State of the art of quantitative (frontier) performance measurement techniques

SAPIENZA
UNIVERSITÀ DI ROMA

RISIS
Research infrastructure for research and innovation policy studies

Cinzia Daraio (daraio@diag.uniroma1.it)

Department of Computer, Control and Management Engineering (DIAG), Sapienza University of Rome, 22 March 2021

Outline

I. Background
II. Distances from the frontier
III. Classical models
IV. The estimation problem
V. Taxonomy of efficient frontier models
VI. Choice of a model

I. Background

Background and notation

- Objective: evaluate the performance of a given sample of units (or decision making units) from the technical efficiency point of view, i.e. their ability to operate close to the boundary of the *production set*, Ψ.
- We assume to have data in *cross-sectional form*, and for each unit we have information about inputs and outputs.
- The measurement of efficiency on these data is done by defining a frontier of the production set and then measuring the distance of any point from this frontier.
- The production set is defined as:

$$\Psi = \{(x, y) \mid x \in \mathbb{R}_+^p,\ y \in \mathbb{R}_+^q, (x, y) \text{ is feasible}\}$$

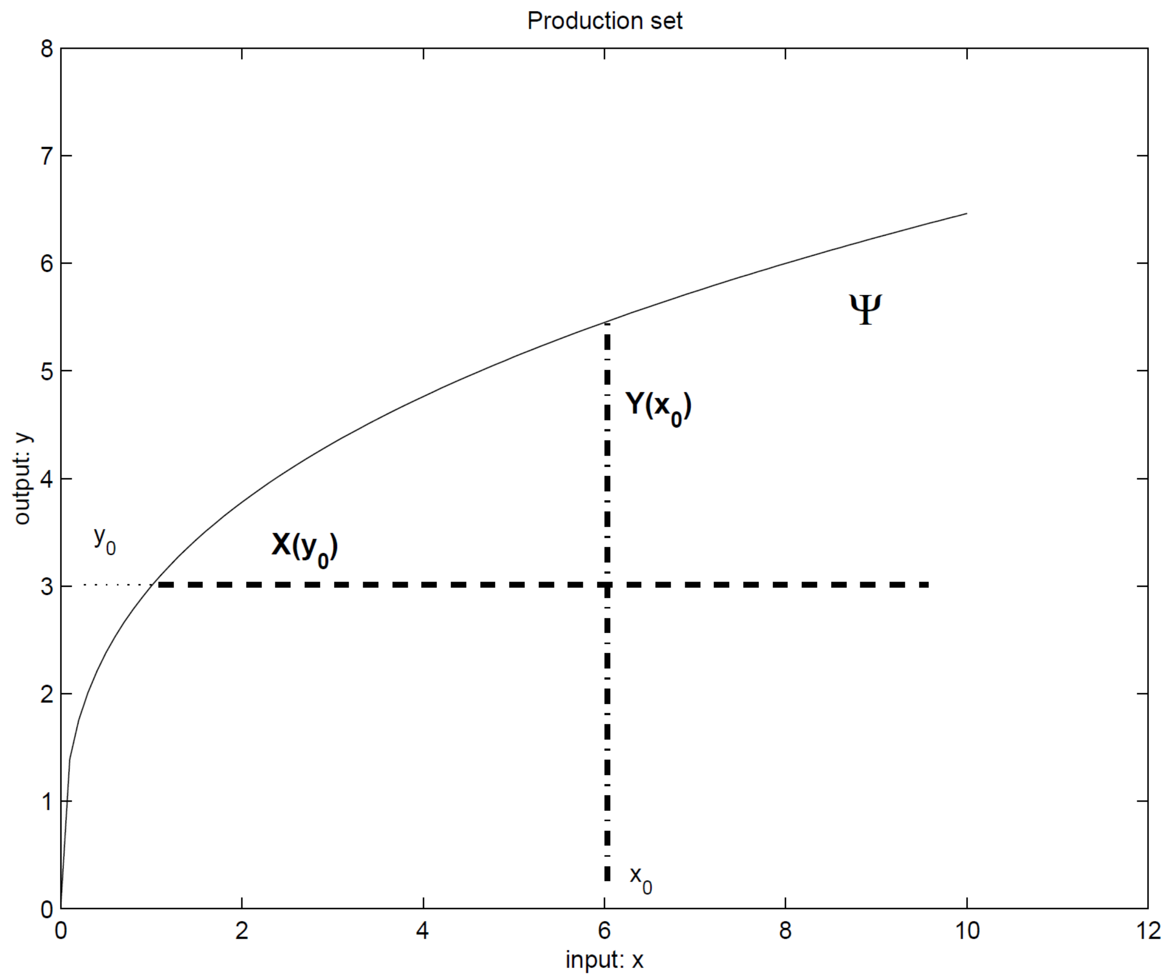
Background and basic concepts

- We can define Ψ in terms of its two sections:
  - *Input requirement* set: $$C(y) = \{x \in \mathbb{R}^p_+ | (x, y) \in \Psi\}$$
  - *Output correspondence* set: $$P(x) = \{y \in \mathbb{R}^q_+ | (x, y) \in \Psi\}$$
- The isoquants of the two sections can be defined in radial terms (Farrell, 1957) as follows:
  - Input space (it is not possible to contract anymore *x* without going outside the production set) $$\partial C(y) = \{x | x \in C(y), \theta x \notin C(y), \forall \theta, 0 < \theta < 1)\}$$
  - Output space (it is not possible to expand anymore *y* without going outside the production set): $$\partial P(x) = \{y | y \in P(x), \lambda y \notin P(x), \forall \lambda > 1\}$$
- Definition of Shephard:

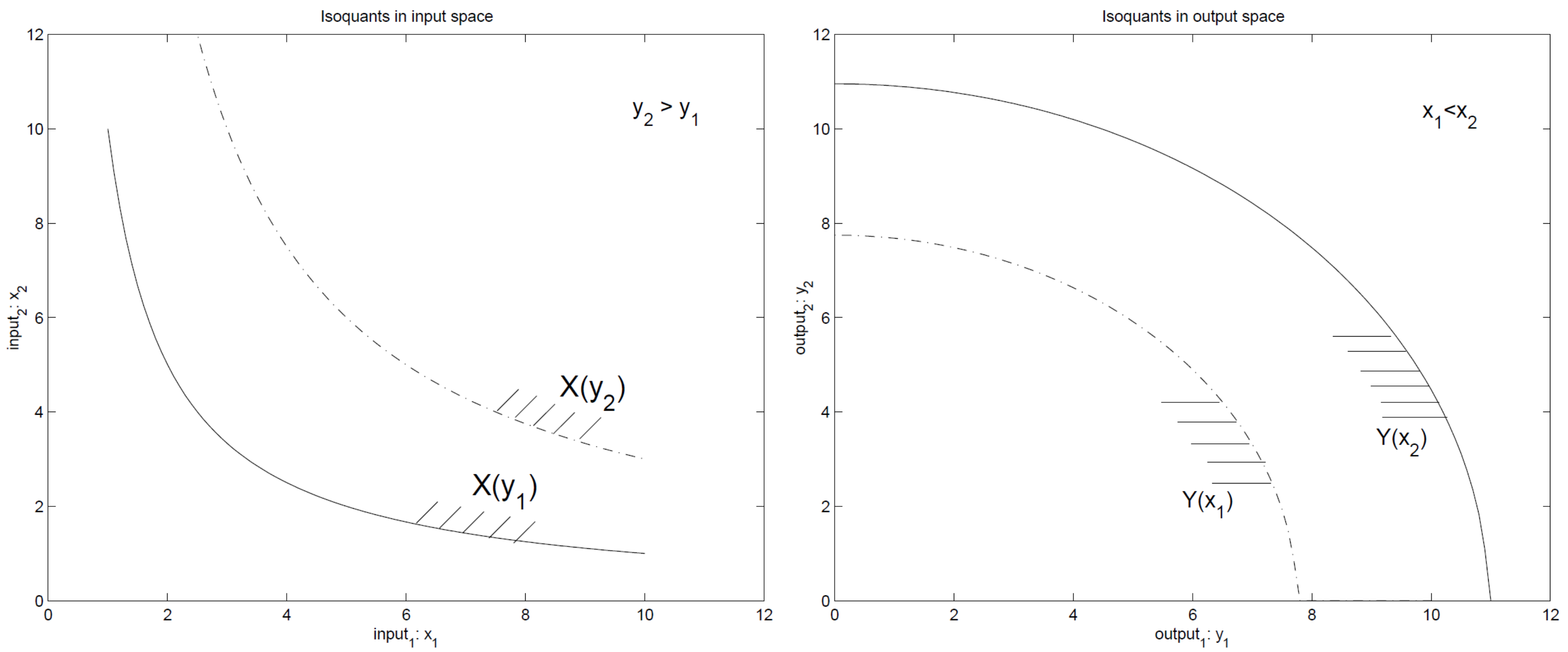
$$\partial C(y) = \{x | x \in C(y), x' \leq x, x' \neq x \Rightarrow x' \notin C(y)\}$$

$$\partial P(x) = \{y | y \in P(x), y' \geq y, y' \neq y \Rightarrow y' \notin P(x)\}.$$

An illustration of a production set



An illustration of two isoquants



Background and basic concepts

- NB. The frontiers of the input requirement set $(\partial C(y))$ and of the output correspondence set $(\partial P(x))$ are two alternatives ways to describe the frontier of the production set that is *unique*
- The frontier can be considered:
  - From the input space (*input orientation*): given the output *y* we look at the maximum contraction of the input usage
  - or from the output space (*output orientation*): given the inputs x we look at the maximum expansion of the output.

Economic Assumptions (EA) on Ψ (Shephard, 1970)

- *EA1:* **No free lunch**. $(x, y) \notin \Psi$ *if* $x = 0, y \geq 0, y \neq 0$.

  This axiom states that inactivity is always possible, *i.e.*, zero output can be produced by any input vector $x \in \mathcal{R}^p_+$, but it is impossible to produce output without any inputs.

- *EA2:* **Free disposability**. *Let* $\widetilde{x} \in \mathcal{R}^p_+$ *and* $\widetilde{y} \in \mathcal{R}^q_+$, *with* $\widetilde{x} \geq x$ *and* $\widetilde{y} \leq y$, *if* $(x, y) \in \Psi$ *then* $(\widetilde{x}, y) \in \Psi$ *and* $(x, \widetilde{y}) \in \Psi$.

  This is the *free disposability* assumption, named also the 'possibility of destroying goods without costs', on the production set Ψ.

- *EA3:* **Bounded**. $P(x)$ *is bounded* $\forall x \in \mathcal{R}^p_+$.

Economic Assumptions (EA) on Ψ (Shephard, 1970)

- *EA4:* **Closeness**. Ψ *is closed,* $P(x)$ *is closed,* $\forall x \in \mathcal{R}^p_+$, $C(y)$ *is closed,* $\forall y \in \mathcal{R}^q_+$.

- *EA5:* **Convexity**. Ψ *is convex. The convexity of* Ψ *can be stated as follows:*

$$\textit{If } (x_1, y_1), (x_2, y_2) \in \Psi, \quad \textit{then } \forall \alpha \in [0,1] \textit{ we have}:$$

$$(x, y) = \alpha(x_1, y_1) + (1 - \alpha)(x_2, y_2) \in \Psi.$$

- *EA6:* **Convexity of the requirement sets**. $P(x)$ *is convex* $\forall x \in \mathcal{R}^p_+$ *and* $C(y)$ *is convex* $\forall y \in \mathcal{R}^q_+$.

  If Ψ is convex, then the inputs and outputs sets are also convex, *i.e.* $EA5$ implies $EA6$.

A further characterization refers to the Returns To Scale (RTS): CRS, IRS, VRS

II. Distances from the frontier

Efficiency measures à la Farrell-Debreu

- Efficiency input oriented:

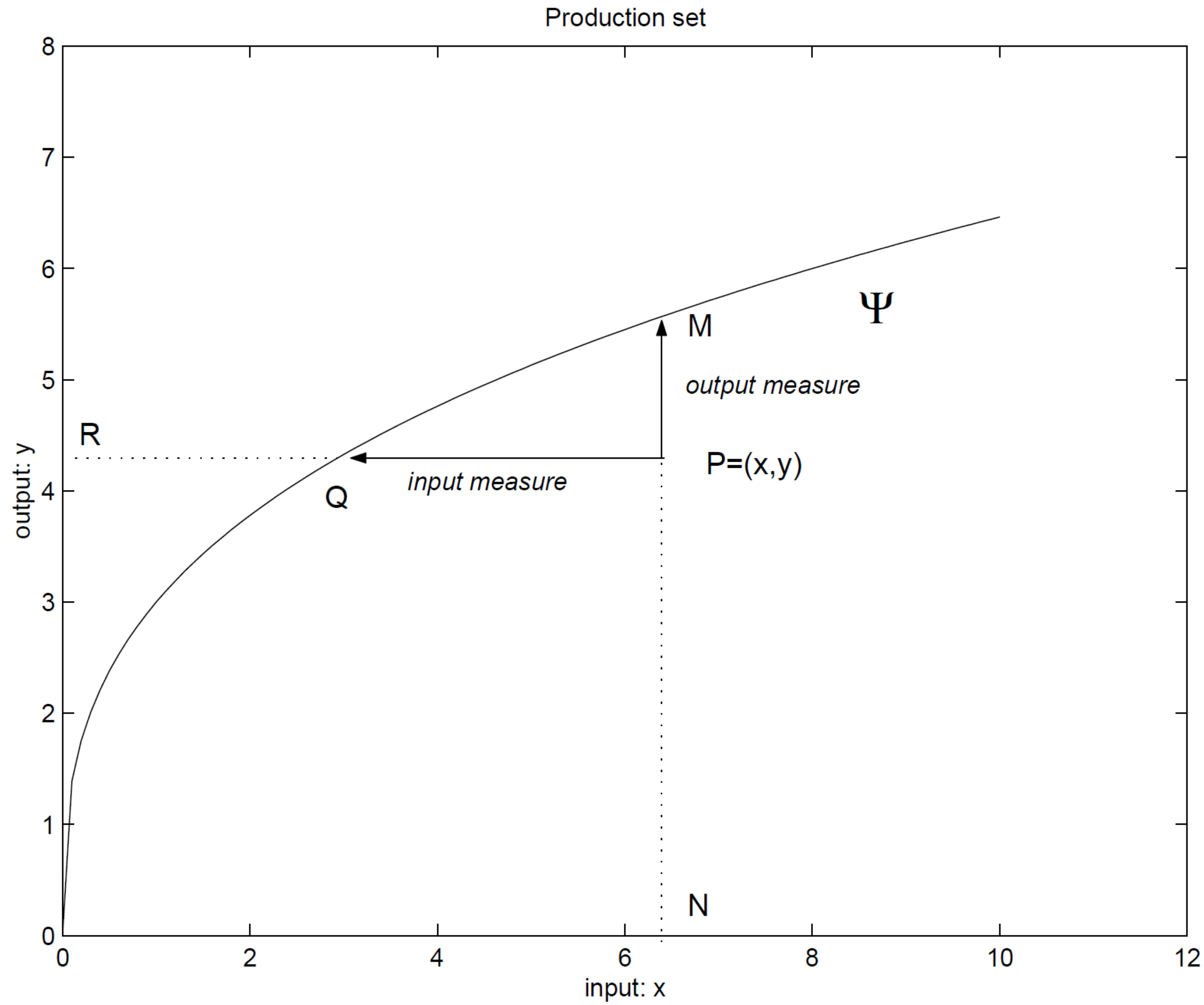
$$\theta(x_0, y_0) = \inf\{\theta | \theta x_0 \in C(y_0)\} = \inf\{\theta | (\theta x_0, y_0) \in \Psi\}$$

- $\theta(x_0, y_0) \leq 1$ radial contraction of the inputs that the firm (x_0, y_0) has to do to be considered *input-efficient*, i.e. $(\theta(x_0, y_0)x_0, y_0)$ is a frontier point.
- Efficiency output oriented:

$$\lambda(x_0, y_0) = \sup\{\lambda | \lambda y_0 \in P(x_0)\} = \sup\{\lambda | (x_0, \lambda y_0) \in \Psi\}$$

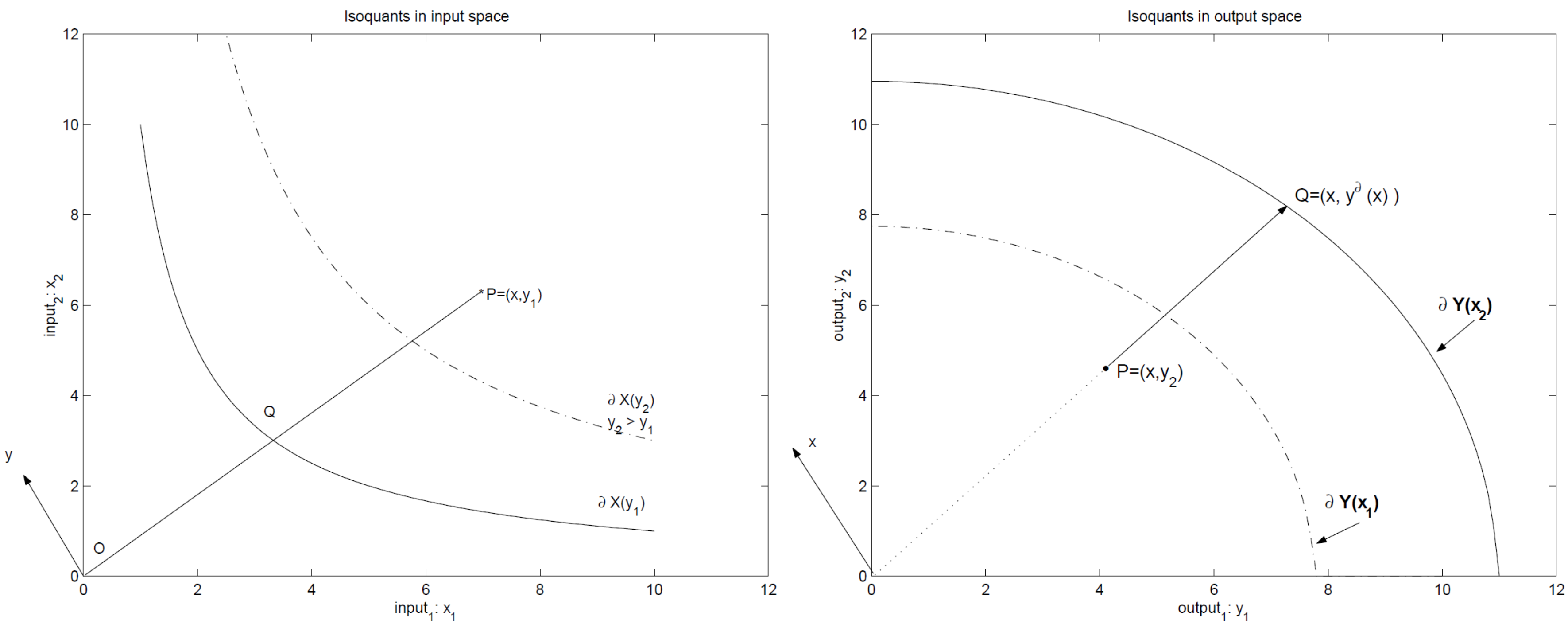
- $\lambda(x_0, y_0) \geq 1$ proportional increase of the output that the firm (x_0, y_0) has to do to be considered *output-efficient, i.e.* $(x_0, \lambda(x_0, y_0)y_0)$ is a frontier point.

An illustration



$$\theta_P = |RQ|/|RP| \leq 1 \text{ and } \lambda_P = |NM|/|NP| \geq 1$$

An illustration



$\theta_P = |OQ|/|OP| \leq 1$ (left) and $\lambda_P = |OQ|/|OP| \geq 1$ (right)

Distance functions à la Shephard

Distance function input-oriented:

- $\delta^{in}(x,y) \equiv sup\{\theta > 0|(\theta^{-1}x,y) \in \Psi\}$ with $\delta^{in}(x,y) \geq 1, \forall(x,y) \in \Psi$
- If $\delta^{in}(x,y) = 1$ then (x,y) is on the frontier of Ψ

Distance function output-oriented:

- $\delta^{out}(x,y) \equiv inf\{\lambda > 0|(x,\lambda^{-1}y) \in \Psi\}$ with $\delta^{out}(x,y) \leq 1$
- If $\delta^{out}(x,y) = 1$ then (x,y) is on the frontier of Ψ

Hyperbolic and Directional Distances

- Hyperbolic distances: operate simultaneously on input and output:

$$\gamma(x, y|\Psi) = \sup\{\gamma > 0|(\gamma^{-1}x, \gamma y) \in \Psi\}$$

- Directional Distances: project (x,y) on the technological frontier in the direction $d = (-d_x, d_y)$

(Chambers et al., 1998, Fare and Grosskopf, 2000):

$$\delta(x, y|d_x, d_y, \Psi) = \sup\{\delta|(x - \delta d_x, y + \delta d_y) \in \Psi\}$$

  - They are additive and allow for negative values of inputs and outputs.

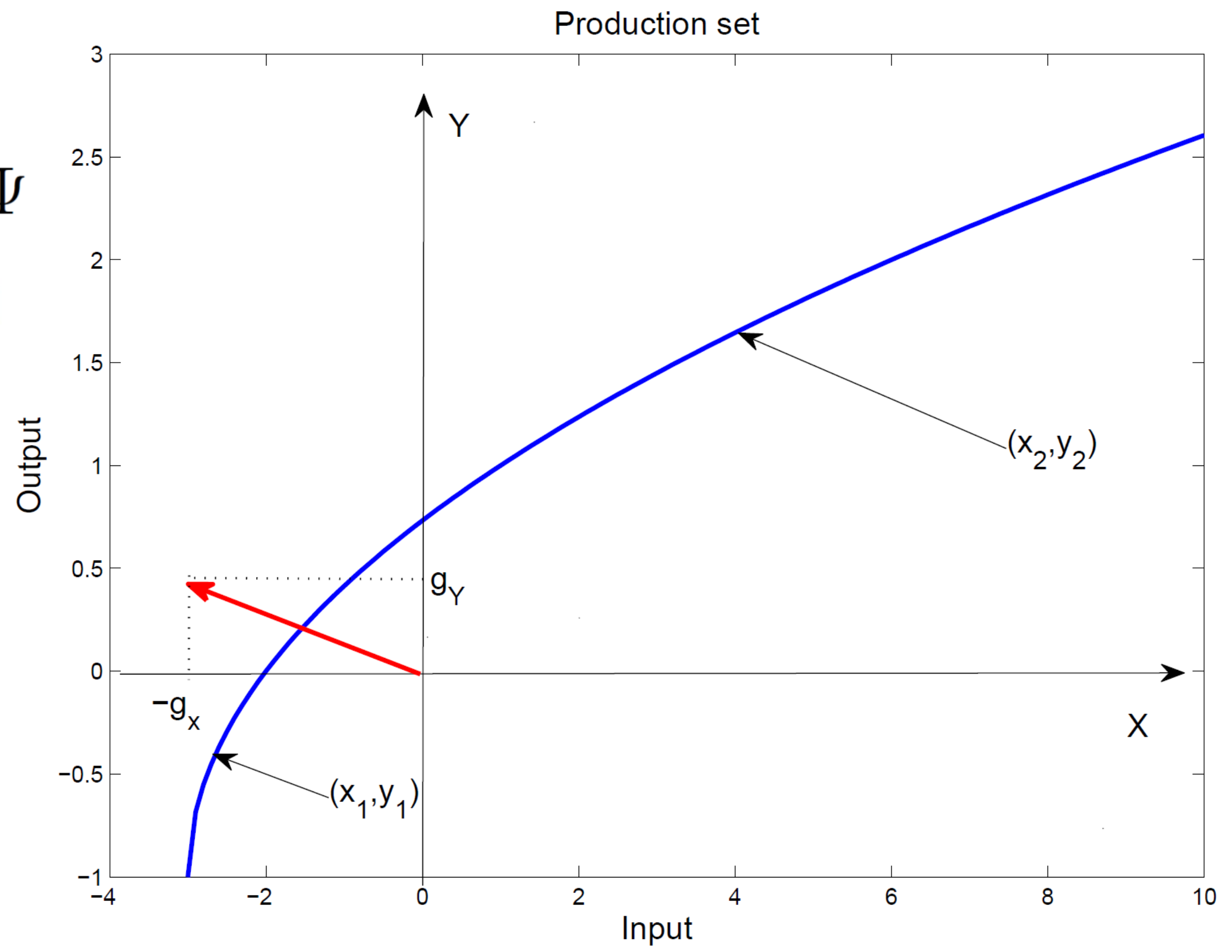
Directional distances

$$\delta(x, y|d_x, d_y, \Psi) \geq 0$$

If and only if $(x, y) \in \Psi$

$$\delta(x, y|d_x, d_y, \Psi) = 0$$

If (x,y) is on the frontier



III. «Classical» Models

Farrell (1957)

- Simple example: the hp. of CRS and knowledge of the efficient unitary isoquant SS' allow to measure the Technical Efficiency (TE), that is the ratio between the input used by a fully efficient unit that produces the same output and the input used by the analysed firm
- If a firm uses a level P of input to produce one unit of output we have:

$$TE = \frac{OQ}{OP} \leq 1.$$

Farrell (1957)

- If we have information on prices, we can determine the Cost Efficiency (CE), that is the ratio between the cost of an efficient firm which produces the same level of output and the cost of the analysed firm:

$$CE = \frac{0R}{0P}$$

- If the ratio among the prices is known (slope of the isocost line AA') we can determine the Allocative Efficiency (AE), that is the ability of the firm to select the optimal combination of inputs which minimize costs:

$$AE = \frac{0R}{0Q}$$

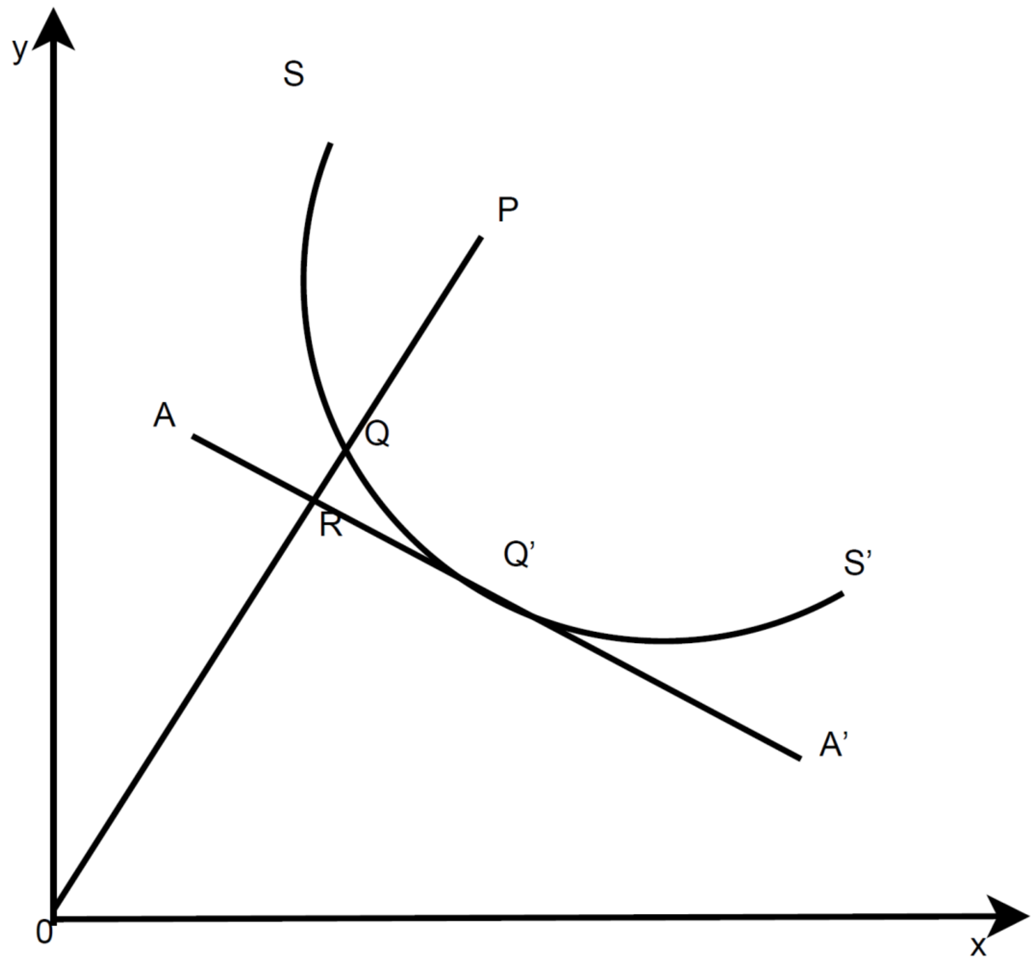
Farrell (1957)

$$TE = \frac{OQ}{OP} \leq 1.$$

$$CE = \frac{0R}{0P}$$

$$AE = \frac{0R}{0Q}$$

$$CE = TE * AE = \frac{0Q}{0P}\frac{0R}{0Q}$$



Charnes, Cooper, and Rhodes (1978)

- Input oriented model with CRS.
- They introduce fro the first time the name Data Envelopment Analysis (DEA): «mathematical programming model applied to observational data that provides a new way of obtaining empirical estimates of extreme relations such as the production functions and/or efficient production possibility surfaces that are a cornerstone of modern economics”
- Banker, Charnes, and Cooper (1984): propose the extension to VRS, distinguishing between Technical Efficiency (TE) and Scale Efficiency (SE)
- TE: radial distance of a firm from the efficient frontier
- SE: ratio between the average production of a firm and the average production of a firm which is efficient from a technical and a scale point of view.

Banker, Charnes, and Cooper (1984)

$$TE_{crs} = \frac{AP_{crs}}{AP},$$

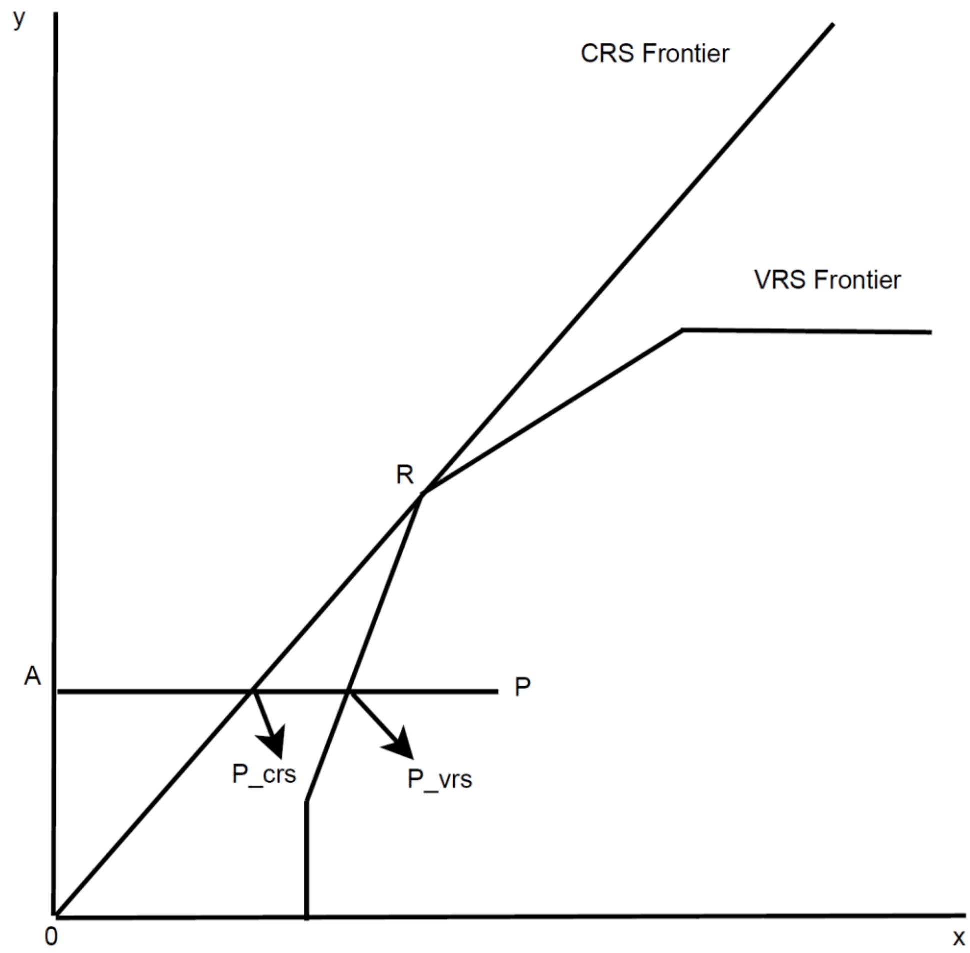
$$TE_{vrs} = \frac{AP_{vrs}}{AP},$$

$$SE = \frac{AP_{crs}}{AP_{vrs}},$$

$$0 \leq TE_{crs}, TE_{vrs}, SE \leq 1$$

$$TE_{crs} = TE_{vrs} SE$$

$$\frac{AP_{crs}}{AP} = \frac{AP_{vrs}}{AP} \frac{AP_{crs}}{AP_{vrs}}$$



IV. The problem of estimation

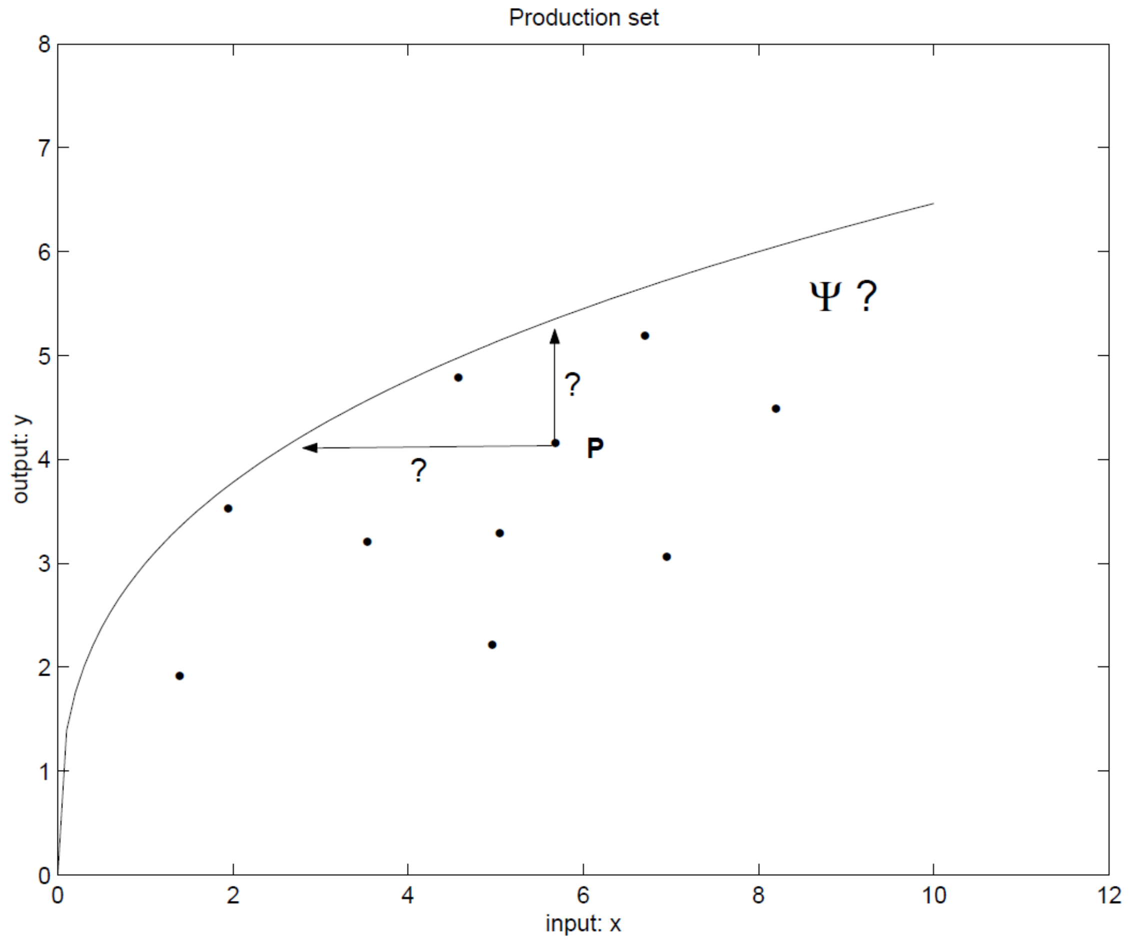
The problem of estimation

- Ψ is unknow, as $C(y), \partial C(y)$, and $\theta(x,y)$ $P(x), \partial P(x)$ $\lambda(x,y)$,....
- We observe a set of data $\mathcal{X} = \{(x_i, y_i),\ i = 1, ..., n\}$
- We need estimators: $\widehat{\Psi}, \widehat{\theta}, \widehat{\lambda}$, etc

Questions:

- How we define the estimators?
- Which properties should have?
- Can we test hypothesis, build confidence intervals, and so on?

The problem of estimation



It is a problem of estimation of a multidimensional frontier
We need to define a Data Generating Process (statistical model)

V. Taxonomy of efficient frontier models

Taxonomy of efficient frontier models

It is based on three criteria:

1. Specification of the functional form of the frontier
2. Presence of noise in the data
3. Type of data analysed

1. Specification of the functional form of the frontier

- Parametric Models: The production set Ψ is defined by a frontier function $g(x, \beta)$ which is a known mathematical function that depends on k unknown parameters, i.e. $\beta \in \mathbb{R}^k$, with y generally univariate.
- Advantages:
  - Economic interpretation of the parameter
  - Statistical properties of the estimators
- Disadvantages:
  - Choice of the functional form for the frontier $g(x, \beta)$
  - Multi-input multi-output

1. Specification of the functional form of the frontier

- Nonparametric Models: they do not assume any functional form for $g(x)$

- Advantages:
  - Robustness to the choice of model
  - Easy management of the multi-input multi-output case
- Disadvantages:
  - Estimate of unknown functionals
  - «Curse» of dimensionality

2. Presence of noise in the data

- Deterministic Models: assume that all the observations (X_i, Y_i) belong to the production set with $Prob\{(X_i, Y_i) \in \Psi\} = 1$, for each $i = 1, ..., n.$
  - Ψ is the support of (X,Y)
  - No noise in data
  - All the distance from the frontier is the inefficiency

- Problems:

  Influence/sensitivity to extreme values and outliers.

2. Presence of noise in the data

- Stochastic Models: there can be noise in the data, i.e. some observations can rely outside the production set Ψ.
- Convolution of $F = F_0 * G$, G is the distribution of noise
  - Ψ is not the support of (X,Y) but of the F_0
  - Distance from the frontier can be n(noise)
  - Problems: more complex and identification of noise from inefficiency.

3. Type of data analysed

- Cross-sectional models: the sample of data available is based on observations on n firms:

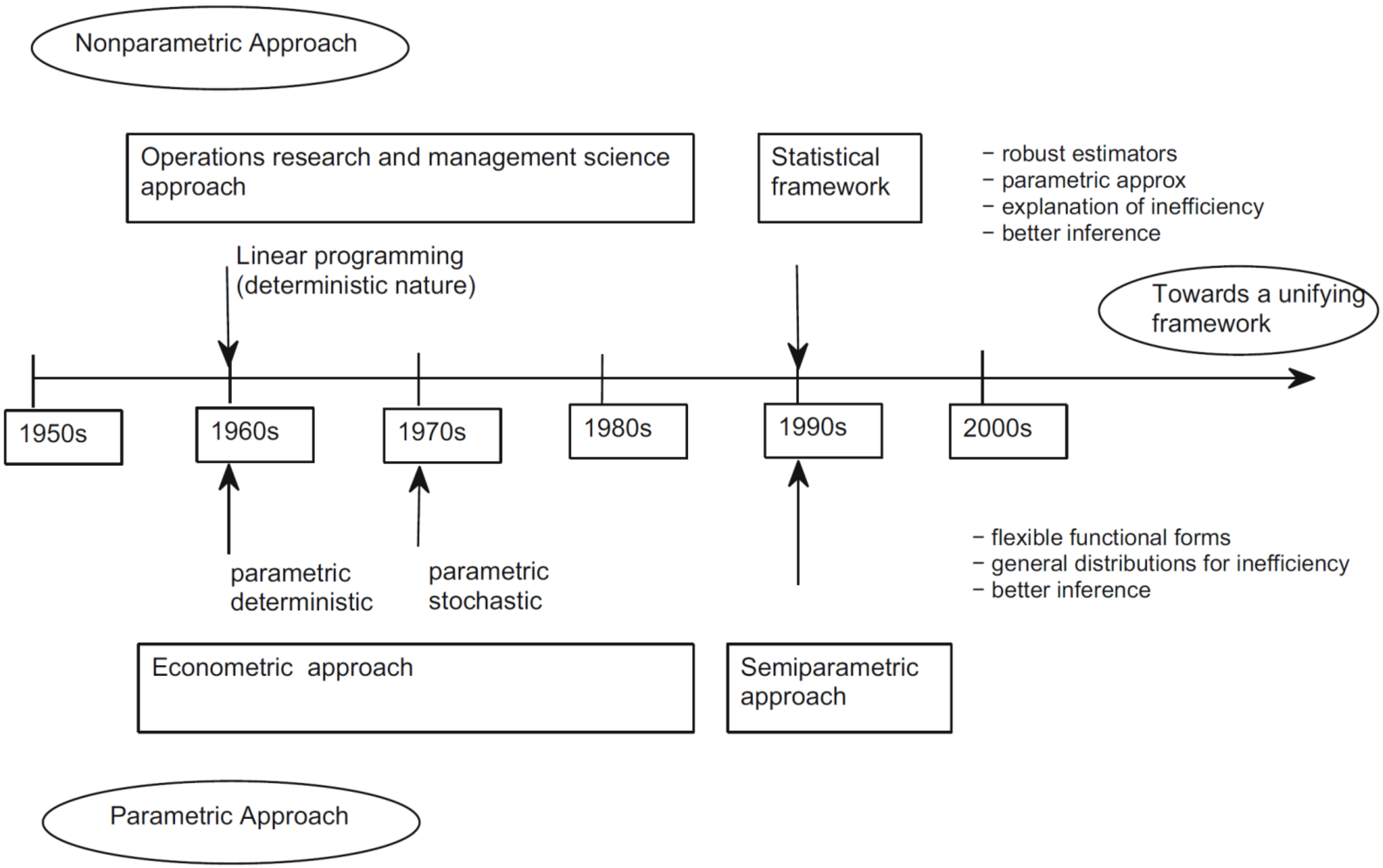
$$\mathcal{X} = \{(X_i, Y_i) | i = 1, ..., n\}$$

- Panel Models: the available data cover n firms observed on T periods:

$$\mathcal{X} = \{(X_{it}, Y_{it}) \mid i = 1, ..., n; t = 1, ..., T\}$$

VI. Choice of a model

Choice of a model



Quality as a latent heterogeneity factor in the efficiency of universities

Next RISIS Seminar Monday 10 May 2021 at 12:30

To participate write to:
serena.fabrizio@ircres.cnr.it

References

- Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, 30(9), 1078-1092.
- Chambers, R. G., Chung, Y., & Färe, R. (1998). Profit, directional distance functions, and Nerlovian efficiency. *Journal of optimization theory and applications*, 98(2), 351-364.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, 2(6), 429-444.
- Daraio C. (2021), A Framework for the Assessment and Consolidation of Productivity Stylized Facts, in C. F. Parmeter, R. C. Sickles (eds.), *Advances in Efficiency and Productivity Analysis*, Springer Proceedings in Business and Economics.
- **Daraio, C., Simar, L. (2007).** ***Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications*****. Springer Science & Business Media.**
- Farrell, M.J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, Series A, 120, 253-281.

References

- Färe, R., & Grosskopf, S. (2000). Theory and application of directional distance functions. *Journal of productivity analysis*, 13(2), 93-103.
- Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, 931-954.
- Shephard, R.W. (1970), *Theory of Cost and Production Function*. Princeton: Princeton University Press.
- Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of econometrics*, 136(1), 31-64.
- **Simar, L., Wilson, P. W. (2015). Statistical approaches for non-parametric frontier models: a guided tour. *International Statistical Review*, 83(1), 77-110.**

Other References

Badin L., Daraio C., Simar L. (2012), How to Measure the Impact of Environmental Factors in a Nonparametric Production Model, *European Journal of Operational Research*, 223, 818–833.

Badin L., Daraio C., Simar L. (2014) Explaining Inefficiency in Nonparametric Production Models: the State of the Art, *Annals of Operations Research*, 214, 5–30.

Daraio C. Simar L. (2014), Directional Distances and their Robust versions. Computational and Testing Issues, *European Journal of Operational Research*, 237, 358-369.

Daraio, C., Simar L. (2016), Efficiency and benchmarking with directional distances: a data-driven approach, *Journal of the Operational Research Society*, 67 (7), 928-944.

Daraio C., Simar L., Wilson P.W. (2018), Central Limit Theorems for Conditional Efficiency Measures and Tests of the “Separability” Condition in Nonparametric, Two-Stage Models of Production, *The Econometrics Journal*, 21, 170-191: Highly Cited WoS paper.

Badin L. Daraio C. Simar L. (2019) A Bootstrap Approach for Bandwidth Selection in Estimating Conditional Efficiency Measures, *European Journal of Operational Research* 277, 784–797.

Daraio C. Simar L. and Wilson P. (2020), Fast and Efficient Computation of Directional Distance Estimators, *Annals of Operations Research*, 288: 805–835