Productivity and Efficiency Analysis

4) Unified approach: StoNED

c) Convex regression

Timo Kuosmanen

Aalto University School of Business

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

Taxonomy of methods

based on Kuosmanen & Johnson (2010), Operations Research

		Parametric	Nonparametric: Local averaging	Nonparametric: Axiomatic
Average curve		*OLS* Gauss (1795), Legendre (1805)	*Kernel regression* Nadaraya (1964), Watson (1964)	*Convex regression* Hildreth (1954), Hanson and Pledger (1976)
Frontier	**Deterministic (Sign constr.)**	*Parametric programming* Aigner and Chu (1968)	*Nonparametric programming* Post et al. (2002)	***DEA*** **Farrell (1957),** **Charnes et al. (1978)**
	Deterministic (2-stage)	*Corrected OLS* Winsten (1957) Greene (1980)	*Corrected kernel* Kneip and Simar (1996)	*Corrected CNLS* Kuosmanen and Johnson (2010)
	Stochastic	***SFA*** **Aigner et al. (1977)** **Meeusen and van den Broeck (1977)**	*Semi-nonparametric SFA* Fan, Li and Weersink (1996)	***StoNED*** **Kuosmanen and Kortelainen (2012)**

Convex regression

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

where

y_i is dependent variable

f is monotonic increasing and concave regression function

$\mathbf{x}_i$ is vector of explanatory variables

ε_i is symmetric error term (zero mean, constant variance)

Source: Kuosmanen, Johnson & Saastamoinen (2014) Stochastic nonparametric approach to efficiency analysis: A Unified Framework, in J. Zhu (Ed) *Handbook on DEA Vol. 2*, Springer.

CNLS problem: conceptual definition

$$\min_{f} \sum_{i=1}^{n} (y_i - f(\mathbf{x}_i))^2$$

subject to

$$f \in F_2$$

Where F_2 is the set of monotonic increasing and concave functions.

If the true function to be estimated is monotonic increasing and concave, then the CNLS estimator is statistically consistent.

CNLS problem: QP formulation

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

subject to

$$y_i = \alpha_i + \boldsymbol{\beta}'_i \mathbf{x}_i + \varepsilon_i^{CNLS} \; \forall i$$

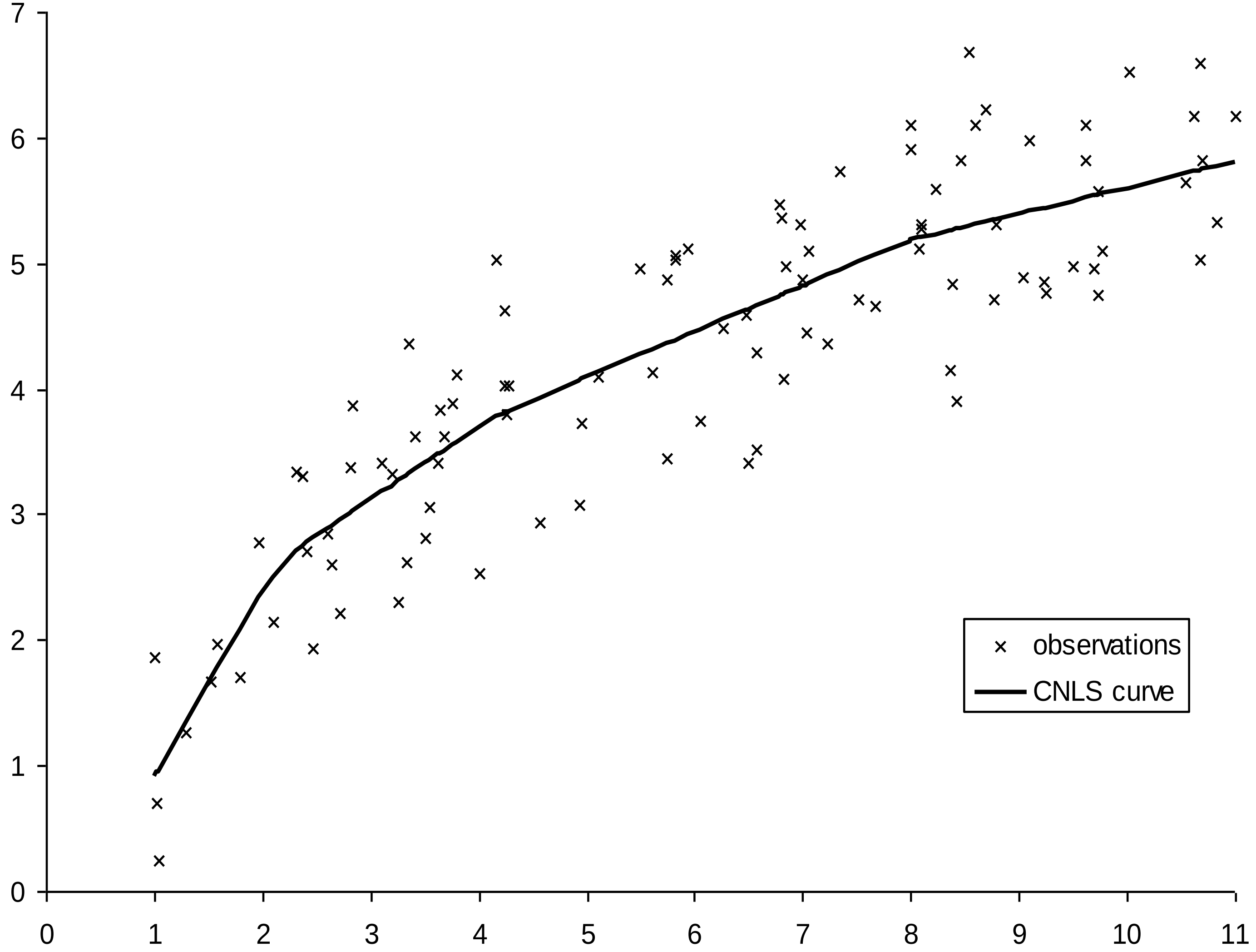
$$\alpha_i + \boldsymbol{\beta}'_i \mathbf{x}_i \leq \alpha_h + \boldsymbol{\beta}'_h \mathbf{x}_i \; \forall h, i$$

$$\boldsymbol{\beta}_i \geq \mathbf{0} \; \forall i$$

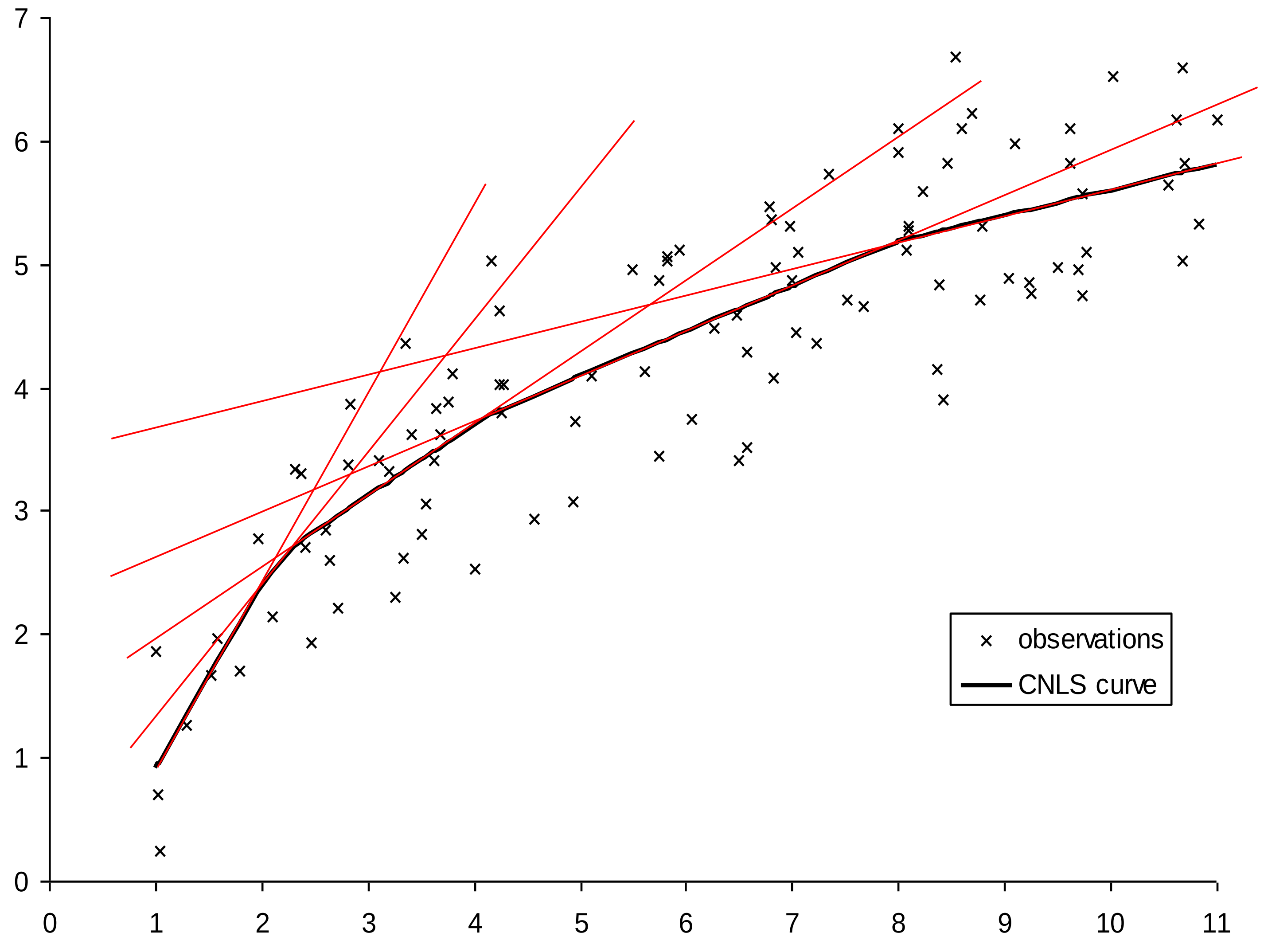
Representation theorem (Kuosmanen, 2008): The QP problem is equivalent to the infinite dimensional CNLS problem: the optimal solutions satisfy

$$f^*(\mathbf{x}_i) = \alpha_i^* + \boldsymbol{\beta}_i^* \mathbf{x}_i$$

CNLS : Illustration



CNLS : Illustration



Interpolation to unobserved points x

- Our object of interest is the regression function *f*.

- Thus far we have only fitted values of *f* in the observed data points $\mathbf{x}_i$, $i = 1,\ldots,n$.

- To interpolate to unobserved points **x**, we can use the lower bound (output-oriented multiplier formulation of DEA)

$$\hat{f}_{\min}^{CNLS}(\mathbf{x}) = \min_{\alpha,\boldsymbol{\beta}} \left\{ \alpha + \boldsymbol{\beta}' x \,\middle|\, \alpha + \boldsymbol{\beta}' \mathbf{x}_i \geq \hat{f}^{CNLS}(\mathbf{x}_i)\ \forall i = 1,\ldots,n \right\} \tag{7.5}$$

Theorem 3 Function $\hat{f}_{\min}^{CNLS}$ stated in Eq. (7.5) is one of the optimal solutions to the infinite dimensional optimization problem (7.2). It is the unique lower bound for the functions that solve problem (7.2), formally

$$\hat{f}_{\min}^{CNLS}(\mathbf{x}) \leq f^*(\mathbf{x}) \text{ for all } \mathbf{x} \in \Re_+^m \text{ and } f^* \in F_2^*.$$

DEA as sign-constrained CNLS

Suppose we impose to the CNLS problem an additional constraint to restrict the sign of the CNLS residuals:

$$\varepsilon_i^{CNLS-} \leq 0 \; \forall i$$

Theorem 4 The sign-constrained CNLS estimator is equivalent to the DEA VRS estimator:

$$\hat{f}_{\min}^{CNLS-}(\mathbf{x}) = \hat{f}^{DEA}(\mathbf{x})$$

Source: Kuosmanen & Johnson (2010) *Operations Research*

Multiplicative NLP formulation

Multiplicative composite error structure is obtained by rephrasing model (7.1) as

$$y_i = f(\mathbf{x}_i) \cdot \exp(\varepsilon_i) = f(\mathbf{x}_i) \cdot \exp(v_i - u_i) \tag{7.24}$$

Applying the log-transformation to Eq. (7.23), we obtain

$$\ln y_i = \ln f(\mathbf{x}_i) + \varepsilon_i \,. \tag{7.25}$$

In the multiplicative case, the CNLS formulation (7.3) can be rephrased as

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

subject to

$$\begin{aligned}
&\ln y_i = \ln(\phi_i + 1) + \varepsilon_i^{CNLS} \; \forall i \\
&\phi_i + 1 = \alpha_i + \boldsymbol{\beta}'_i \mathbf{x}_i \; \forall i \\
&\alpha_i + \boldsymbol{\beta}'_i \mathbf{x}_i \le \alpha_h + \boldsymbol{\beta}'_h \mathbf{x}_i \; \forall h, i \\
&\boldsymbol{\beta}_i \ge \mathbf{0} \; \forall i
\end{aligned} \tag{7.26}$$

Returns to scale in CNLS

CNLS formulations above are stated in the variable returns to scale (VRS) case.

Analogous to DEA, alternative returns to scale specifications are obtained by imposing additional constraints:

Constant returns to scale (CRS): impose $\alpha_i = 0 \ \forall i$
Non-increasing returns to scale (NIRS): impose $\alpha_i \geq 0 \ \forall i$
Non-decreasing returns to scale (NDRS): impose $\alpha_i \leq 0 \ \forall i$

Note: CRS requires multiplicative formulation to ensure that $f(\mathbf{0}) = 0$.

Relaxing convexity: INLS formulation

Analogous to FDH, we can relax convexity and rely on monotonicity

Define the binary matrix $\mathbf{P} = [p_{ij}]_{n\times n}$ as follows

$$p_{ij} = \begin{cases} 1 & \text{if } \mathbf{x}_i \preccurlyeq \mathbf{x}_j \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} \min \quad & \sum_{i=1}^{n} \varepsilon_i^2 \\ \text{s.t.} \quad & y_i = \alpha_i + \boldsymbol{\beta}_i' \mathbf{x}_i + \varepsilon_i \quad i = 1, \ldots, n, \\ & p_{ij}(\alpha_i + \boldsymbol{\beta}_i' \mathbf{x}_i) \leqslant p_{ij}\left(\alpha_j + \boldsymbol{\beta}_j' \mathbf{x}_j\right), i, \quad j = 1, \ldots, n, \\ & \boldsymbol{\beta}_i \geqslant 0 \quad i = 1, \ldots, n \end{aligned}$$

Source: Keshvari & Kuosmanen (2013) *EJOR*

Next lesson

4d) Decomposing the residual