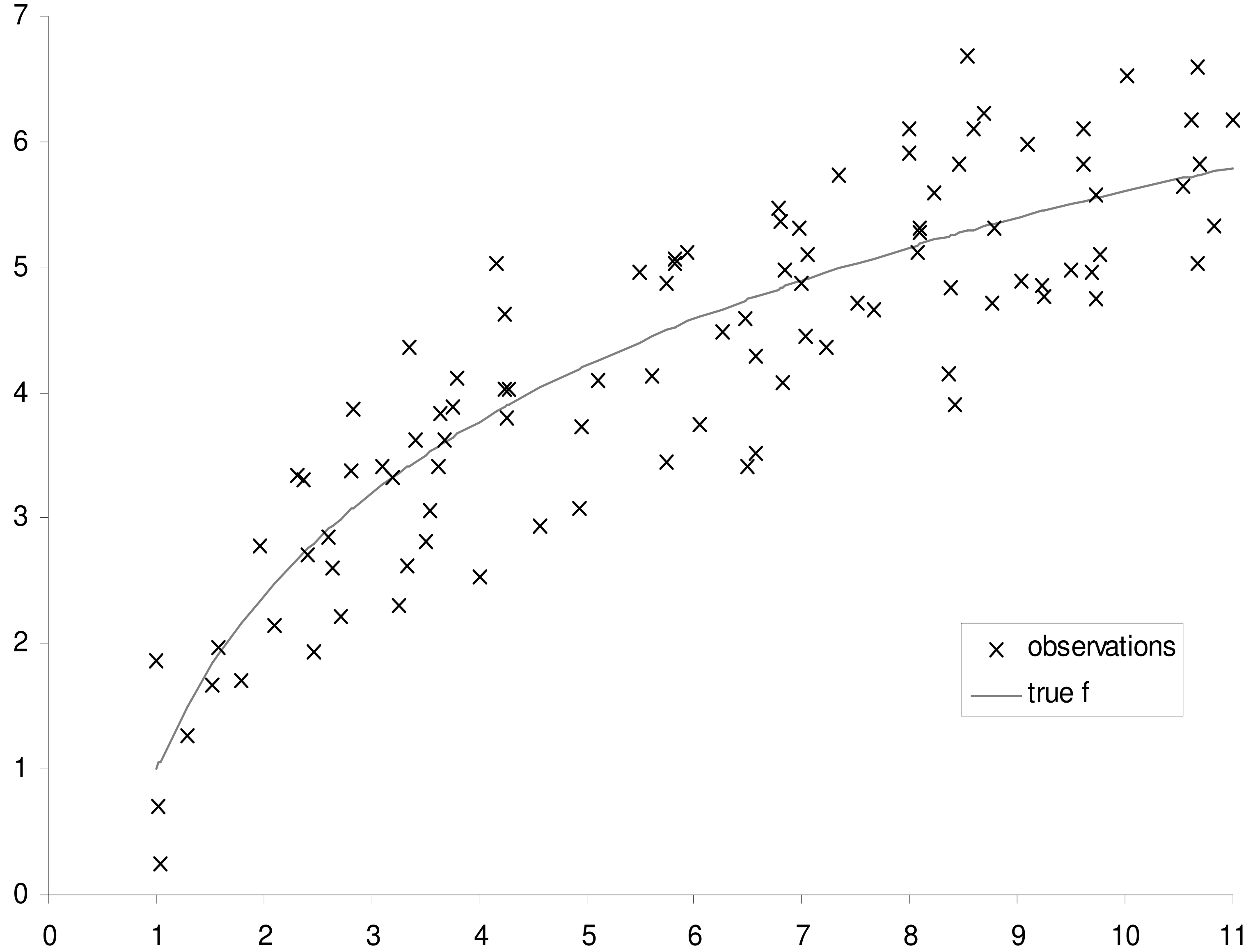
$$\alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_i \leq \alpha_h + \boldsymbol{\beta}_h'\mathbf{x}_i \ \forall h,i = 1,...,n \ \text{(concavity)}$$

Kuosmanen, T. (2008): Representation Theorem for Convex Nonparametric Least Squares, *Econometrics Journal* 11, 308-325.

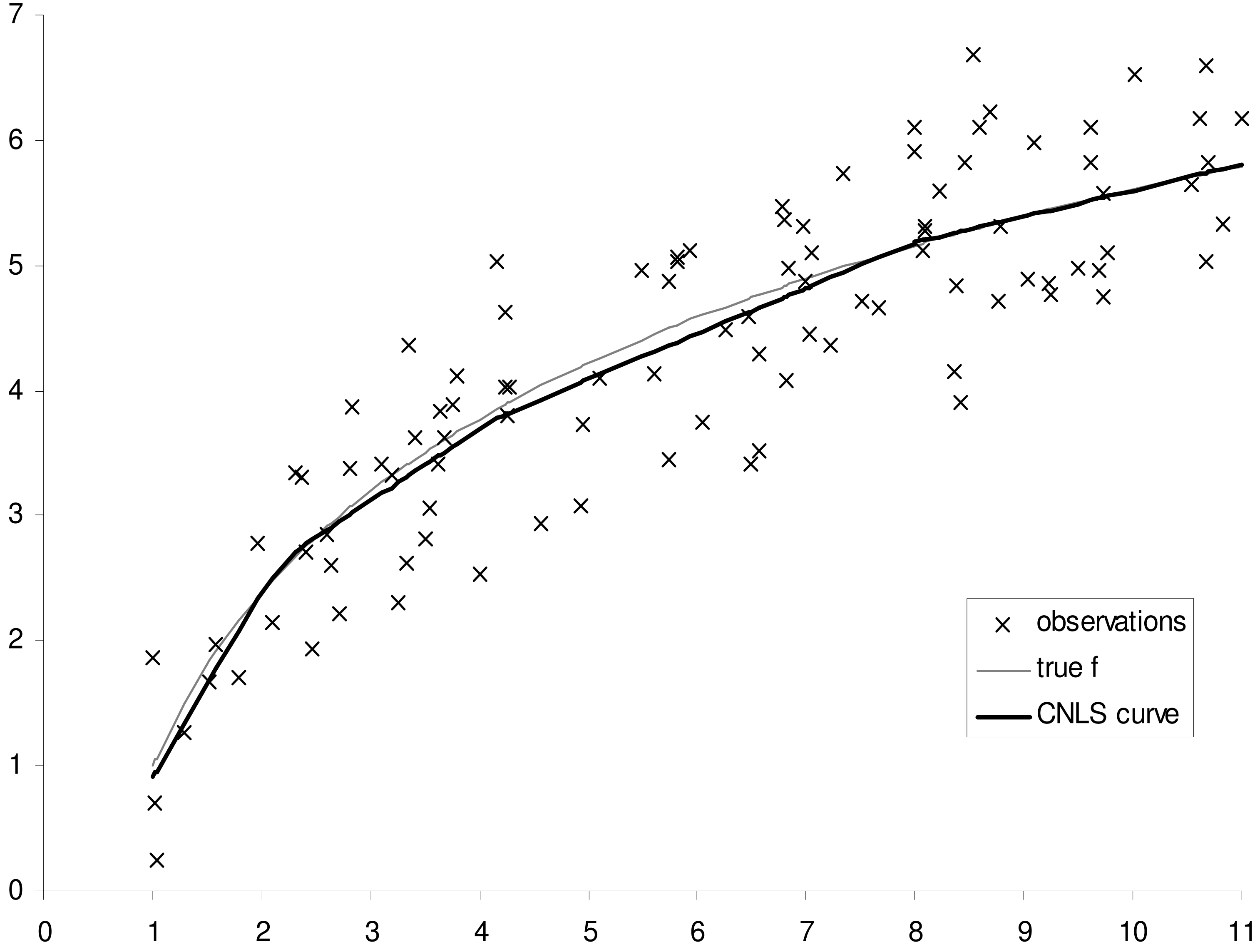
Simulation example

- True production function $y = \ln(x) + 2$
- Inputs x randomly drawn from $Uni[1,11]$.
- Error term randomly drawn from $N(0,0.6^2)$
- Sample size 100.

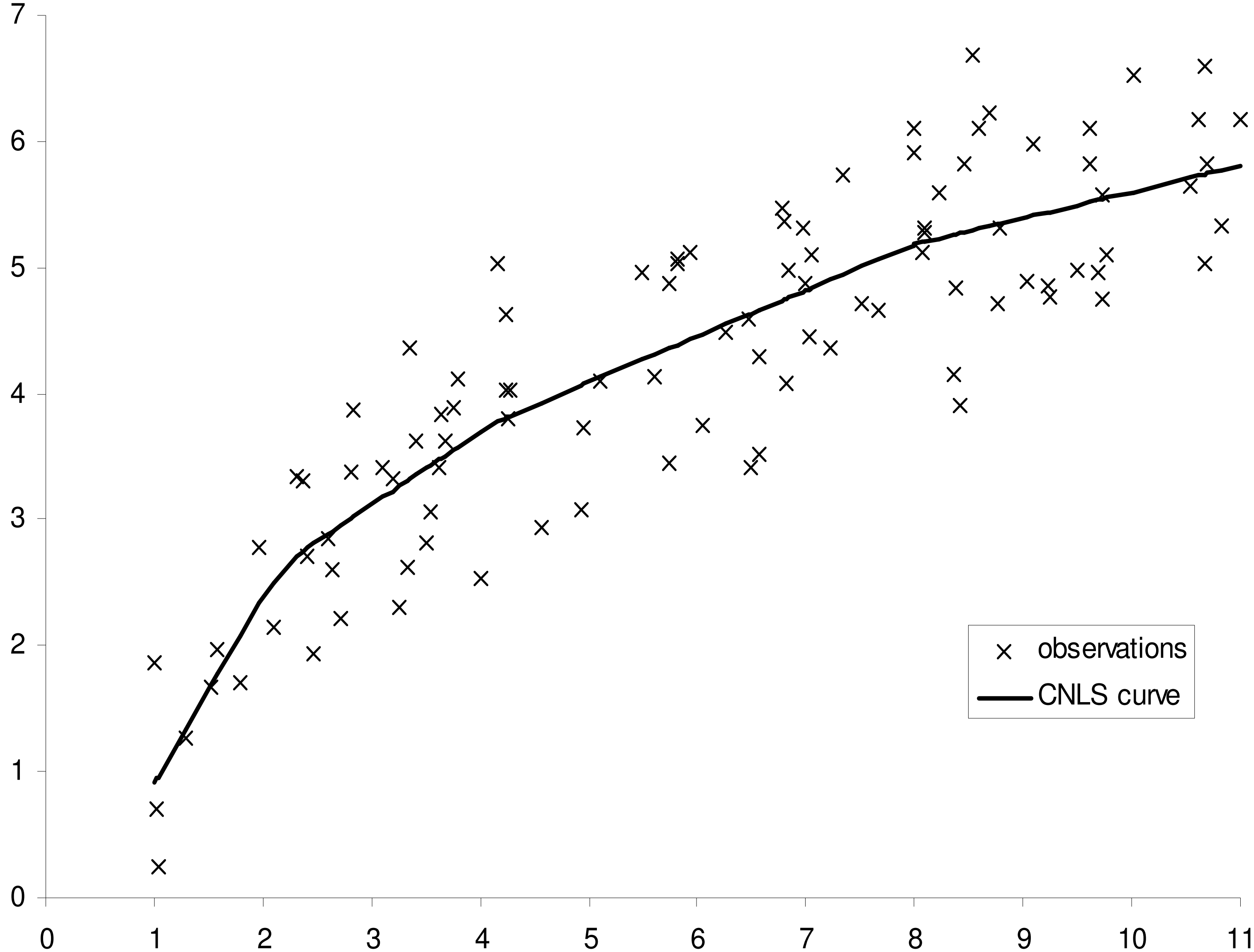
Data and true *f*



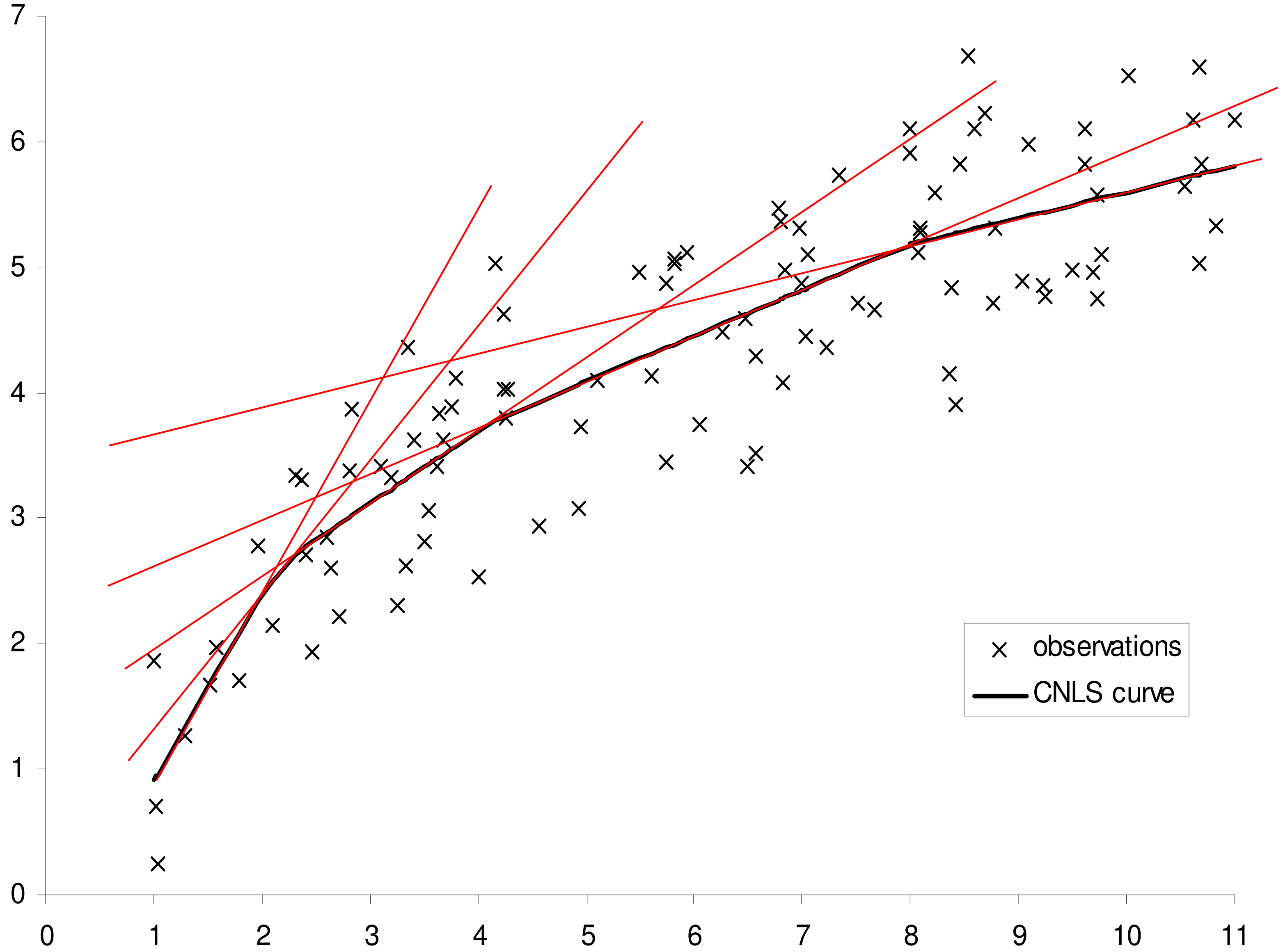
CNLS and true *f*



CNLS regression



CNLS regression



Classification

	Parametric	*Non-parametric*
Central Tendency	**OLS** Gauss, Legendre,…	**CNLS** Hildreth (1954) Hanson & Pledger (1976)
Deterministic frontier; Sign- Constraints	**PP** Aigner and Chu (1968)	**DEA** Farrell (1957) Charnes et al. (1978)
Deterministic frontier; 2-stage Estimation	**COLS** Richmond (1974) Greene (1980)	**C^2NLS** Kuosmanen & Johnson (2008)
Stochastic frontier; composite error term	**SFA** Aigner et al. (1977) Meeusen & Vanden Broeck (1977)	**StoNED** Kuosmanen (2006) Kuosmanen and Kortelainen (2007)

Sign-constrained CNLS

- Recall that PP model (Aigner&Chu) differs from OLS in that an additional sign constraint $\boldsymbol{\varepsilon} \leq \mathbf{0}$ is imposed

- Suppose we impose the same sign constraint in nonparanetric CNLS

  => sign-constrained CNLS

Sign-constrained CNLS

$$\min_{f \in F_2} \sum_{l=1}^{n} \varepsilon_i^2$$

$$s.t.$$

$$y_i = f(\mathbf{x}_i) + \varepsilon_i$$

$$\varepsilon_i \leq 0$$

Sign-constrained CNLS

$$\min_{\alpha,\boldsymbol{\beta}} \sum_{I=1}^{n} \varepsilon_i^2$$

$$s.t.$$

$$y_i = \alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_i + \varepsilon_i \quad \text{(regression equation)}$$

$$\boldsymbol{\beta}_i \geq \mathbf{0} \ \forall i = 1,...,n \quad \text{(monotonicity)}$$

$$\alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_i \leq \alpha_h + \boldsymbol{\beta}_h'\mathbf{x}_i \ \forall h,i = 1,...,n \quad \text{(concavity)}$$

$$\varepsilon_i \leq 0 \ \forall i = 1,...,n \quad \text{(sign constraint)}$$

DEA

- VRS frontier (Afriat 1972; Banker 1993):

$$f^{DEA}(\mathbf{x}) = \max_{\lambda \in \mathbb{R}_+^n} \left\{ y \,\middle|\, y \le \sum_{h=1}^{n} \lambda_h y_h ; \mathbf{x} \ge \sum_{h=1}^{n} \lambda_h \mathbf{x}_h ; \sum_{h=1}^{n} \lambda_h = 1 \right\}$$

- Additive output efficiency:

$$\max_{\boldsymbol{\lambda} \ge \mathbf{0}} \phi$$

$$y_i + \phi \le \sum_{h=1}^{n} \lambda_h y_h$$

$$\mathbf{x}_i \ge \sum_{h=1}^{n} \lambda_h \mathbf{x}_h$$

$$\sum_{h=1}^{n} \lambda_h = 1$$

Main result

Theorem 3.1: *For all real valued data, the sign-constrained nonparametric least squares problem is equivalent to the output-oriented DEA VRS problem.*

Both measure efficiency relative to the same DEA frontier.

$$f^{DEA}(\mathbf{x}) = \max_{\lambda \in \mathbb{R}_+^n} \left\{ y \,\middle|\, y \leq \sum_{h=1}^{n} \lambda_h y_h ; \mathbf{x} \geq \sum_{h=1}^{n} \lambda_h \mathbf{x}_h ; \sum_{h=1}^{n} \lambda_h = 1 \right\}$$

Classification

	Parametric	*Non-parametric*
Central Tendency	**OLS** Gauss, Legendre,…	**CNLS** Hildreth (1954) Hanson & Pledger (1976)
Deterministic frontier; Sign- Constraints	**PP** Aigner and Chu (1968)	**DEA** Farrell (1957) Charnes et al. (1978)
Deterministic frontier; 2-stage Estimation	**COLS** Richmond (1974) Greene (1980)	**C^2NLS** Kuosmanen & Johnson (2008)
Stochastic frontier; composite error term	**SFA** Aigner et al. (1977) Meeusen & Vanden Broeck (1977)	**StoNED** Kuosmanen (2006) Kuosmanen and Kortelainen (2007)

Corrected Convex Nonparametric Least Squares (C²NLS)

- Nonparametric analogue to COLS
  - Just replace OLS by CNLS

- Two-step approach:

Step 1: Apply CNLS to estimate the central tendency (average production function)

Step 2: Shift the estimated CNLS curve upwards until all observations are enveloped

Corrected Convex Nonparametric Least Squares (C^2NLS)

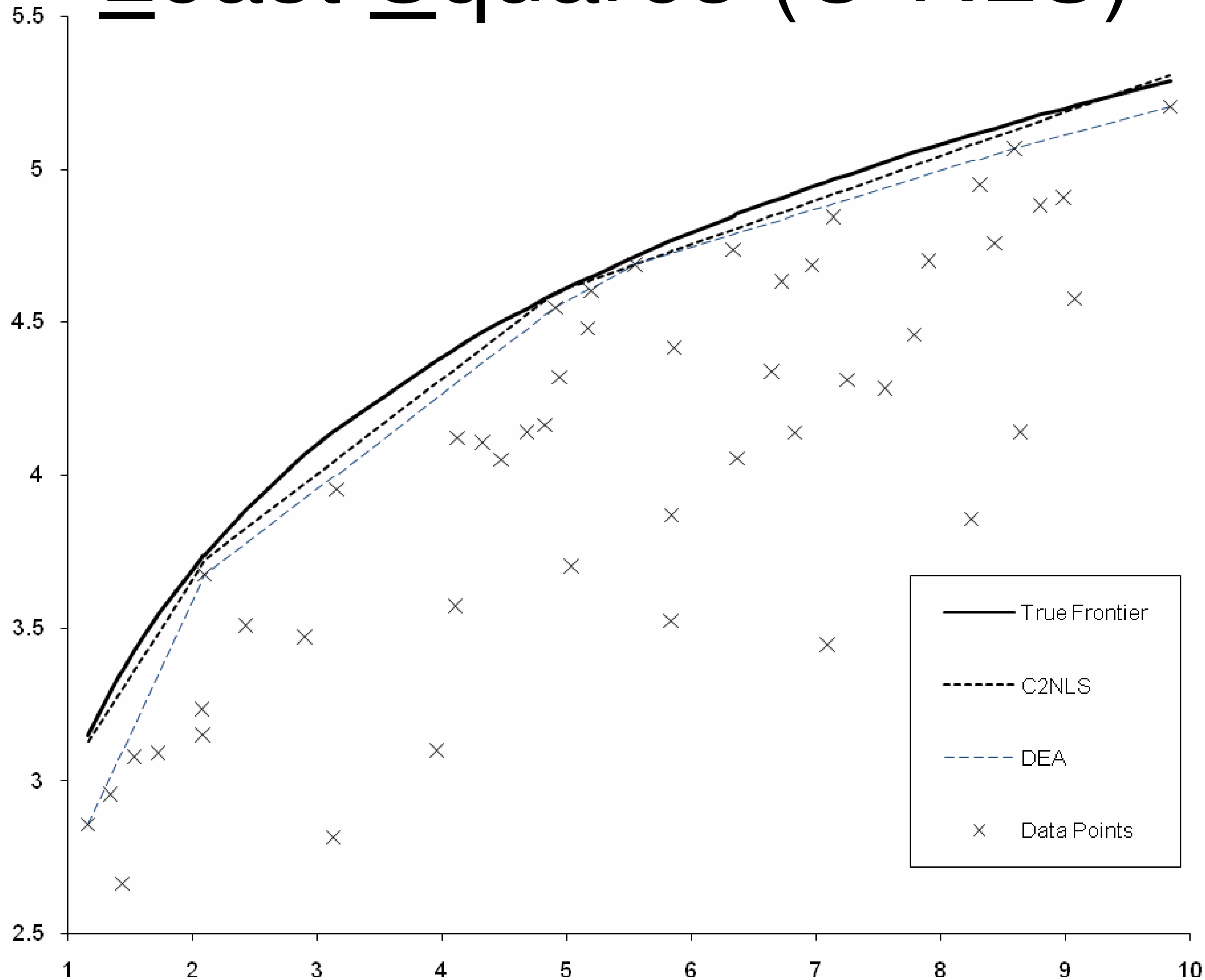
Theorem 4.1: *For any sequence of independent observations* **X**,**y** *generated by production function f and i.i.d inefficiency terms* **ε** *that are uncorrelated with* **X** *and have a positive density at* **ε**=**0**, *the C^2NLS estimator is statistically consistent.*

Corrected Convex Nonparametric Least Squares (C²NLS)

***Theorem 4.2**: For any real valued data set, the discriminatory power of C²NLS is always greater than or equal to that of DEA in the sense that*

$$\hat{\varepsilon}_i^{C2NLS} \leq \varepsilon_i^{DEA} \leq 0 \ \forall i = 1,\ldots,n$$

Corrected Convex Nonparametric Least Squares (C²NLS)



Monte Carlo simulations

- 6 scenarios, n = 50, 100, 150

	Inputs	Functional Form
A)	x	$y = \ln(x)+3 - u$
B)	x	$y = 3 + x^{1/2} + \ln(x) - u$
C)	x_1, x_2	$y = 0.1x_1 + 0.1x_2 + 0.3(x_1x_2)^{1/2} - u$
D)	x_1, x_2, x_3	$y = 0.1x_1 + 0.1x_2 + 0.1x_3 + 0.3(x_1x_2x_3)^{1/3} - u$
E)	x_1, x_2	$y = 0.1x_1 + 0.1x_2 + 0.3(x_1x_2)^{1/3} - u$
F)	x_1, x_2, x_3	$y = 0.1x_1 + 0.1x_2 + 0.1x_3 + 0.3(x_1x_2x_3)^{1/4} - u$

Monte Carlo simulations

- 4 methods: C^2NLS, DEA, PP, and COLS
- 100 replications of each scenario
- Performance criteria

$$\text{BIAS} = \sum_{t=1}^{M}\sum_{i=1}^{n}\left(\hat{f}(x_i) - f(x_i)\right) \Big/ nM$$

$$\text{MSE} = \sum_{t=1}^{M}\sum_{i=1}^{n}\left(\hat{f}(x_i) - f(x_i)\right)^2 \Big/ nM$$

Results

Scenario	Number of Obs.	Mean Squared Error PP	COLS	DEA	C2NLS	Fraction of Trials DEA > C2NLS	Bias PP	COLS	DEA	C2NLS	Fraction of Trials abs(DEA) > abs(C2NLS)
A)	50	0.003	0.014	0.009	0.005	0.76	0.007	0.062	-0.068	0.015	0.87
A)	100	0.003	0.012	0.004	0.004	0.47	0.018	0.071	-0.040	0.034	0.62
A)	150	0.003	0.014	0.003	0.004	0.49	0.025	0.079	-0.035	0.032	0.62
B)	50	0.002	0.020	0.011	0.008	0.66	-0.003	0.077	-0.074	0.028	0.81
B)	100	0.001	0.019	0.005	0.006	0.46	0.007	0.089	-0.048	0.043	0.62
B)	150	0.001	0.022	0.003	0.005	0.36	0.009	0.103	-0.038	0.038	0.52
C)	50	0.038	0.132	0.032	0.013	0.95	0.115	0.283	-0.130	0.003	0.96
C)	100	0.057	0.208	0.019	0.011	0.87	0.180	0.387	-0.092	0.039	0.83
C)	150	0.068	0.242	0.013	0.009	0.72	0.210	0.428	-0.074	0.049	0.68
D)	50	0.027	0.085	0.070	0.025	0.99	0.069	0.216	-0.207	-0.024	0.99
D)	100	0.041	0.133	0.046	0.017	0.98	0.138	0.306	-0.162	0.023	0.97
D)	150	0.052	0.195	0.035	0.015	0.97	0.173	0.386	-0.138	0.049	0.97
E)	50	0.015	0.054	0.030	0.015	0.91	0.047	0.170	-0.121	0.018	0.94
E)	100	0.022	0.079	0.016	0.011	0.85	0.101	0.233	-0.083	0.044	0.80
E)	150	0.026	0.084	0.011	0.008	0.72	0.118	0.247	-0.068	0.046	0.71
F)	50	0.022	0.069	0.061	0.022	0.99	0.058	0.198	-0.191	-0.015	0.99
F)	100	0.032	0.101	0.041	0.017	0.95	0.117	0.265	-0.148	0.038	0.94
F)	150	0.037	0.110	0.030	0.013	0.94	0.142	0.293	-0.124	0.043	0.92

Robustness to noise

- Scenario D with 3 inputs, $n = 100$

- Stochastic frontier model:

  $y = \mathrm{f}(\mathbf{x}) - u + v$

  u is i.i.d. half-normal inefficiency term

  v is i.i.d. normal inefficiency term

- Signal to noise $\tilde{\lambda} = \sigma_u / \sigma_v$

Robustness to noise

Scenario	Mean Squared Error (MSE)				Fraction of Trials	Bias				Fraction of Trials
Lambda	PP	COLS	DEA	C2NLS	DEA > C2NLS	PP	COLS	DEA	C2NLS	abs(DEA) > abs(C2NLS)
0.83	0.33	0.64	0.08	0.27	0.00	0.53	0.75	0.12	0.48	0.00
1.66	0.15	0.32	0.04	0.09	0.1	0.34	0.51	-0.02	0.26	0.01
2.49	0.09	0.25	0.04	0.05	0.42	0.24	0.45	-0.08	0.17	0.2
3.32	0.06	0.2	0.04	0.03	0.78	0.2	0.39	-0.12	0.1	0.57
4.15	0.06	0.18	0.04	0.03	0.85	0.18	0.37	-0.13	0.09	0.72

Robustness to noise

Scenario	Rank Correlation		Fraction of Trials
Lamda	DEA	C2NLS	DEA > C2NLS
0.83	0.26	0.37	0.96
1.66	0.43	0.58	1.00
2.49	0.50	0.70	1.00
3.32	0.55	0.76	1.00
4.15	0.59	0.79	1.00

Classification

	Parametric	*Non-parametric*
Central Tendency	**OLS** Gauss, Legendre,…	**CNLS** Hildreth (1954) Hanson & Pledger (1976)
Deterministic frontier; Sign- Constraints	**PP** Aigner and Chu (1968)	**DEA** Farrell (1957) Charnes et al. (1978)
Deterministic frontier; 2-stage Estimation	**COLS** Richmond (1974) Greene (1980)	**C^{2}NLS** Kuosmanen & Johnson (2008)
Stochastic frontier; composite error term	**SFA** Aigner et al. (1977) Meeusen & Vanden Broeck (1977)	**StoNED** Kuosmanen (2006) Kuosmanen and Kortelainen (2007)

SFA model

- Estimated equation

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

- Production function is linear in parameters **β**
- Distributional assumptions

$$u_i \underset{iid}{\sim} \left| N\left(0, \sigma_u^2\right) \right|$$

$$v_i \underset{iid}{\sim} N\left(0, \sigma_v^2\right)$$

Estimation of SFA model

SFA model can be estimated by

- Maximum likelihood (ML)
- Modified OLS (MOLS)
  1) Estimate central tendency by OLS
  2) Estimate $E(u)$ based on OLS residuals and shift the frontier upward

- In both methods, the firm-specific efficiency scores are obtained by using the conditional expectation $E(u_i / \varepsilon_i)$ (Jondrow et al. 1982)

- Monte Carlo simulations by Olson et al. (1980): MOLS and ML almost equally efficient

Classification

	Parametric	*Non-parametric*
Central Tendency	**OLS** Gauss, Legendre,…	**CNLS** Hildreth (1954) Hanson & Pledger (1976)
Deterministic frontier; Sign- Constraints	**PP** Aigner and Chu (1968)	**DEA** Farrell (1957) Charnes et al. (1978)
Deterministic frontier; 2-stage Estimation	**COLS** Richmond (1974) Greene (1980)	**C^2NLS** Kuosmanen & Johnson (2008)
Stochastic frontier; composite error term	**SFA** Aigner et al. (1977) Meeusen & Vanden Broeck (1977)	**StoNED** Kuosmanen (2006) Kuosmanen and Kortelainen (2007)

StoNED model

- Estimated equation

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

- *f* is monotonic increasing and concave
- Distributional assumptions

$$u_i \underset{iid}{\sim} \left| N\left(0, \sigma_u^2\right) \right|$$

$$v_i \underset{iid}{\sim} N\left(0, \sigma_v^2\right)$$

Estimation of StoNED model

In principle, StoNED model can be estimated by

- Maximum likelihood (ML)
  - Banker & Maindiratta 1992, JPA
- Modified CNLS
  1) Estimate central tendency by CNLS
  2) Estimate $E(u)$ based on CNLS residuals and shift the frontier upward

- In both methods, the firm-specific efficiency scores must be obtained by using the conditional expectation $E(u_i/\varepsilon_i)$ (Jondrow et al. 1982)

ML problem

- Banker & Maindiratta (1992): *JPA* 3, 401-415.

$$\max_{\substack{y_1^f,\ldots,y_n^f \\ \sigma,\lambda}} \quad \frac{n}{2}\ln(2/\pi) - n\ln\sigma + \sum_{i=1}^{n}\ln\Phi\left[\frac{-\left(y_i - y_i^f\right)\lambda}{\sigma}\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - y_i^f\right)^2$$

$$s.t.$$

$$y_i^f - \boldsymbol{\beta}_i'\mathbf{x}_i \geq y_j^f - \boldsymbol{\beta}_i'\mathbf{x}_j \;\; \forall i,j = 1,\ldots,n \quad \text{(concavity)}$$

$$\boldsymbol{\beta}_i \geq 0 \;\; \forall i = 1,\ldots,n, \quad \text{(monotonicity)}$$

$$y_i^f \geq 0 \;\; \forall i = 1,\ldots,n; \;\; \sigma,\lambda \geq 0.$$

Least squares estimation of StoNED model

In two stages:

Stage 1: Estimate the conditional expectation $E(y_i / \mathbf{x}_i)$ by CNLS

Stage 2: Given the CNLS residuals, estimate the variance parameters of the inefficiency and error distribution by either

- method of moments
- maximum pseudolikelihood

Method of moments

Estimate the variance parameters σ_v^2, σ_u^2 based on the 2nd and 3rd moments of the residual distribution.

$$\hat{\sigma}_u = \sqrt[3]{\frac{\sum_{i=1}^{n}\left(e_i - \sum_{i=1}^{n}(e_i / n)\right)^3}{n\left(\sqrt{\frac{2}{\pi}}\right)\left[1-\frac{4}{\pi}\right]}}$$

$$\hat{\sigma}_v = \sqrt{\frac{\sum_{i=1}^{n}\left(e_i - \sum_{i=1}^{n}(e_i / n)\right)^2}{n} - \left[\frac{\pi-2}{\pi}\right]\hat{\sigma}_u^2}$$

Maximum pseudolikelihood

Based on Fan et al. (1996, JBES).

Estimate parameters σ_v^2, σ_u^2 by maximizing the concentrated likelihood function $\ln(\lambda)$:

$$\max_{\lambda} \ln L(\lambda) = \max_{\lambda} \left\{ -n \ln \hat{\sigma} + \sum_{i=1}^{n} \ln \Phi \left[\frac{-\hat{\varepsilon}_i \lambda}{\hat{\sigma}} \right] - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \right\}$$

$$\hat{\varepsilon}_i = e_i - \left(\sqrt{2}\lambda\hat{\sigma}\right) \Big/ \left[\pi\left(1+\lambda^2\right)\right]^{1/2}$$

$$\hat{\sigma} = \left\{ \frac{1}{n} \sum_{j=1}^{n} e_i^2 \Big/ \left[1 - \frac{2\lambda^2}{\pi(1+\lambda)} \right] \right\}^{1/2}$$

Estimating inefficiency

Expected value of inefficiency

$$E(u_i) = \hat{\mu} = \hat{\sigma}_u \sqrt{2/\pi}$$

Conditional expected value of DMU *i's* inefficiency term is obtained by the Jondrow et al. formula

$$\hat{E}(u_i \left| \hat{\varepsilon}_i \right.) = -\frac{\hat{\varepsilon}_i \hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_v^2} + \frac{\hat{\sigma}_u^2 \hat{\sigma}_v^2}{\hat{\sigma}_u^2 + \hat{\sigma}_v^2} \left[\frac{\phi(\hat{\varepsilon}_i / \hat{\sigma}_v^2)}{1 - \Phi(\hat{\varepsilon}_i / \hat{\sigma}_v^2)} \right]$$

Estimating inefficiency

Note:

- Distributional assumptions regarding *u* and *v* do not influence the relative ranking of units
  - StoNED and C^2NLS methods yield exactly the same efficiency rankings

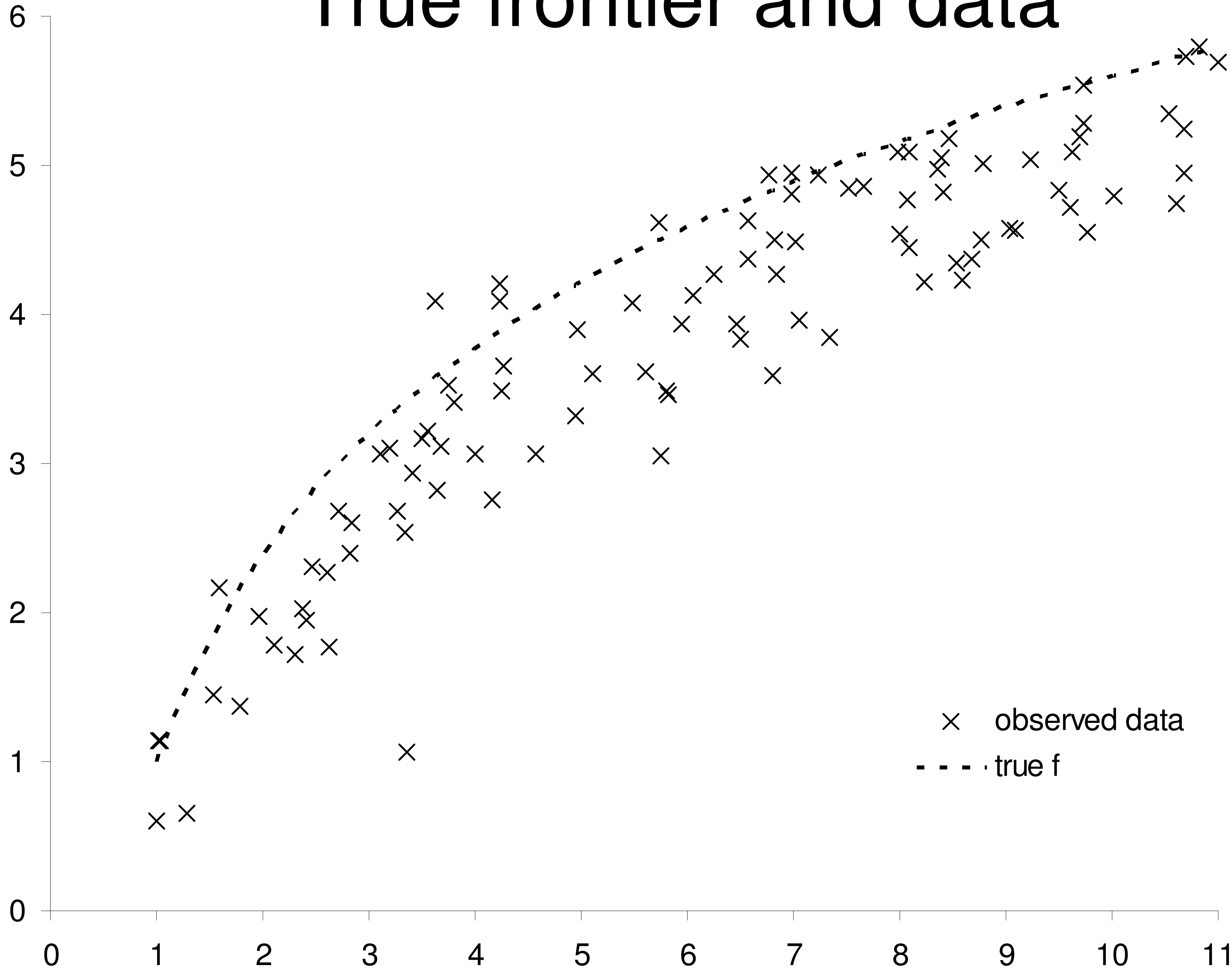
- Level of inefficiency depends on the distributional assumptions

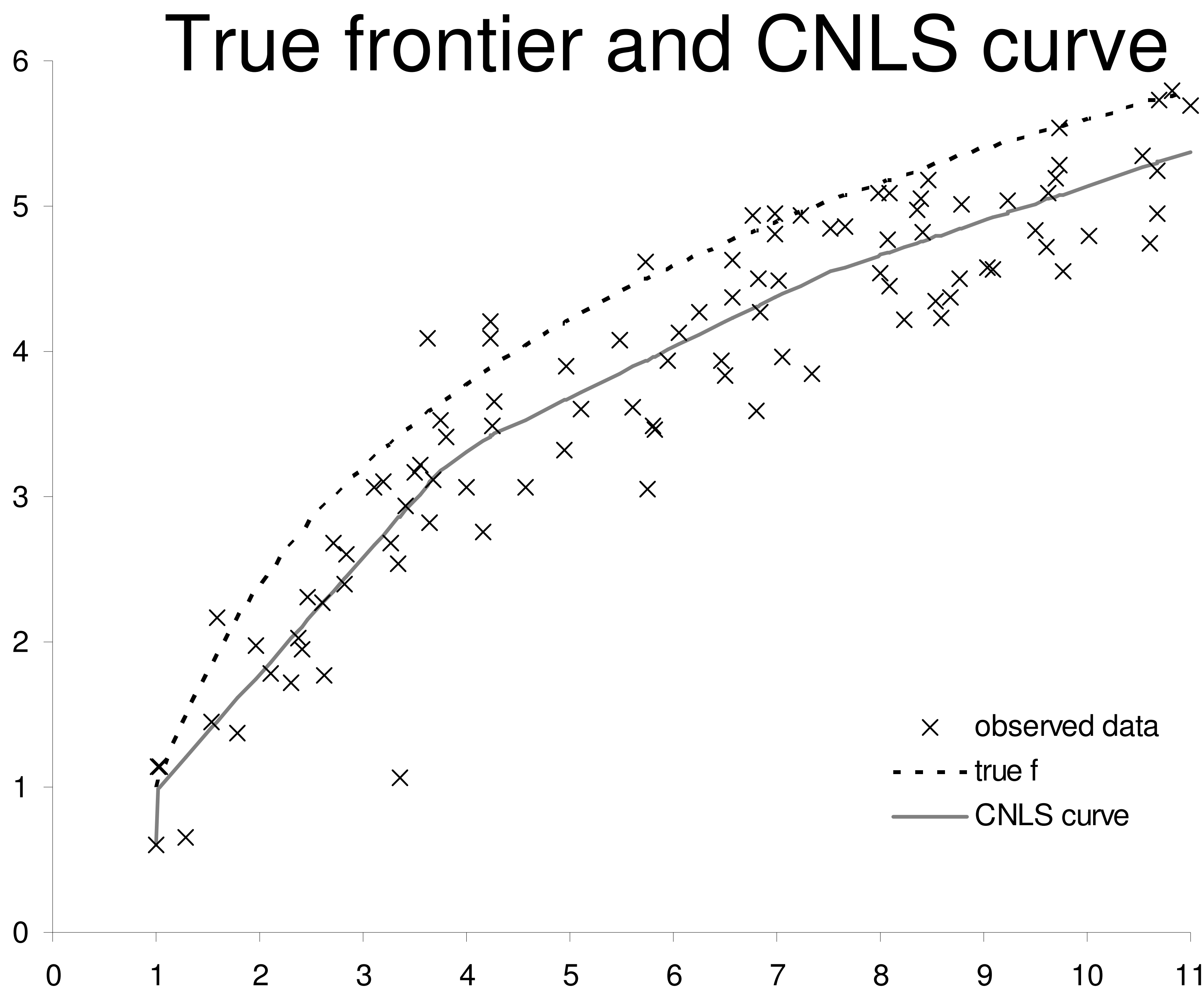
- In panel data setting, no distributional assumptions are necessary

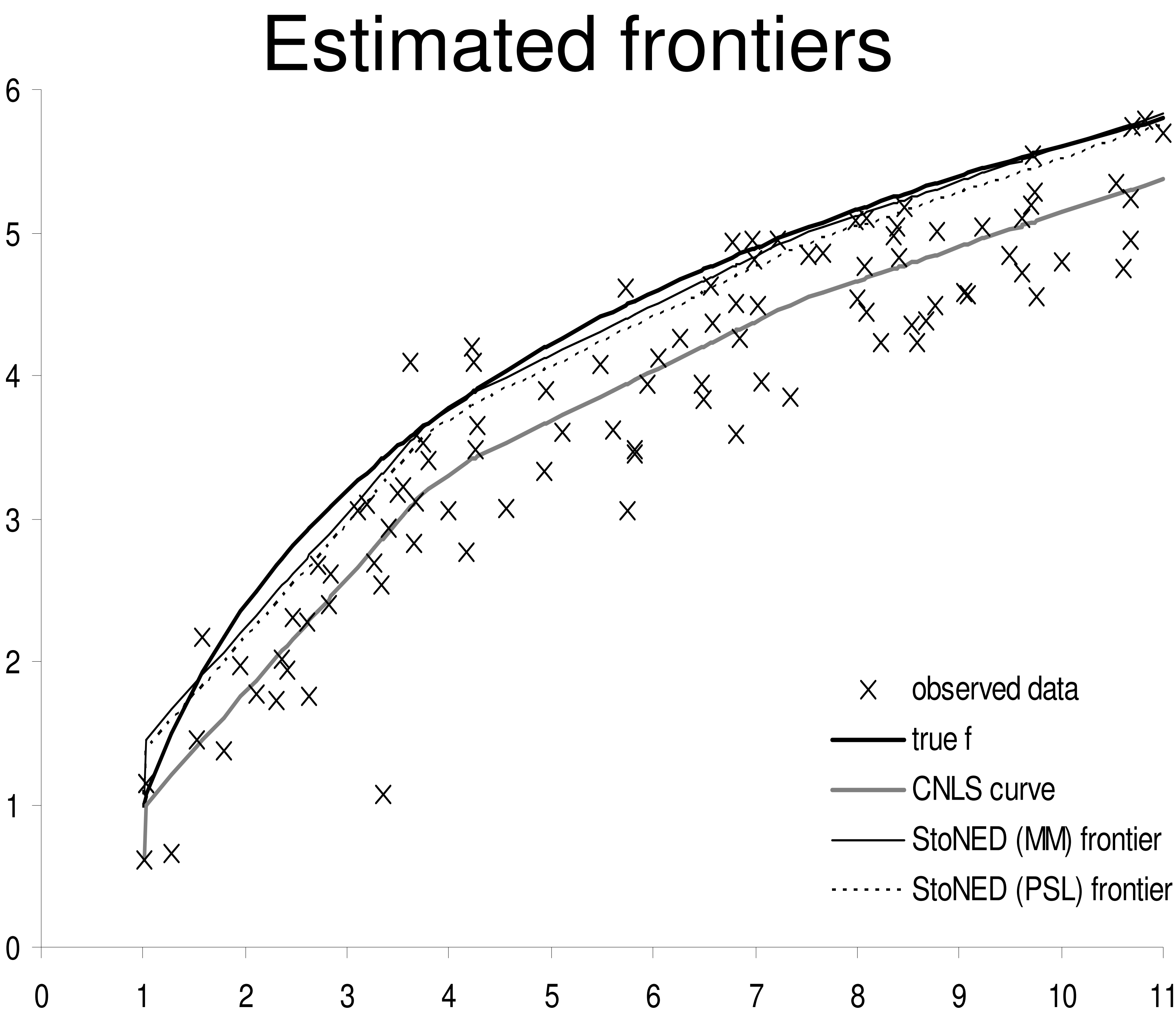
Simulation example

- Production function $y = \ln(x) + 2$
- Inputs x randomly drawn from $Uni[1,11]$.
- Inefficiency randomly drawn from $|N(0,0.6^2)|$
- Error term randomly drawn from $N(0,0.3^2)$
- Sample size 100.

True frontier and data




True frontier and CNLS curve
0
1
2
3
4
5
6
0
1
2
3
4
5
6
7
8
9
10
11
observed data
true f
CNLS curve


Estimated frontiers
0
1
2
3
4
5
6
0
1
2
3
4
5
6
7
8
9
10
11
observed data
true f
CNLS curve
StoNED (MM) frontier
StoNED (PSL) frontier

Monte Carlo simulations

		scenario A 1 input N=50		scenario B 1 input N=100		scenario C 2 inputs N=100		scenario D 3 inputs N=100	
		MSE	**BIAS**	**MSE**	**BIAS**	**MSE**	**BIAS**	**MSE**	**BIAS**
StoNED	MM	0,030	-0,105	0,010	-0,049	0,081	-0,266	0,296	-0,484
	PSL	0,022	-0,055	0,025	-0,132	0,085	-0,272	0,417	-0,594
SFA	CD	0,102	-0,023	0,189	-0,243	0,361	-0,544	0,092	0,190
	Trnslg	0,101	-0,023	0,110	-0,307	0,371	-0,547	0,145	0,047
DEA	CRS	21,594	3,868	23,490	3,980	2,571	1,313	1,078	0,830
	VRS	0,097	0,260	0,151	0,364	1,065	0,802	0,581	0,427
semiparam. kernel		0,297	0,508	0,075	-0,093	0,103	0,272	0,943	-0,850

Conclusions

- DEA can be recast and understood as a nonparametric regression method
  - Barriers between DEA and regression analysis are lower than earlier assumed
- DEA is a nonparametric generalization of PP (Aigner & Chu 1968)
- Parallel development of parametric and nonparametric models, including stochastic StoNED model

Parallel development

	Parametric	*Non-parametric*
Central Tendency	**OLS** Gauss, Legendre,…	**CNLS** Hildreth (1954) Hanson & Pledger (1976)
Deterministic frontier; Sign- Constraints	**PP** Aigner and Chu (1968)	**DEA** Farrell (1957) Charnes et al. (1978)
Deterministic frontier; 2-stage Estimation	**COLS** Richmond (1974) Greene (1980)	**C^2NLS** Kuosmanen & Johnson (2008)
Stochastic frontier; composite error term	**SFA** Aigner et al. (1977) Meeusen & Vanden Broeck (1977)	**StoNED** Kuosmanen (2006) Kuosmanen and Kortelainen (2007)

Conclusions

- In contrast to the ML interpretation (Banker), the least squares interpretation of DEA can be utilized in many ways
- Examples
  - More efficient C^2NLS estimator in the deterministic setting
  - Probabilistic treatment of inefficiency and noise in the stochastic setting (StoNED)
  - Avoiding problems in two-step semiparametric estimation of contextual variables **z** that influence efficiency (cf. critique by Simar & Wilson)
    - 1-stage DEA regression incl. both **x** and **z**
    - C^2NLS and StoNED models with **z** variables

Conclusions

- StoNED model melds together
  - Nonparametric frontier of DEA ($f(\mathbf{x})$)
  - Stochastic composite error of SFA ($\varepsilon_i = v_i - u_i$)

- Least-squares interpretation of DEA enables estimation of StoNED model in practice

- Combining the virtues of SFA and DEA is possible
  - New opportunities as well as challenges

Immediate extensions

- returns to scale
  - VRS, CRS, NIRS, NDRS
- cost functions, distance functions, etc.
- statistical inference by bootstrapping
- quantile estimators
  - order-*m* C^2NLS frontiers
- panel data models
  - fixed and random effects

Further work needed…

- Finite sample properties of CNLS
  - unbiasedness, efficiency?
- Nonradial noise in multi-output setting
- Efficient computational strategies for solving the CNLS problem
- Quasiconcavity, non-convexities
- Modelling endogeneity
  - nonparametric GMM
- Modelling heteroskedasticity and autocorrelation
- etc., etc..

Thank you for your attention!

See the StoNED homepage:

http://www.nomepre.net/stoned

for papers, computer codes, the latest news etc.

- Questions and comments are welcome:
  - E-mail: Timo.Kuosmanen@mtt.fi

Proof of Theorem 3.1

Consider the output-oriented DEA VRS model with additive efficiency (Afriat 1972; Banker 1993)

$$\max_{\boldsymbol{\lambda} \geq \mathbf{0}} \phi$$

$$y_i + \phi \leq \sum_{h=1}^{n} \lambda_h y_h$$

$$\mathbf{x}_i \geq \sum_{h=1}^{n} \lambda_h \mathbf{x}_h$$

$$\sum_{h=1}^{n} \lambda_h = 1$$

Proof of Theorem 3.1

Step 1: Derive equivalent dual (multiplier) problem

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}}(-\varepsilon)$$

$$s.t.$$

$$y_i = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon$$

$$y_h \le \alpha + \boldsymbol{\beta}'\mathbf{x}_h \ \ \forall h = 1,\ldots,n$$

$$\boldsymbol{\beta} \ge \mathbf{0}$$

Proof of Theorem 3.1

Step 2: Sum over all observations *i*

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \sum_{i=1}^{n} -\varepsilon_i$$

$$s.t.$$

$$y_i = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i$$

$$y_h \leq \alpha + \boldsymbol{\beta}'\mathbf{x}_h \;\; \forall h = 1,\ldots,n$$

$$\boldsymbol{\beta} \geq \mathbf{0}$$

Proof of Theorem 3.1

Step 3: Add sign constraint $\varepsilon_i \leq 0$

$$
\begin{aligned}
&\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \sum_{i=1}^{n} -\varepsilon_i \\
&s.t. \\
&y_i = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i \\
&y_h \leq \alpha + \boldsymbol{\beta}'\mathbf{x}_h + \varepsilon_h \ \ \forall h = 1,\ldots,n \\
&\boldsymbol{\beta} \geq \mathbf{0} \\
&\varepsilon_i \leq 0 \ \forall i = 1,\ldots,n
\end{aligned}
$$

Proof of Theorem 3.1

Step 4: Insert y_h into the concavity constraint

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \sum_{i=1}^{n} -\varepsilon_i$$

$$s.t.$$

$$y_i = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i$$

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

$$\boldsymbol{\beta} \ge \mathbf{0}$$

$$\varepsilon_i \le 0 \;\; \forall i = 1,\ldots,n$$

Proof of Theorem 3.1

Step 5: apply quadratic transformation to the objective function

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

$$s.t.$$

$$y_i = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i$$

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

$$\boldsymbol{\beta} \ge \mathbf{0}$$

$$\varepsilon_i \le 0 \;\; \forall i = 1, \ldots, n$$

Robustness to heteroskedasticity

Same 8 scenarios with multiplicative inefficiency terms

- Earlier: $y = \mathrm{f}(\mathbf{x}) + \varepsilon$

- Now: $y = \mathrm{f}(\mathbf{x})/(1+\varepsilon)$

Robustness to heteroskedasticity

Scenario	Number of Obs.	Mean Squared Error PP	COLS	DEA	C2NLS	Fraction of Trials DEA > C2NLS	Bias PP	COLS	DEA	C2NLS	Fraction of Trials abs(DEA) > abs(C2NLS)
A)	50	0.003	0.014	0.009	0.007	0.67	0.007	0.060	-0.069	0.024	0.82
A)	100	0.003	0.016	0.004	0.005	0.44	0.020	0.079	-0.044	0.035	0.59
A)	150	0.003	0.015	0.003	0.004	0.46	0.024	0.081	-0.035	0.034	0.55
B)	50	0.002	0.023	0.011	0.008	0.75	-0.006	0.080	-0.075	0.026	0.83
B)	100	0.001	0.021	0.005	0.006	0.41	0.005	0.095	-0.048	0.041	0.59
B)	150	0.001	0.022	0.004	0.005	0.41	0.010	0.100	-0.039	0.040	0.53
C)	50	0.039	0.125	0.031	0.016	0.93	0.116	0.279	-0.126	0.029	0.94
C)	100	0.056	0.220	0.019	0.010	0.87	0.179	0.399	-0.093	0.036	0.83
C)	150	0.073	0.259	0.013	0.010	0.71	0.220	0.443	-0.074	0.058	0.62
D)	50	0.027	0.084	0.065	0.025	0.98	0.075	0.222	-0.201	-0.019	0.98
D)	100	0.039	0.144	0.046	0.019	0.98	0.132	0.323	-0.161	0.038	0.98
D)	150	0.050	0.171	0.034	0.015	0.96	0.168	0.361	-0.138	0.045	0.95
E)	50	0.039	0.075	0.204	0.076	0.95	-0.098	0.096	-0.346	0.009	0.95
E)	100	0.020	0.063	0.123	0.062	0.83	-0.029	0.134	-0.253	0.101	0.80
E)	150	0.015	0.048	0.090	0.076	0.69	0.008	0.138	-0.204	0.159	0.64
F)	50	0.081	0.128	0.529	0.150	1.00	-0.152	0.129	-0.590	-0.097	1.00
F)	100	0.036	0.097	0.338	0.116	0.97	-0.054	0.180	-0.452	0.087	0.96
F)	150	0.027	0.097	0.249	0.124	0.88	-0.004	0.206	-0.374	0.171	0.85

Head-to-head comparison of the MM and PSL estimators

	Scenario 1		Scenario 2		Scenario 3	
	σ^2	λ	σ^2	λ	σ^2	λ
	1,88	1,66	1,63	1,24	1,35	0,83
PSL better	47,0 %	49,6 %	43,0 %	40,2 %	33,8 %	34,0 %
MM better	45,8 %	43,2 %	39,8 %	41,4 %	27,6 %	27,4 %
equally good	7,2 %	7,2 %	17,2 %	18,4 %	38,6 %	38,6 %

Distributional assumptions

Schmidt, P. (1985). "Frontier Production Functions." *Econometric Reviews* **4**(2): 289-328.

In my opinion the only serious intrinsic problem with stochastic frontiers is that the separation of noise and inefficiency ultimately hinges on strong (and arbitrary) distributional assumptions. This is not easy to defend.

Distributional assumptions

Schmidt, P. (1985). "Frontier Production Functions." *Econometric Reviews* **4**(2): 289-328.

However, in defense of stochastic frontier models, it is clear that this problem is not avoided by assuming the frontier to be deterministic. Assuming statistical noise not to exist is itself a strong distributional assumption, and one that is empirically false in data sets that I have analyzed.