Maa- ja elintarviketalouden tutkimuskeskus | Agrifood Research Finland
Forskningscentralen för jordbruk och livsmedelsekonomi
MTT

HSE HELSINGIN KAUPPAKORKEAKOULU
HELSINKI SCHOOL OF ECONOMICS

DEA as Nonparametric Least Squares Regression

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

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)	

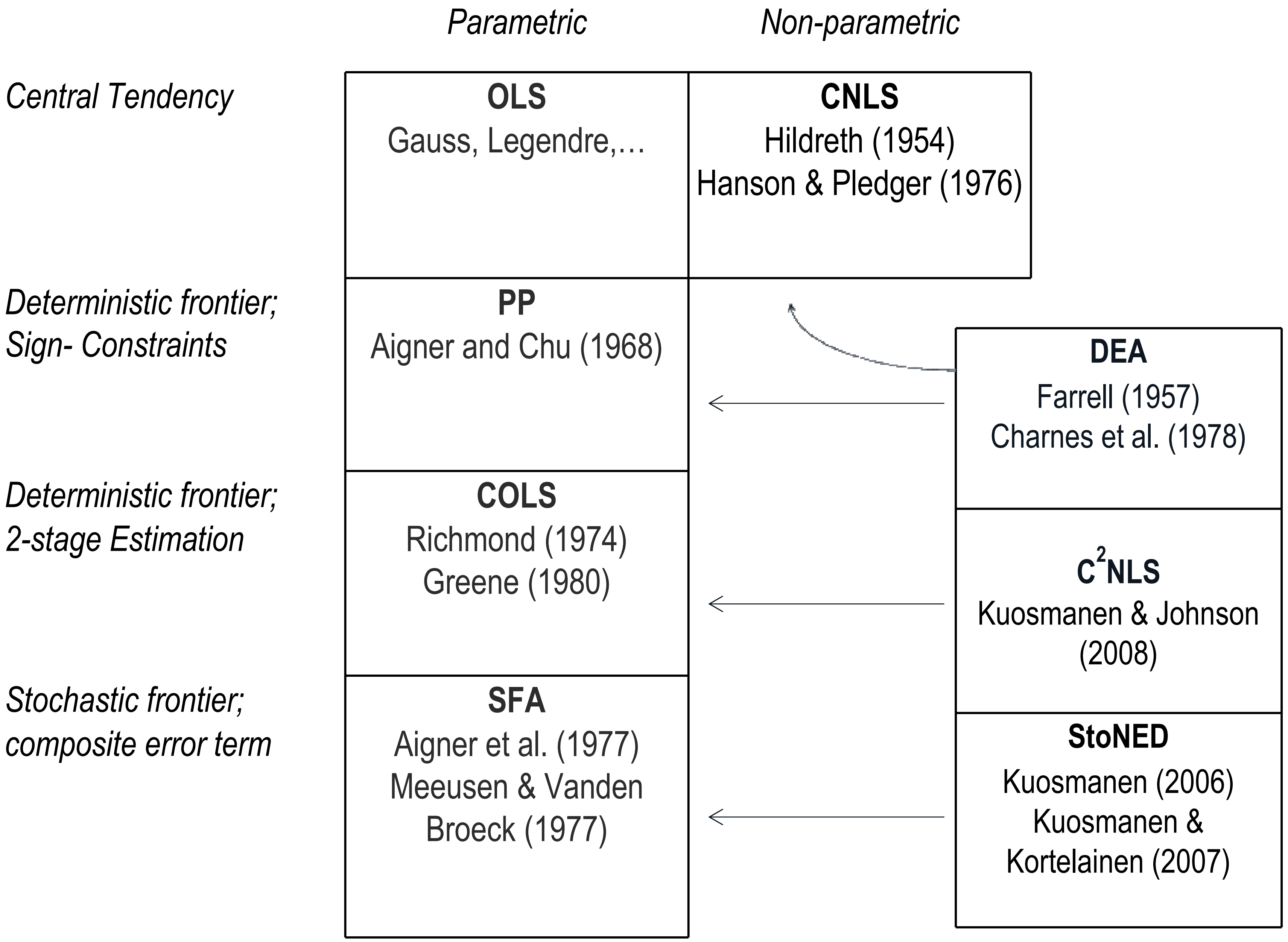
Classification

	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^{2}NLS** Kuosmanen & Johnson (2008)
Stochastic frontier; composite error term	**SFA** Aigner et al. (1977) Meeusen & Vanden Broeck (1977)	**StoNED** Kuosmanen (2006) Kuosmanen & 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 (1954)

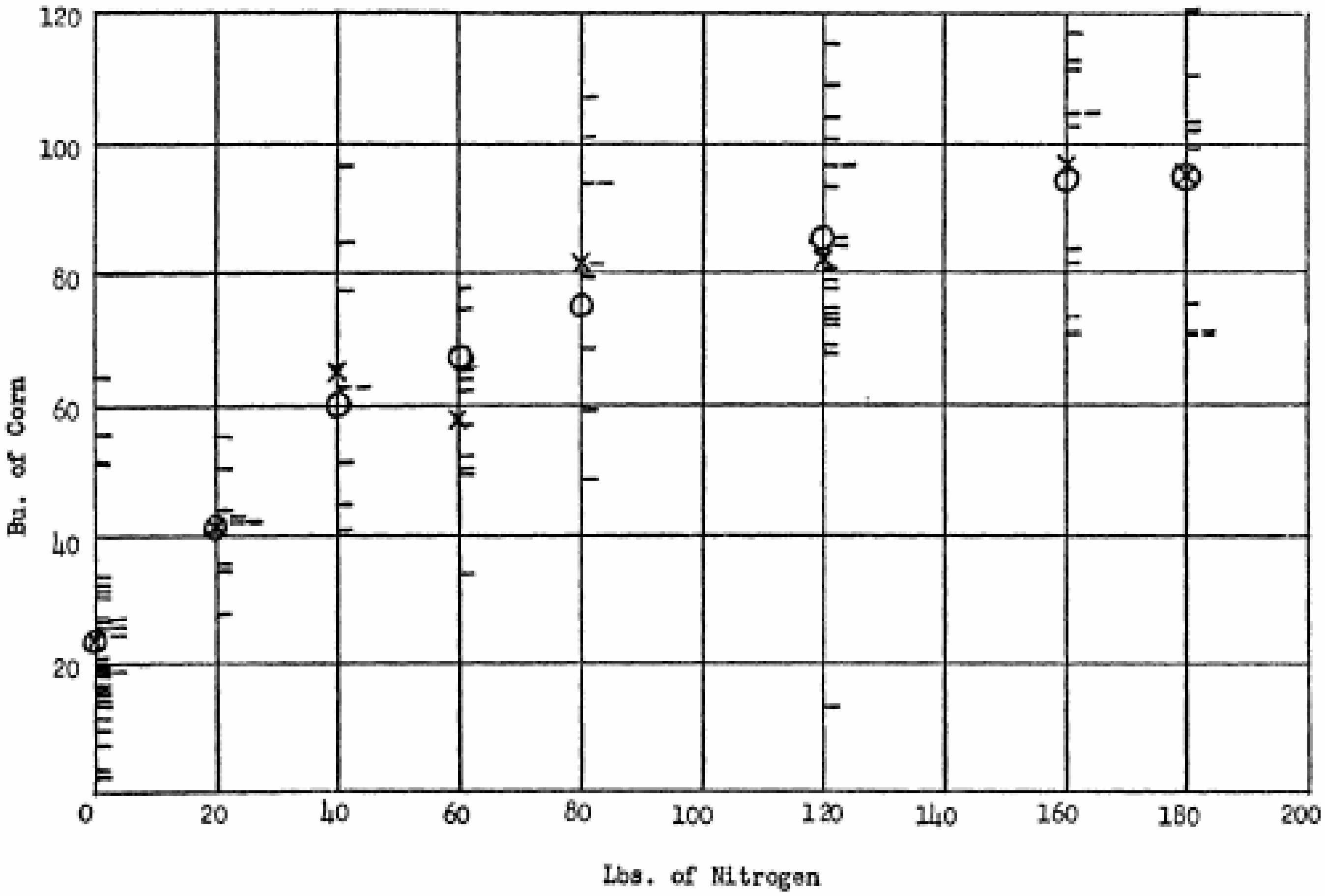


FIG. 1. Observations and Estimates.

— Observed yield (y_{nt})

× Mean yield ($\bar{y}_n$)

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

Modified illustration

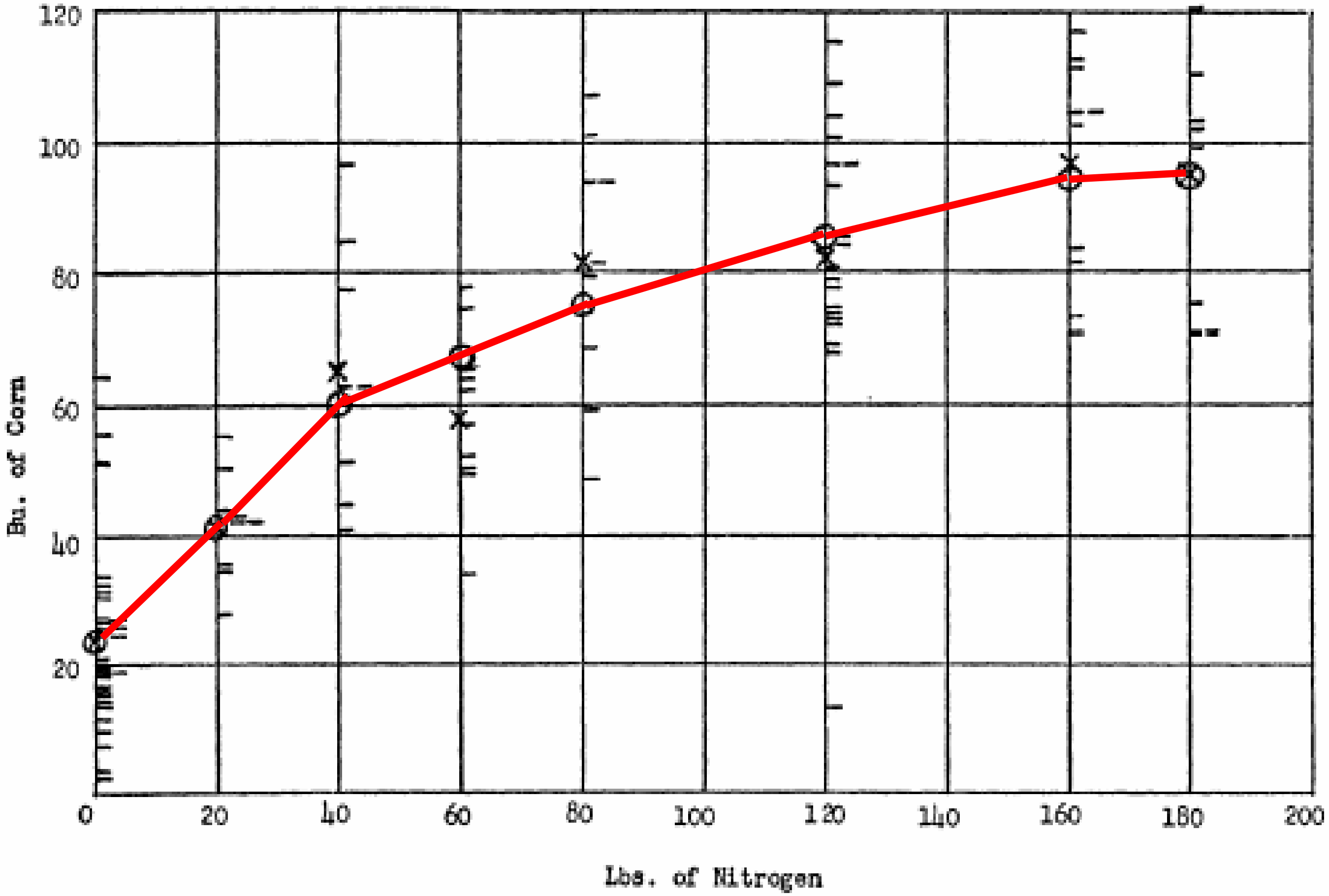


FIG. 1. Observations and Estimates.

– Observed yield (y_{nt})

× Mean yield ($\bar{y}_n$)

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

CNLS model

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

Assumptions:

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

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

CNLS problem

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

$$s.t.$$

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

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)

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

$$s.t.$$

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

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

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

CNLS problem with multiple **x**

(Kuosmanen 2008)

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

$$s.t.$$

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

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

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

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

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

$$s.t.$$

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

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

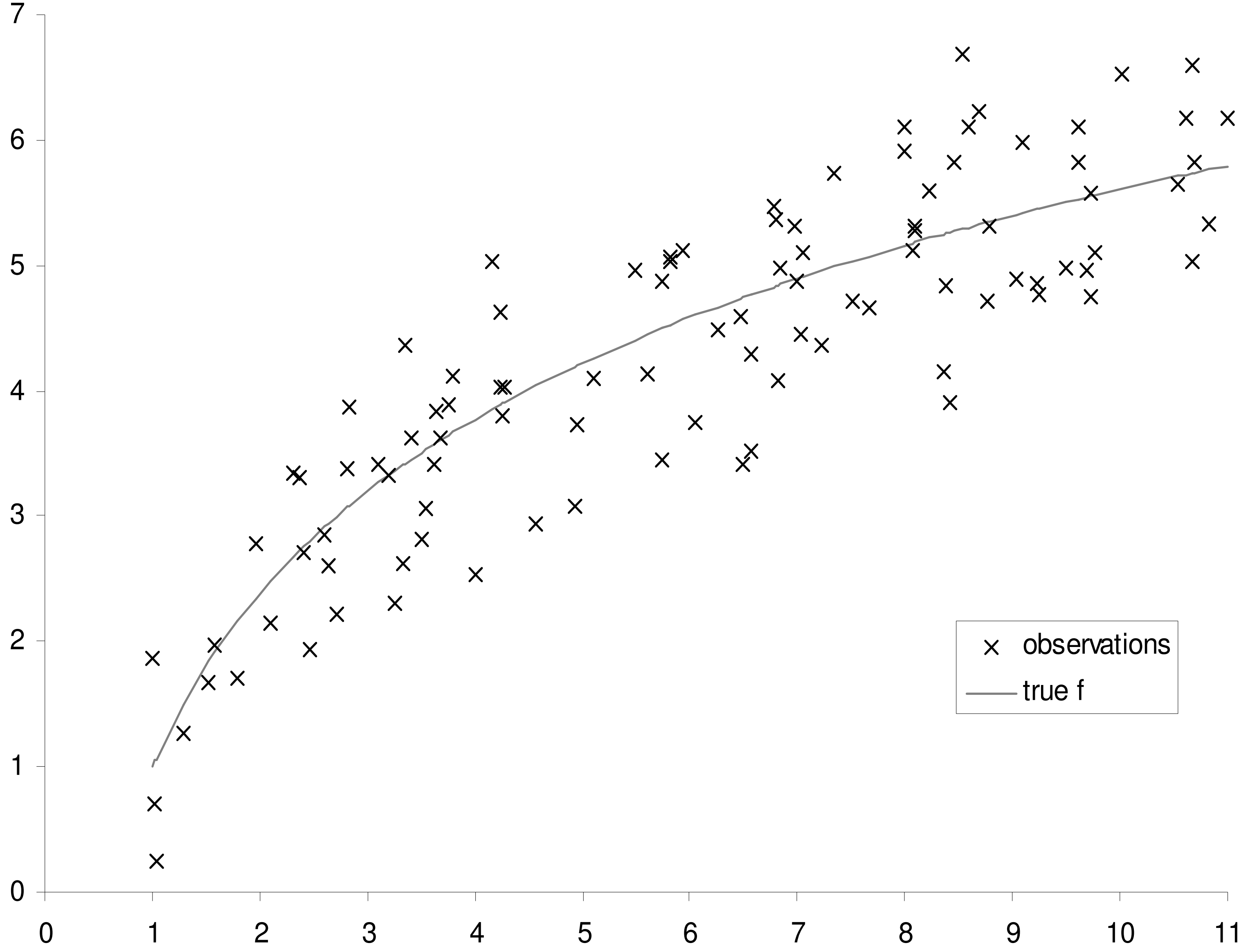
$$\alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_i \leq \alpha_h + \boldsymbol{\beta}_h'\mathbf{x}_i \ \forall h,i = 1,\ldots,n \ \text{(concavity)}$$

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

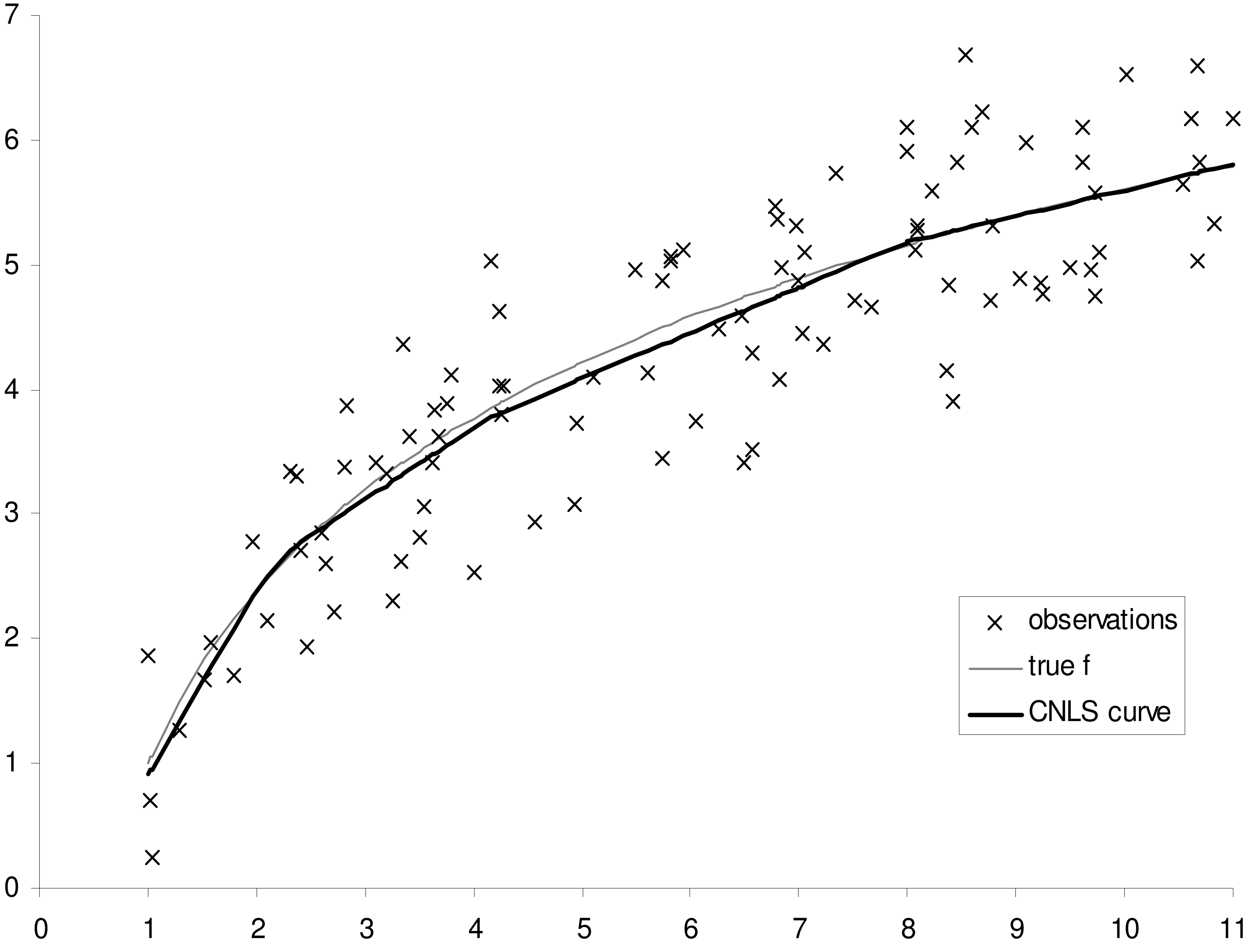
Simulation example

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

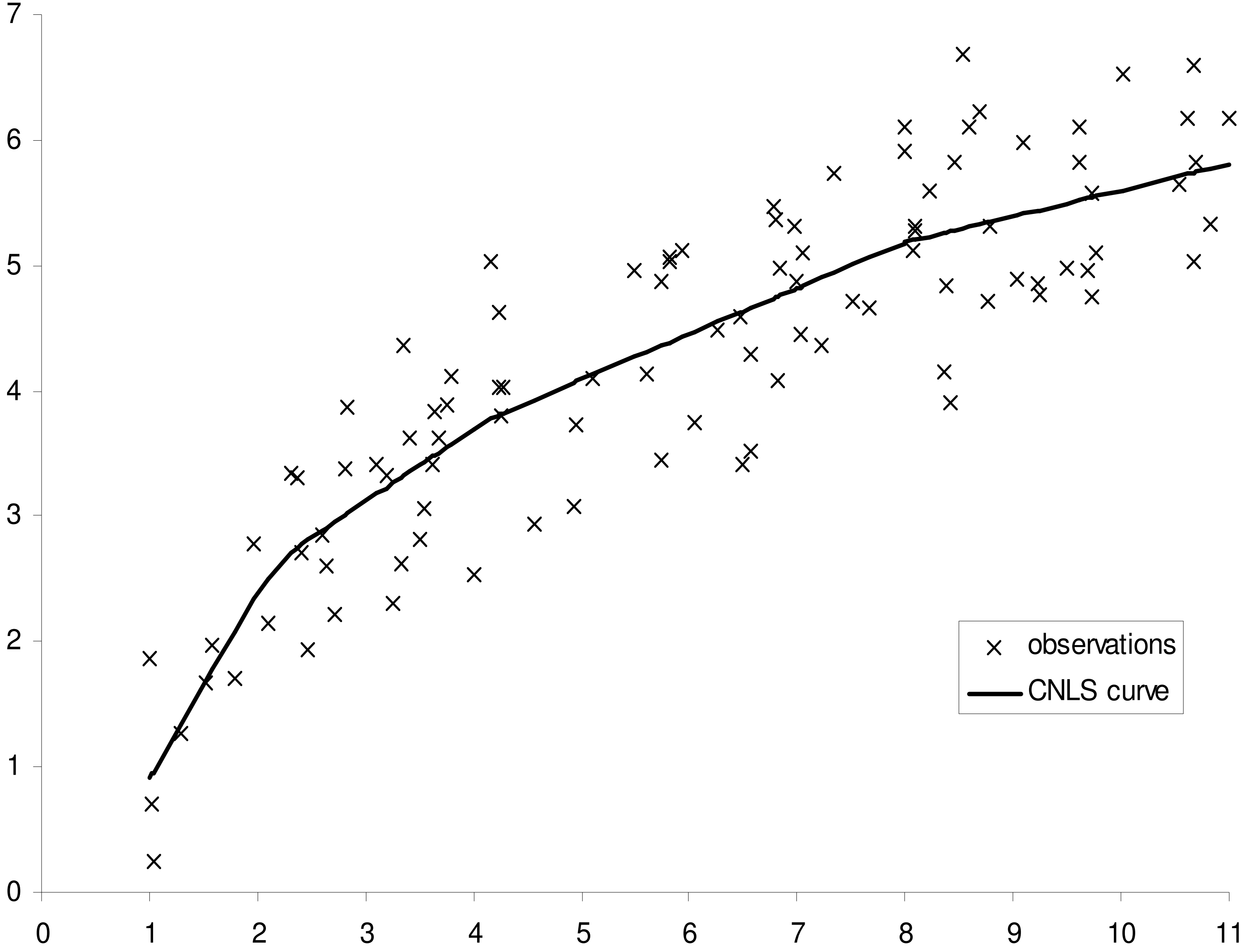
Data and true *f*



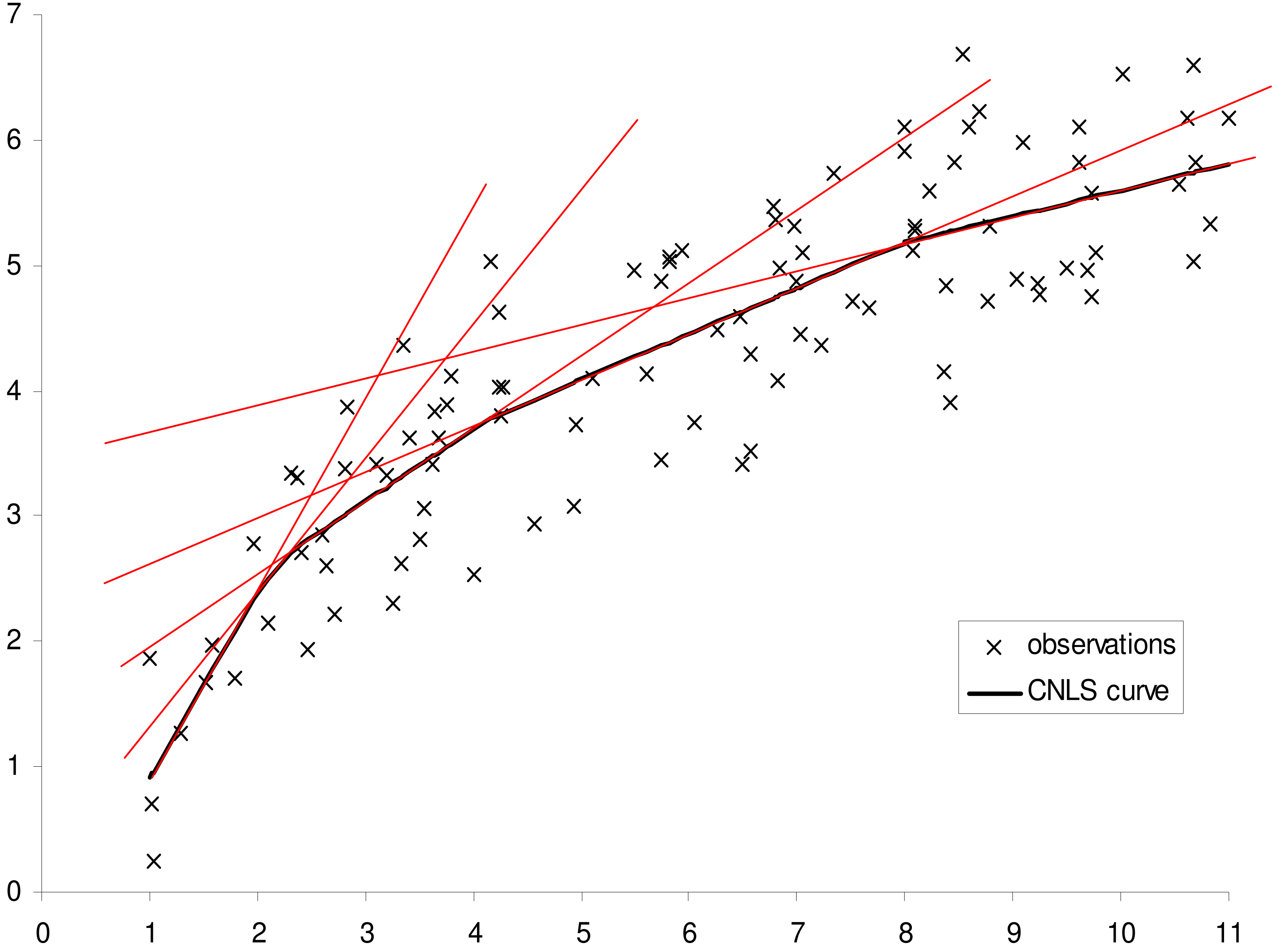
CNLS and true *f*



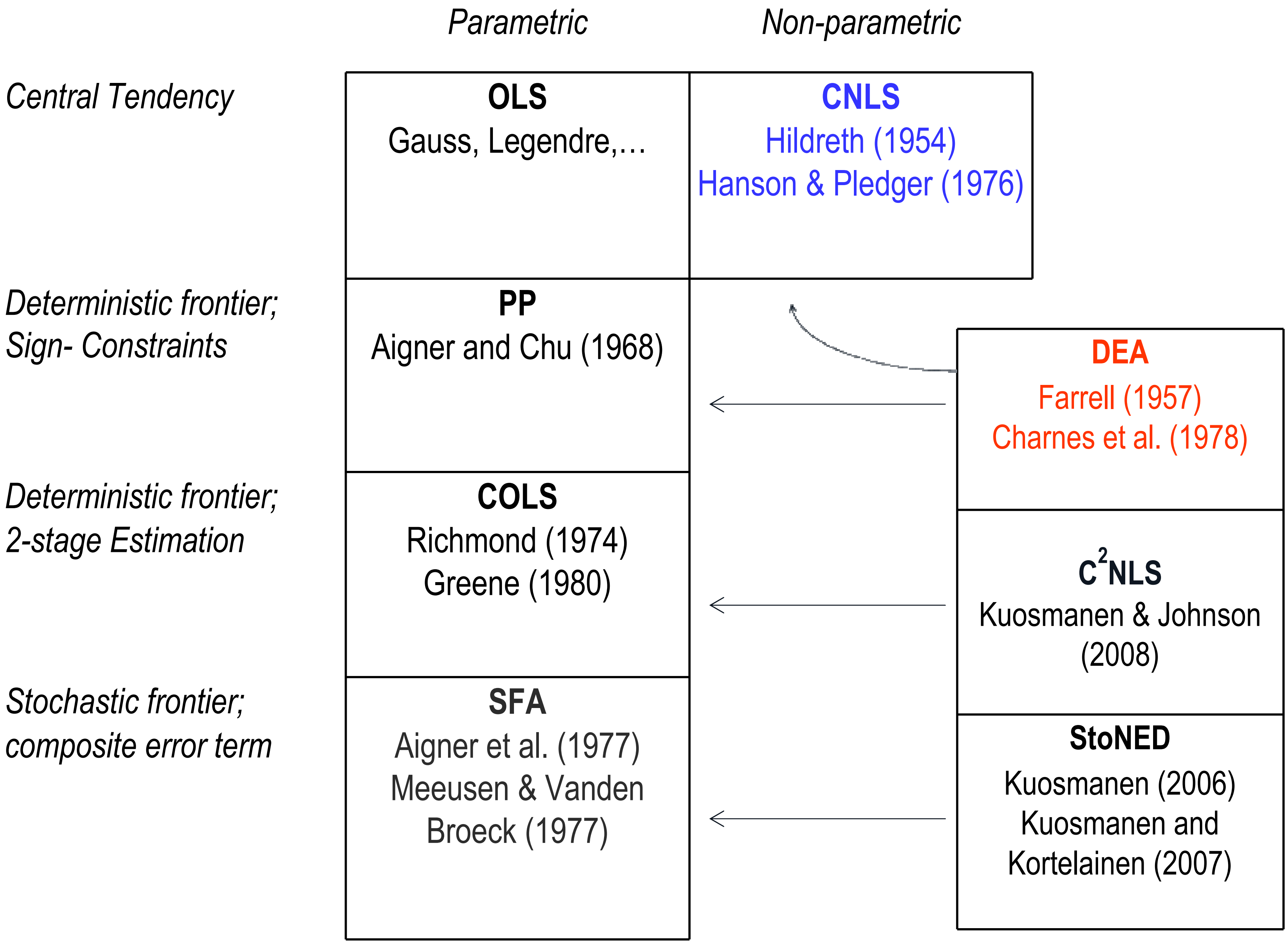
CNLS regression



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

- Parametric Programming model (Aigner&Chu 1968) 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

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

$$s.t.$$

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

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

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

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

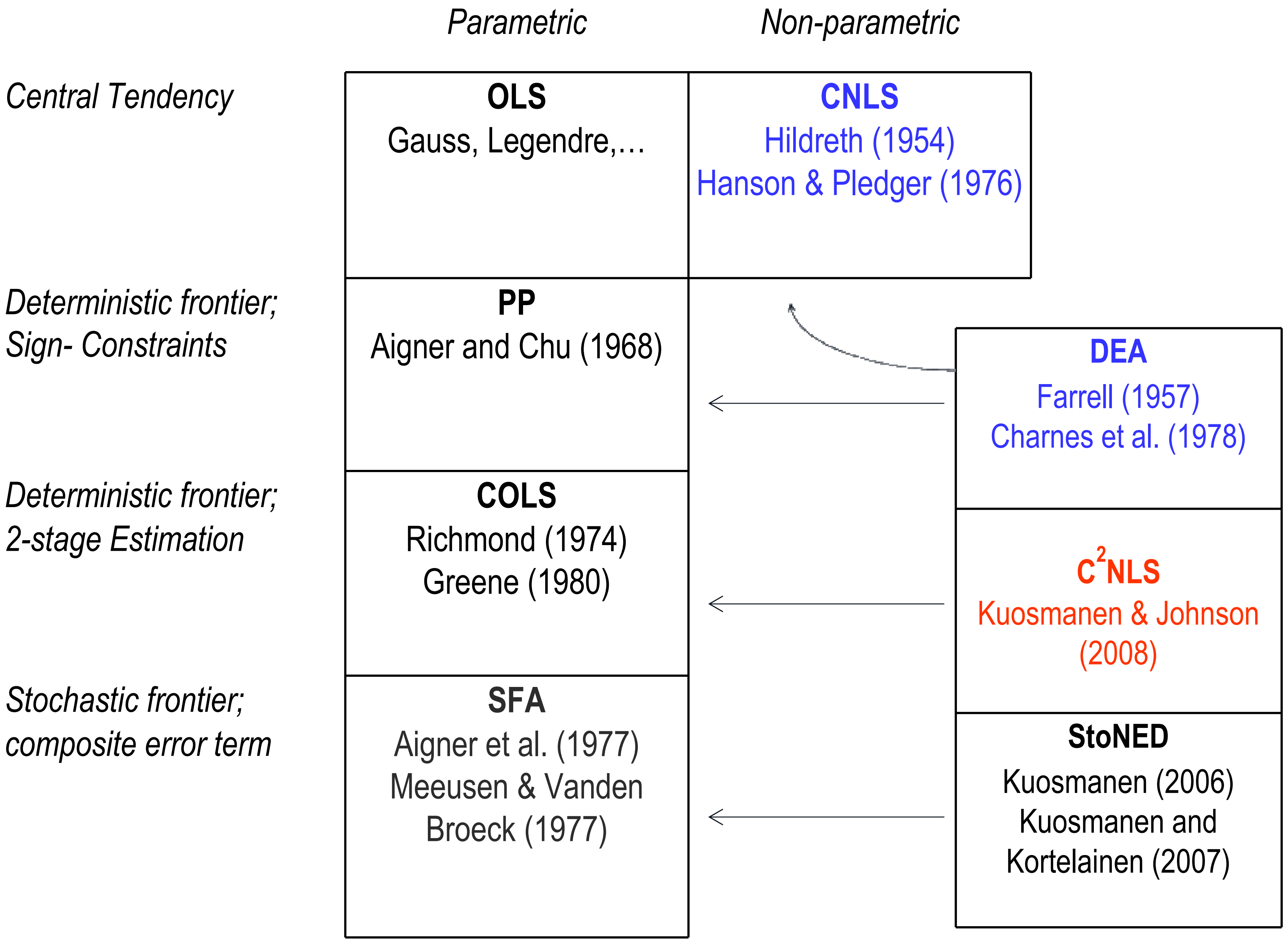
”Sign-constrained CNLS” = DEA

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

Both measure efficiency relative to the same DEA frontier.

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

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 (C^{2}NLS)

- 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)

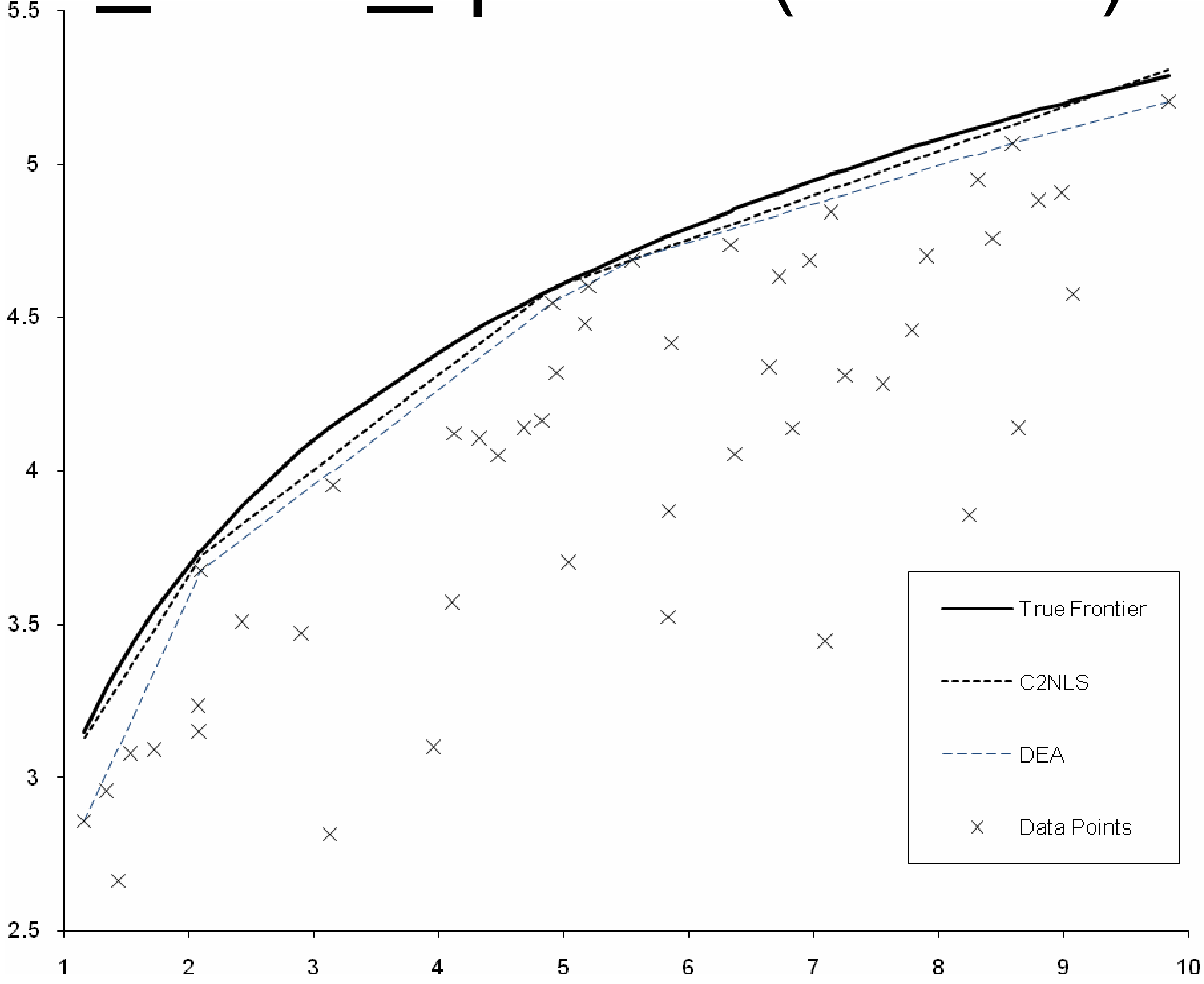
Theorem 4.1: *For any sequence of independent observations* $\mathbf{X},\mathbf{y}$ *generated by production function f and i.i.d inefficiency terms* $\boldsymbol{\varepsilon}$ *that are uncorrelated with* $\mathbf{X}$ *and have a positive density at* $\boldsymbol{\varepsilon}=\mathbf{0}$*, the* C^2*NLS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

Corrected Convex Nonparametric Least Squares (C^2NLS)

***Theorem 4.2**: For any real valued data set, the discriminatory power of C^2NLS is always greater than or equal to that of DEA in the sense that*

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

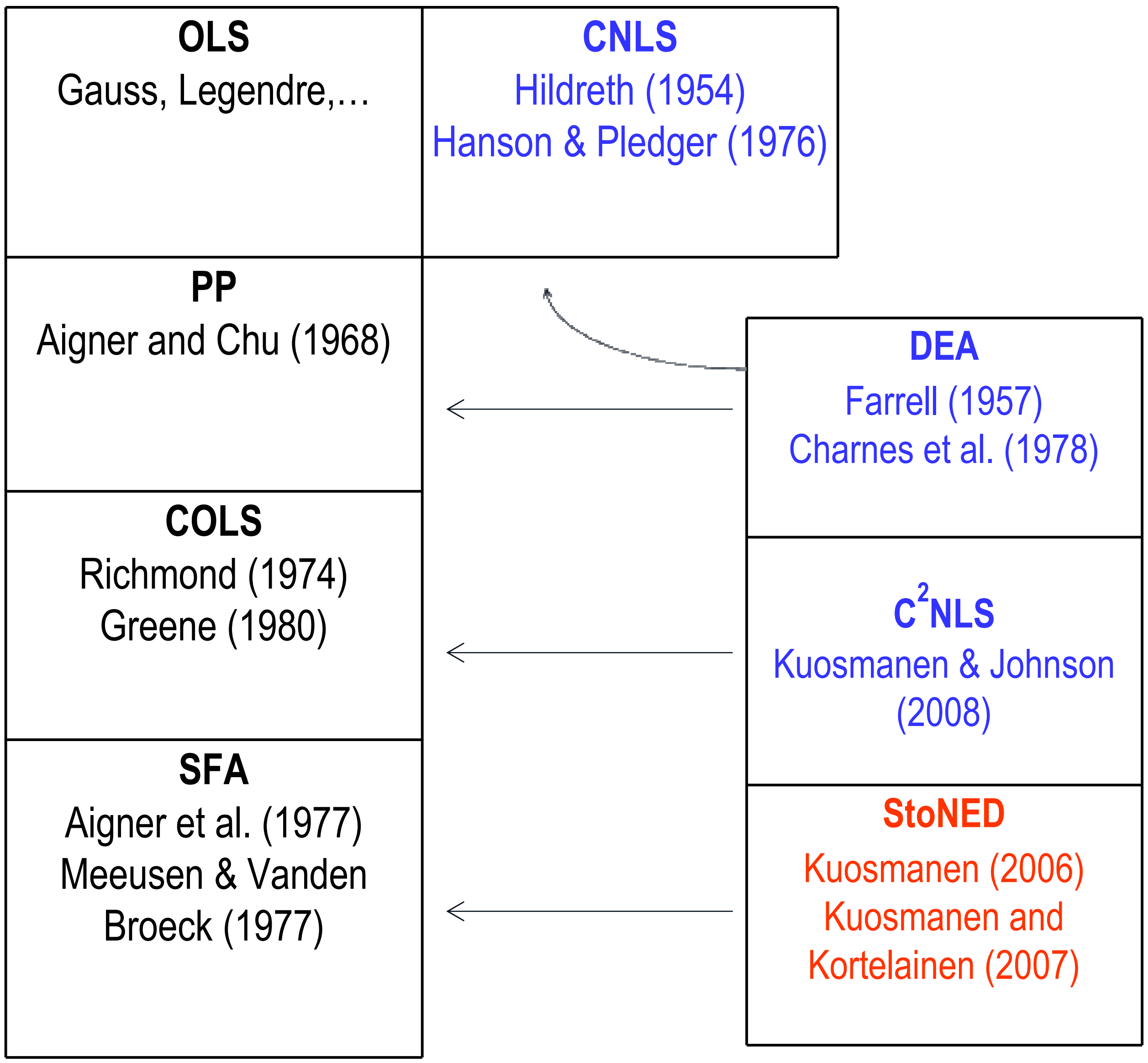
Corrected Convex Nonparametric Least Squares (C^{2}NLS)



Pros and cons of C^{2}NLS

- C^{2}NLS estimates the shape of the frontier making use of all observed data points
  - more efficient than DEA
- C^{2}NLS has a higher discriminatory power than DEA
- C^{2}NLS efficiency estimates are more sensitive to outliers and noise than DEA
- Relative efficiency rankings of C^{2}NLS are more robust than those of DEA

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/ ε_i) (Jondrow et al. 1982)

ML estimator

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

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

$$s.t.$$

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

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

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

Least squares estimation of StoNED model

In two steps

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

Step 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(λ):

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

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

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

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

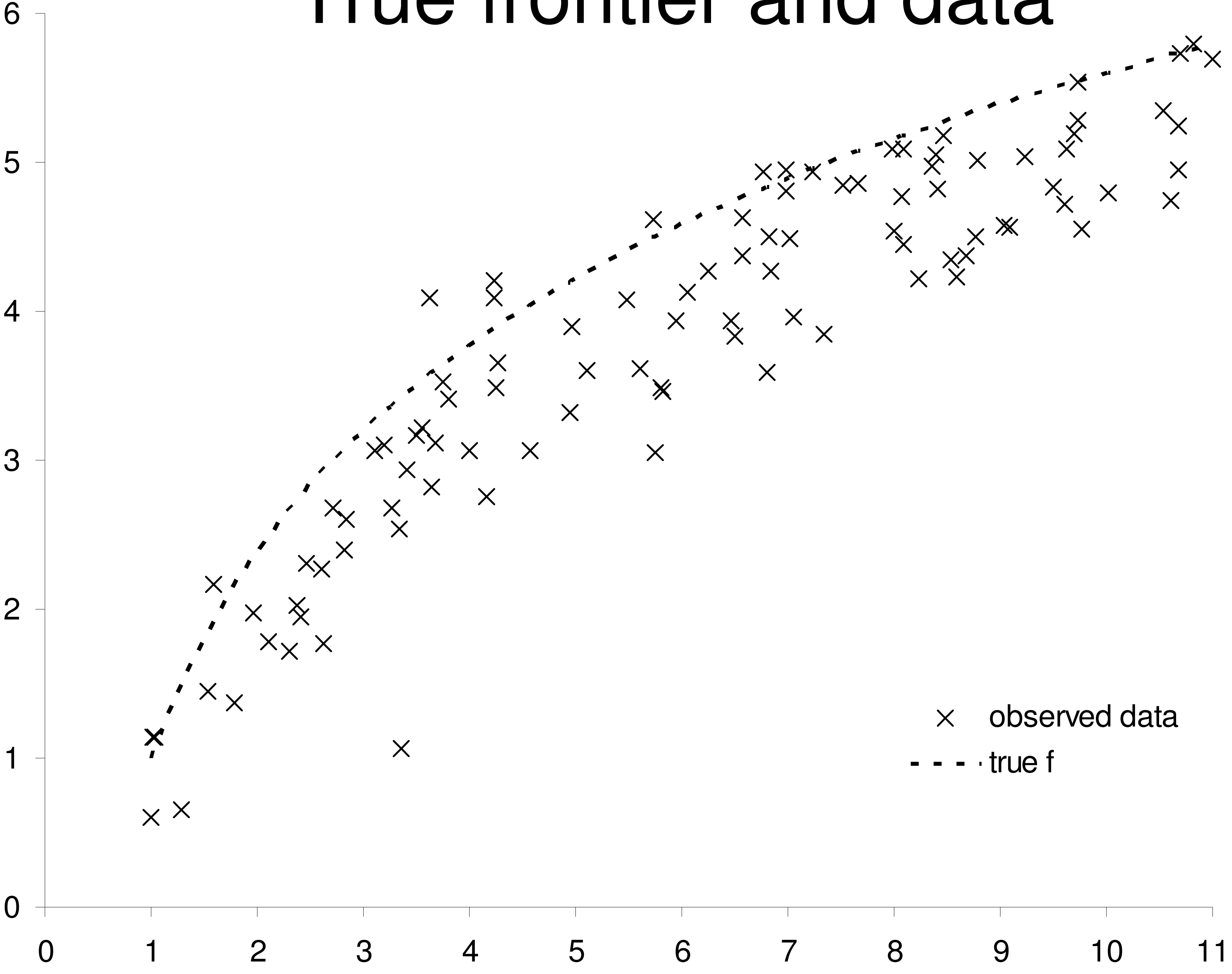
- Level of inefficiency depends on the distributional assumptions

- In panel data setting, distributional assumptions can be relaxed

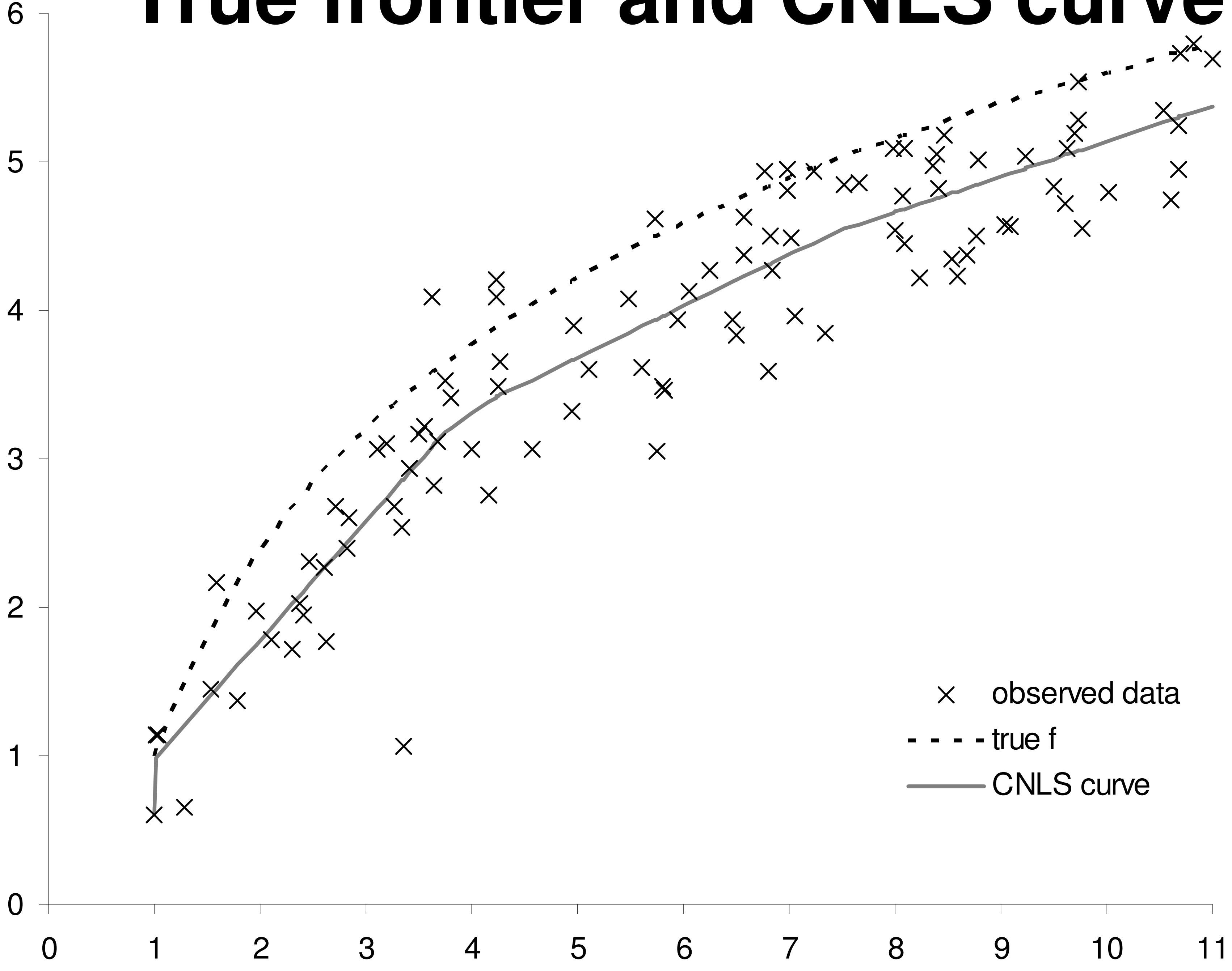
Simulation example

- 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.

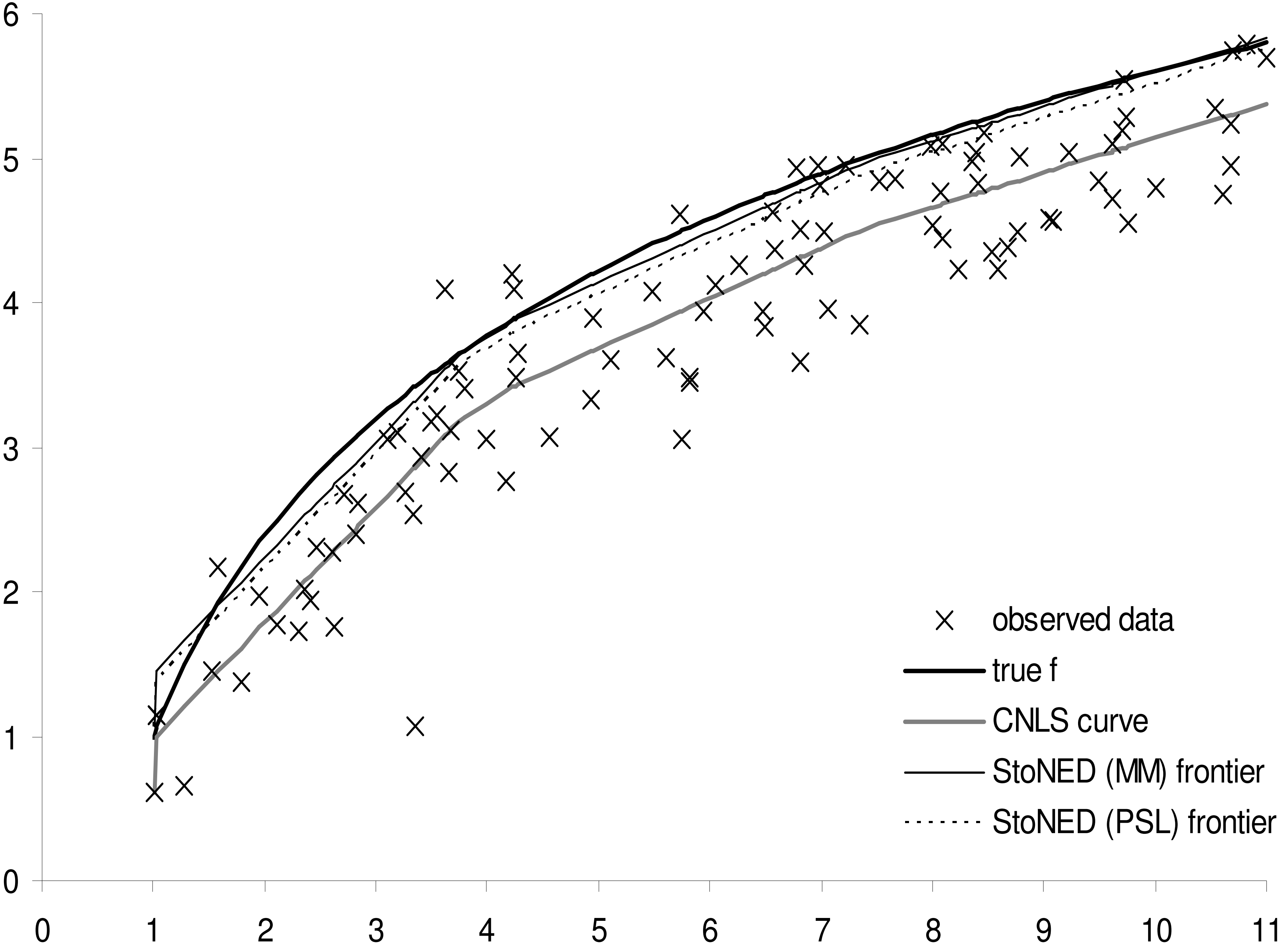
True frontier and data



True frontier and CNLS curve



Estimated frontiers



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 can be seen as 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^{2}NLS 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 $\mathbf{z}$ that influence efficiency (cf. critique by Simar & Wilson)
    - 1-stage DEA regression incl. both $\mathbf{x}$ and $\mathbf{z}$
    - C^{2}NLS and StoNED models with $\mathbf{z}$ 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

$$
\begin{aligned}
&\min_{\boldsymbol{\alpha},\boldsymbol{\beta}}(-\varepsilon)\\
&s.t.\\
&y_i = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon\\
&y_h \le \alpha + \boldsymbol{\beta}'\mathbf{x}_h \;\; \forall h = 1,\dots,n\\
&\boldsymbol{\beta} \ge \mathbf{0}
\end{aligned}
$$

Proof of Theorem 3.1

Step 2: Sum over all observations *i*

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

Proof of Theorem 3.1

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

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

$$s.t.$$

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

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

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

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

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

$$
\begin{aligned}
&\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \sum_{i=1}^{n} \varepsilon_i^2 \\
&s.t. \\
&y_i = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i \\
&\alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_i \le \alpha_h + \boldsymbol{\beta}_h'\mathbf{x}_i \ \ \forall h, i = 1,\ldots,n \\
&\boldsymbol{\beta} \ge \mathbf{0} \\
&\varepsilon_i \le 0 \ \forall i = 1,\ldots,n
\end{aligned}
$$