Quality as a latent heterogeneity factor in the efficiency of universities

Cinzia Daraio

Department of Computer, Control and Management Engineering *A. Ruberti* (DIAG)
Sapienza University of Rome, Italy
E-mail: cinzia.daraio@uniroma1.it
From a joint work with Leopold Simar and Paul W. Wilson

SAPIENZA
UNIVERSITÀ DI ROMA

RISIS Seminar
10 May 2021

Outline

"Quality" and its Impact on Productivity

- Quality is important but not easy to measure.
- Quality is linked to productivity and performance but there may be trade-offs between quality and efficiency.
- The role played by Quality is far from being unambiguously determined.
- We propose an approach to identify unobserved Quality factors, difficult to be observed in a direct way, but may have an impact on the productivity and performance, although its impact is not a priori known and must be empirically estimated.

Existing approaches

The most used approaches within the nonparametric efficiency literature are:

- one-stage approach, in which the Quality variables are included in the efficiency estimation as outputs;
- two-stage approach, in which the *unconditional* efficiency scores are estimated including only the inputs and outputs and afterwards are then regressed against Quality variables;
- conditional approach that includes Quality variables by conditioning to their values the production process.

Introduction: Our Approach to Quality

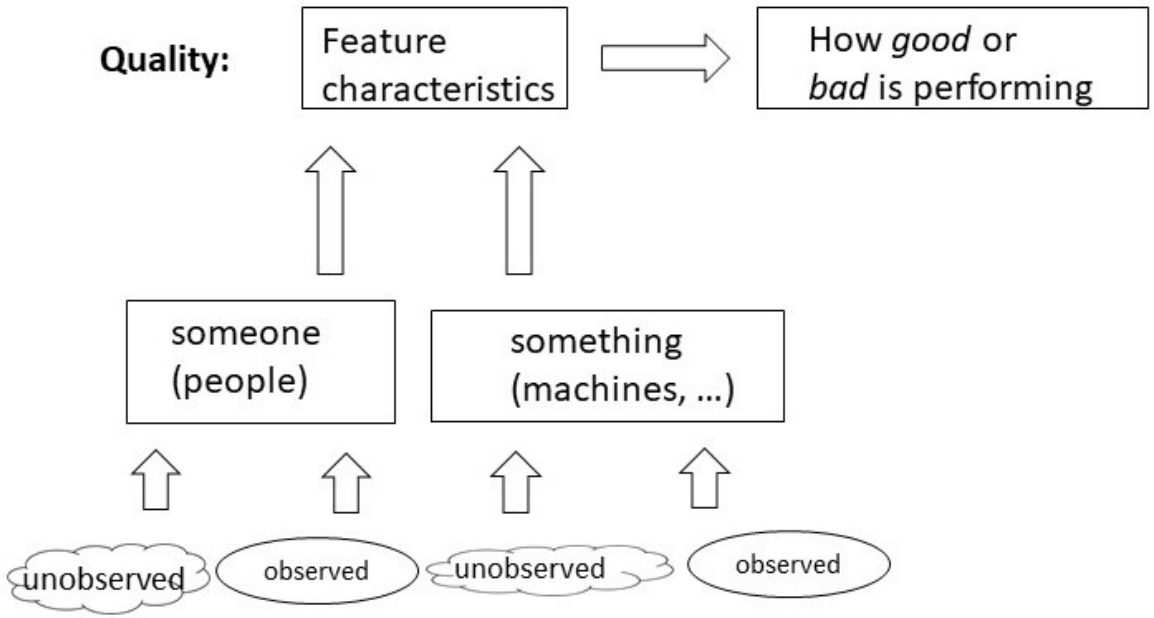


Figure: *An illustration of the concept of Quality and its ambiguity.*

Introduction: Our Approach to Quality

- Although there have been recent developments in the literature, the issue of Quality and its relationship to productivity and performance remains crucial and open in many fields of study, not only in the frontier-based quantitative analysis of performance.
- The selection of the Quality dimensions to be included in the empirical analyzes and the methods on how to include them, avoiding too many variables or too few, is a crucial operational issue at the intersection of the economics of production, management science and operational research with other disciplines.
- In the paper we provide an *overview of the interested literature*.

Introduction: Our Approach to Quality

- Recent works (Bounfour and Edvinsson, 2012) show that measuring and managing the intellectual capital of communities has the potential to change how public sector planning and development is done. Intangibles and intellectual capital have always been considered as relevant factors to the productivity and competitiveness of the private sector as well as of the public sector (Guthrie and Dumay, 2015; Dumay, Guthrie and Puntillo, 2015; Secundo, Lombardi, and Dumay, 2018).
- The measurement of intellectual capital (Bryl, 2018) is an emerging research area in *knowledge management* (Tiwana, 2000; Alavi and Leidner, 2001 and Liebowitz, 2012). However, being at its infant stage, it still lacks a rigorous methodology for being assessed, as it is the case for managerial quality, that remains difficult to be directly measured and included in a more general performance measurement system.

Our approach to quality and its impact on productivity

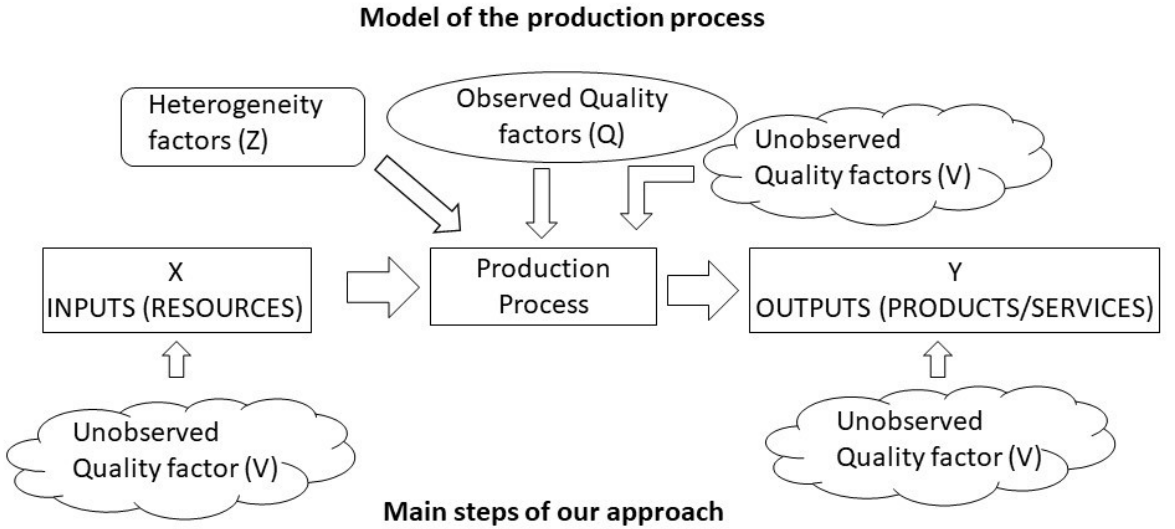


Figure: *A simple illustration of our approach.*

Quality in Higher Education

Table: Selected references on Quality in Higher Education.

Description	References
Conceptualization of Quality	Harvey and Green (1993); Sarrico et al. (2010) Williams and de Rassenfosse (2018)
Total Quality in HE	Lewis and Smith (1994)
Quality Assurance and regulation in HE	Westerheijden et al. (2007)
Total Quality Management in Education	Sallis (2002)
Quality Management in HE	Manatos, Sarrico and Rosa (2016); Sarrico (2018)
Econometric modelling of Quality	Daraio (2017, 2018a,b)
Human capital management and efficiency in HE	Kucharčíková et al. (2015)
Academic Quality (reputation and excellence)	Paradeise and Thoenig (2015)

Quality in Higher Education

- Universities carry out a complex production process. Multiple activities, such as teaching, research and third mission are realized by combining different resources: human capital, financial stocks and infrastructures to produce heterogeneous outputs, such as: undergraduate degrees, PhD degrees, scientific publications, citations, service contracts, patents, spin off and so on, within an heterogeneous environment in which size and subject mix play an important role (e.g. Daraio et al. 2015 JI and Daraio et al. 2015 EJOR and the references cited there).
- The concept of quality of HEIs is tricky (Sarrico et al. 2010), elusive and complex (Westerheijden et al. 2007) and multidimensional (Blackmur 2007).
- It's modeling in quantitative analysis is compelling and challenging (econometric modeling of quality, Daraio, 2017a).
- *In search of Academic Quality*(Paradeise and Thoenig, 2015): "Academic quality still remains a black box not only with regard to assessing the outputs, but also in terms of the formal and informal social, cultural and organizational processes adopted by specific university governance regimes". In this book quality is linked to the academic staff.

Modelling strategy: Conditional Efficiency

- Model for Production Process in the presence of observable heterogeneous factors $Z \in \mathcal{Z} \subseteq \mathbb{R}^d$, with inputs $X \in \mathbb{R}^p_+$ and one output $Y \in \mathbb{R}_+$

$$\begin{aligned} H_{X,Y|Z}(x,y|Z=z) &= P(X \leq x, Y \geq y|Z=z) \\ &= S_{Y|X,Z}(y|X \leq x, Z=z) F_{X|Z}(x|Z=z), \end{aligned}$$

- Conditional support of (X, Y), conditionally on $Z = z$

$$\Psi(z) = \{(x,y) \in \mathbb{R}^p_+ \times \mathbb{R}_+ \mid x \text{ can produce } y, \text{ when } Z = z\}.$$

- Marginal support of (X, Y) is given by

$$\Psi = \{(x,y) \in \mathbb{R}^p_+ \times \mathbb{R}_+ \mid x \text{ can produce } y\} = \bigcup_{z \in \mathcal{Z}} \Psi(z).$$

- Ψ is the (marginal) support of (X, Y), e.g. of

$$H_{X,Y}(x,y) = P(X \leq x, Y \geq y).$$

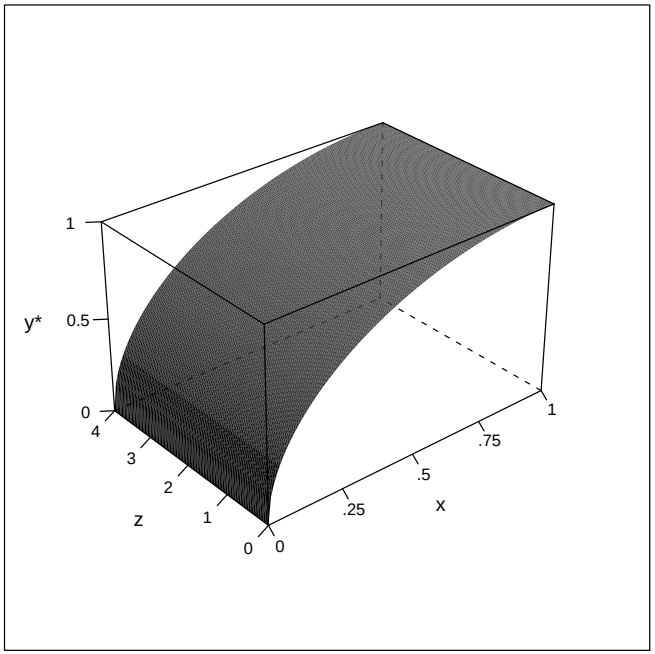
Separability condition

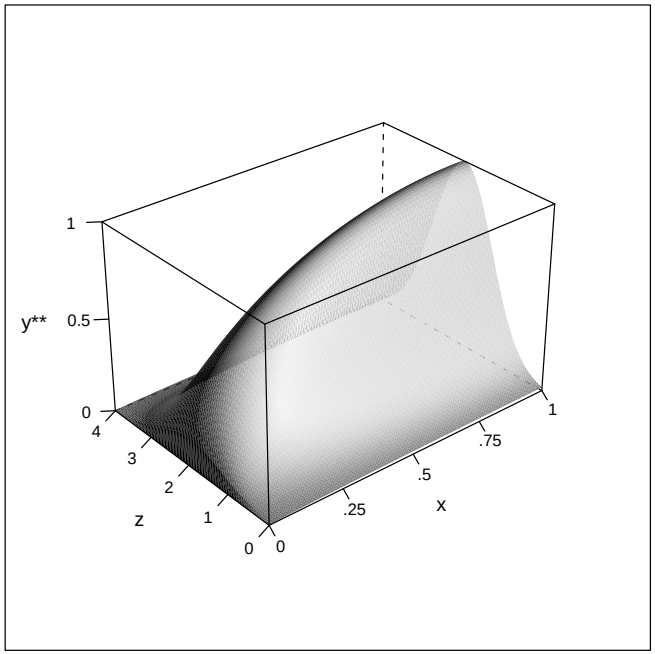
- "Separability" condition (Simar-Wilson, 2007):

$$\Psi(z) = \Psi \quad \text{for all } z \in \mathcal{Z}.$$

- Example from Simar-Wilson (2011): $g(X) = [1 - (X-1)^2]^{1/2}$

$$Y^* = g(X)e^{-(Z-2)^2 U} \qquad Y^{**} = g(X)e^{-(Z-2)^2} e^{-U}$$





Left Panel: **Separable**, Right Panel: **Not Separable**. In addition, the distribution of inefficiency U can or cannot be dependent on Z.

Conditional Efficiency

- If $\exists z \in \mathcal{Z}$ with $\Psi(z) \neq \Psi$ usual Marginal efficiency score has no practical economic meaning. It is given by

$$\begin{aligned}\lambda(x,y) &= \sup\{\lambda \mid (x,\lambda y) \in \Psi\} \\ &= \sup\{\lambda > 0 \mid H_{XY}(x,\lambda y) > 0\}\end{aligned}$$

- What we need: Conditional efficiency scores

$$\begin{aligned}\lambda(x,y|z) &= \sup\{\lambda > 0 \mid (x,\lambda y) \in \Psi(z)\} \\ &= \sup\{\lambda > 0 \mid H_{XY|Z}(x,\lambda y|Z=z) > 0\}\end{aligned}$$

  NB: The production frontier $\phi(x,z) = \lambda(x,y|z)\,y$

- Nonparametric Estimation

$$\widehat{H}_{XY|Z}(x,y|Z=z) = \frac{\sum_{i=1}^{n} \mathbb{1}(X_i \leq x, Y_i \geq y) K_{h_z}(Z_i - z)}{\sum_{i=1}^{n} K_{h_z}(Z_i - z)},$$

- Estimator of the conditional efficiency score (here FDH)

$$\widehat{\lambda}(x,y|z) = \max_{\{i \ | X_i \leq x, ||Z_i - z|| \leq h_z\}} \left(\frac{Y_i}{y}\right),$$

Conditional Efficiency

- Robust version (In SVVK 2016 order-m and univariate Y)

$$\phi_m(x,z) = E(\max(Y_1,...,Y_m)|X \leq x, Z = z)$$
$$= \int_0^\infty \left[1 - F_{Y|X,Z}(y|X \leq x, Z = z)^m\right] dy.$$

- Order-m conditional efficiency score

$$\lambda_m(x,y|z) = \int_0^\infty \left[1 - F_{Y|X,Z}(uy|X \leq x, Z = z)^m\right] du$$

  we have $\phi_m(x,z) = \lambda_m(x,y|z)\, y$.

- Nonparametric estimator

$$\widehat{\lambda}_m(x,y|z) = \int_0^\infty \left[1 - \widehat{F}_{Y|X,Z}(uy|X \leq x, Z = z)^m\right] du.$$

- Statistical properties: Cazals, Florens, Simar (2002) and Jeong, Park, Simar (2010), Bădin, Daraio, Simar (2012).

Conditional Efficiency and Separability

- Analysis of the effect of Z on the production process (see Daraio, Simar, 2005, 2007, and Badin, Daraio, Simar, 2012; Daraio, Simar, 2014)
  - Effect on the support of (X, Y) (efficient frontier)
    - Comparison of conditional and unconditional full efficiency scores (ratios and nonparametric regressions)
    - Robust versions : order-m (and order-α quantile) frontiers for large m (and $\alpha \to 1$)
  - Effect on the distribution of the inefficiencies by using e.g. $m = 1$ and $\alpha = 0.5$
  - See Badin, Daraio, Simar, 2012, 2014, Daraio and Simar, 2014, for details and tests.
- Test of the separability condition (Daraio, Simar and Wilson, 2018): $H_0 : \Psi^z = \Psi$, *for all* $z \in Z$ against $H_1 : \Psi^z \neq \Psi$, *for some* $z \in Z$.
- What to do if Z (or some component of Z) is latent (not observed)?

Unobserved Heterogeneity

Simar, Vanhems, Van Keilegom (2016):

- $V \in \mathbb{R}$ is not observed but may also influence the frontier level

$$\phi(x, v, z) = \sup\{y | F_{Y|XVZ}(y|X \leq x, V = v, Z = z) < 1\}$$

Simar, Vanhems and Van Keilegom follow Matzkin (2003):

- V is linked to one of the inputs X^1 through a flexible nonparametric nonseparable model $X^1 = \psi(W, V)$
- W is an auxiliary variable, correlated with X^1 and independent of V
- ψ is monotone increasing with V and $V \sim \text{Unif}(0, 1)$
- NB: We could also consider several relations $X^j = \psi_j(W^j, V^j)$, $j = 1, \ldots, p$.

Unobserved Heterogeneity

- Interpretation: V can be viewed as the part of X^1 that is independent of W.
  - In our case: W is the size of the university, X^1 is the input factor (or academic staff) and V is the part of X^1 not explained by the size
  - V could be linked to some proxies (see the following slides).
- According to Matzkin (2003), V is identified by the conditional cdf of X^1 given W: we have

$$V = F_{X^1|W}.$$

- In $X^1 = \psi(W, V)$, ψ is unknown and V is like a "residual", identified (under the monotonicity assumptions) by $V = F_{X^1|W}$.
- The choice of $V \sim \mathrm{Unif}(0, 1)$ is just a rescaling and so ψ can be interpreted as a quantile function.

Nonparametric Estimation

- Estimation of V_i

$$\widehat{V}_i = \widehat{F}_{X^1|W}(X_i^1|W_i) = \frac{\sum_{k=1}^{n} \mathbb{1}(X_k^1 \leq X_i^1) K_{h_w}(W_i - W_k)}{\sum_{k=1}^{n} K_{h_w}(W_i - W_k)}.$$

- Efficiency scores

$$\widehat{\lambda}(x, y|v, z) = \sup\{\lambda \mid \widehat{F}_{Y|X,\widehat{V},Z}(\lambda y|X \leq x, (\widehat{V}, Z) = (v, z)) < 1\}$$

$$\widehat{\lambda}_m(x, y|v, z) = \int_0^{\infty} \left[1 - \widehat{F}_{Y|X,\widehat{V},Z}(uy|X \leq x, (\widehat{V}, Z) = (v, z))^m\right] du,$$

$$\widehat{F}_{Y|X,\widehat{V},Z}(y|x, v, z) = \frac{\sum_{i=1}^{n} \mathbb{1}(X_i \leq x, Y_i \leq y) K_{h_v}(\widehat{V}_i - v) K_{h_z}(Z_i - z)}{\sum_{i=1}^{n} \mathbb{1}(X_i \leq x) K_{h_v}(\widehat{V}_i - v) K_{h_z}(Z_i - z)}.$$

- In practice we compute $\widehat{\lambda}(X_i, Y_i, |\widehat{V}_i, Z_i)$ and $\widehat{\lambda}_m(X_i, Y_i, |\widehat{V}_i, Z_i)$.
- We extend the separability test of Daraio, Simar and Wilson (2018) to unobserved heterogeneous factors.

Data sources

- Pioneering project Aquameth (Bonaccorsi and Daraio,eds. 2007; Daraio et al. 2011) and the feasibility study EUMIDA (Bonaccorsi,eds., 2014)
- Micro data validated by NSAs and certified by a Data Quality Report are available now for the European Higher Education Institutions (European Tertiary Education Register, ETER) Project, link to download the data for 2011-2016: http://eter.joanneum.at/imdas-eter/.
- ETER IV just started and will gather data on the next 3 years (2017–2019): check the website for updates!
- Interesting across-country comparison.
- A rich documentation available.
- Data integrated with other sources. Include Scimago Global Ranks, and other outputs such as PUB, %IC, NI and % Q1.

Data: Inputs, Outputs, Auxiliary variable and Factors

Table: Variables about European Universities used in the illustration.

Role	Acron.	Description	Source
Inputs	ACAD	Total number of academic staff	ETER
	NONAC	Total number of non-academic staff	ETER
	TEXP	Total expenditures in Euro PPP	ETER
Outputs	TDEG	Total number of degrees ISCED5-7	ETER
	PUB	Total number of publications	SCIMAGO
	PHD	Total number of PhD degrees	ETER
Qual./Ext. Factors	$V = UQUAL$	estimated by $\widehat{V}_i \in [0, 1]$ (see below)	our elab.
	$Z = SPEC$	Degree of specialization $\in [0, 1]$	SCIMAGO
Aux. var.	SIZE	Total number of enrollments	ETER
Obs. quality prox.	%REVTHIRD	Share of third party funds	ETER
	F. Year	Foundation year	ETER
	%IC	International Collaboration rate	SCIMAGO
	NI	Normalized Citation Impact	SCIMAGO
	%Q1	High Quality Publication ratio	SCIMAGO
	%Exc.	Excellence ratio	SCIMAGO
	%EwL.	Excellence with Leadership ratio	SCIMAGO
	WR	Scimago World Ranking	SCIMAGO
	RR	Scimago European Ranking	SCIMAGO

Unobserved quality: correlations with outputs and some observed proxies

Y_1	0.0609
Y_2 (FY)	0.5817
%REVTHIRD	0.5589
%IC	0.4405
NI	0.4508
%Q1	0.4785
%EXC	0.4751
%EWL	0.3549
WR	-0.6139
RR	-0.6050
F. Year	-0.2626

Test of the separability condition with Z unobserved

- $H_0 : \Psi^z = \Psi$, *for all* $z \in Z$
- against
- $H_1 : \Psi^z \neq \Psi$, *for some* $z \in Z$
- We extend Daraio, Simar and Wilson (2018) to unobserved Z
- p-value: 0.0037
- We reject H_0: we have a non-separable production process!

Also convexity highly rejected (Kneip, Simar and Wilson, 2015), p-value: 0.0000166!

Estimates of Efficiency: direction is egalitarian

Table: Estimates of Efficiency, direction is egalitarian: averages by country and standard deviations of the conditional measures $\beta(x, y|z)$ and $\beta_m(x, y|z)$.

Country	#obs	$\widehat{\beta}(x,y)$	$\widehat{\beta}(x,y\|z)$	std	$\widehat{\beta}_m(x,y)$	$\widehat{\beta}_m(x,y\|z)$	std
BE	5	0.1687	0.1152	0.1293	0.1196	0.1152	0.1293
CH	11	0.5883	0.2051	0.2207	0.5129	0.2051	0.2207
DE	73	0.9908	0.6996	0.6140	0.8801	0.6887	0.6066
DK	8	0.7121	0.4228	0.3848	0.6213	0.4179	0.3797
HU	7	1.0406	0.5463	0.4533	0.9870	0.4237	0.2954
IE	10	0.1293	0.0637	0.0990	0.1159	0.0637	0.0990
IT	60	0.1976	0.1137	0.1788	0.1504	0.1060	0.1681
LT	4	0.7334	0.2923	0.2242	0.7021	0.2923	0.2242
NL	13	0.3250	0.0579	0.0959	0.2190	0.0576	0.0954
NO	10	0.5508	0.4408	0.5428	0.5045	0.4373	0.5412
PT	17	0.1219	0.0723	0.1059	0.1075	0.0721	0.1059
SE	20	0.3445	0.2262	0.2866	0.3191	0.2260	0.2863
UK	96	0.0972	0.0621	0.1305	0.0184	0.0522	0.1156
EU	337	0.4072	0.2582		0.3374	0.2488	

Estimates of Efficiency: autonomous direction

Table: Estimates of Efficiency, direction is autonomous: averages by country and standard deviations of the conditional measures $\beta(x, y|z)$ and $\beta_m(x, y|z)$.

Country	#obs	$\widehat{\beta}(x,y)$	$\widehat{\beta}(x,y\|z)$	std	$\widehat{\beta}_m(x,y)$	$\widehat{\beta}_m(x,y\|z)$	std
BE	5	0.1609	0.0648	0.0857	0.1443	0.0648	0.0857
CH	11	0.3912	0.1699	0.2238	0.3411	0.1699	0.2238
DE	73	0.6984	0.4914	0.4472	0.6416	0.4880	0.4480
DK	8	0.5091	0.2981	0.3417	0.4633	0.2944	0.3408
HU	7	1.1710	0.4907	0.3665	1.1074	0.3996	0.2903
IE	10	0.1721	0.1264	0.2065	0.1618	0.1233	0.1982
IT	60	0.2779	0.1638	0.2909	0.2574	0.1579	0.2879
LT	4	1.6082	0.4993	0.4257	1.5668	0.4719	0.3993
NL	13	0.2042	0.0375	0.0615	0.1562	0.0372	0.0608
NO	10	0.7585	0.3205	0.3045	0.7342	0.3115	0.3059
PT	17	0.3204	0.2158	0.3035	0.3122	0.2147	0.3038
SE	20	0.4443	0.2684	0.2906	0.4223	0.2665	0.2882
UK	96	0.0896	0.0496	0.0998	0.0668	0.0457	0.0976
EU	337	0.3804	0.2245		0.3486	0.2184	

Impact on the efficient frontier

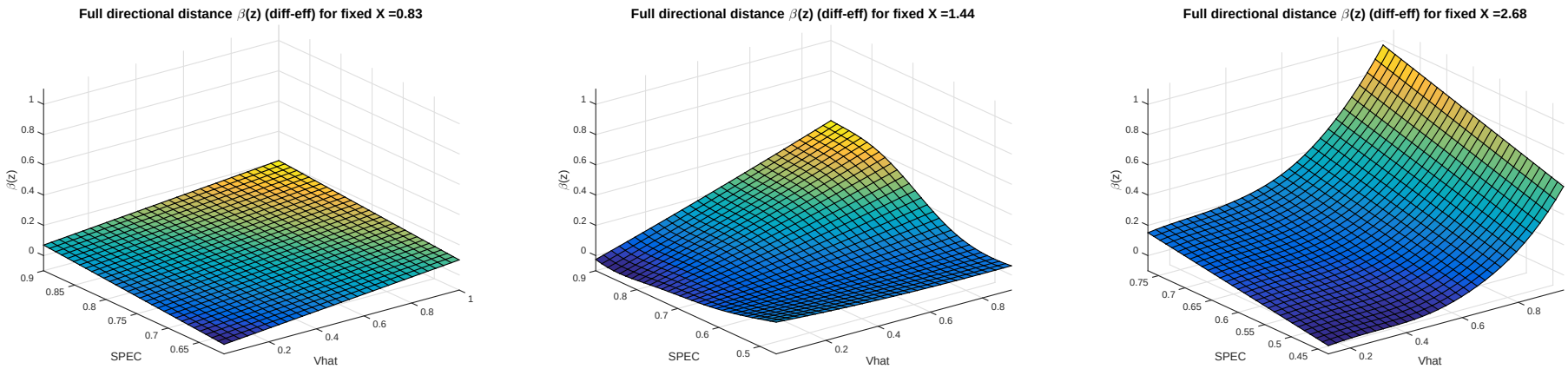
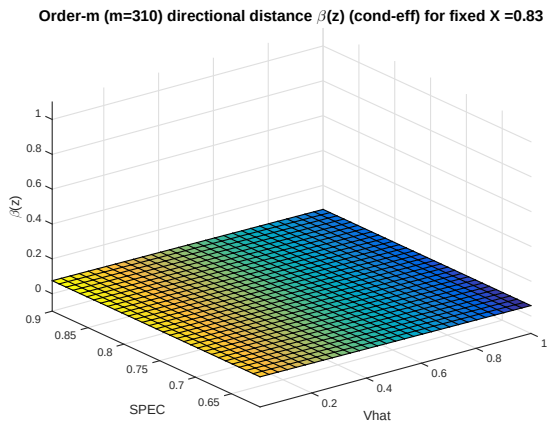
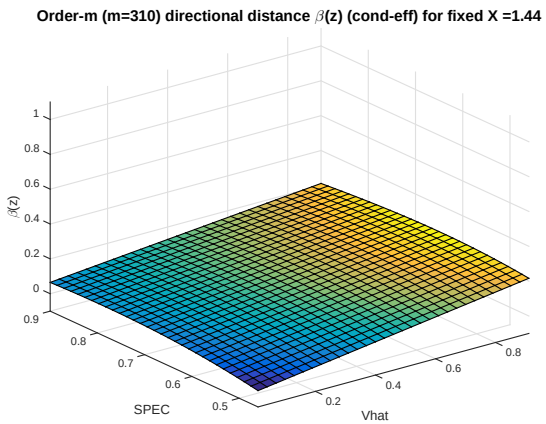


Figure: Impact of $\widehat{V} = QUAL$ and $Z = SPEC$ on the shift of the full frontier $\beta(x, y; 0, d) - \beta(x, y; 0, d|z, v)$, where $d = \text{med}(y)$ for fixed values of the Input Factor at the 3 quartiles: from the left to the right, small, median and large level of labor (academic staff).

Impact on the distribution of the (in-)efficiencies





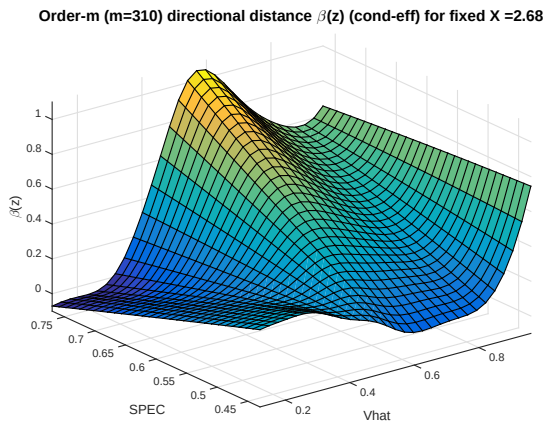


Figure: Impact of $\widehat{V} = QUAL$ and $Z = SPEC$ on conditional order-m efficiency measures $\beta_m(x, y; 0, d|z, v)$, where $d = \mathrm{med}(y)$ for fixed values of the Input Factor at the 3 quartiles: from the left to the right, small, median and large level of labor (academic staff).

Estimated gaps in the outputs

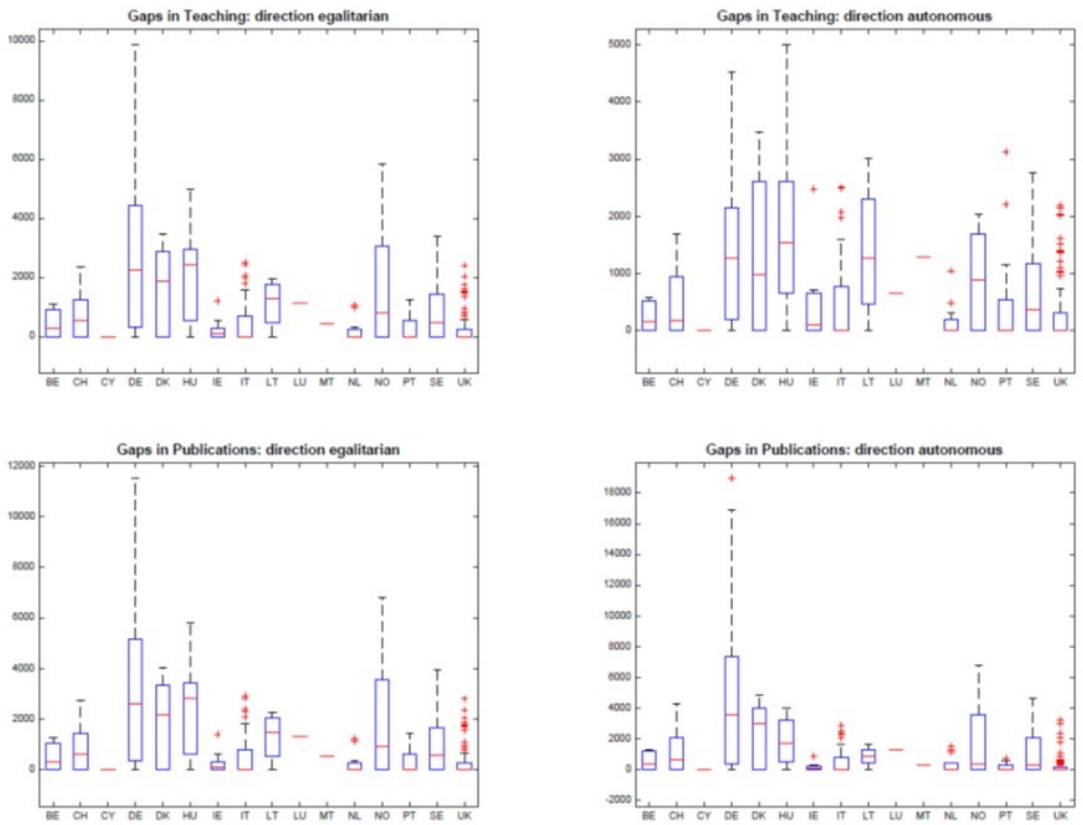


Figure: *Estimated gaps in the outputs. Top panels report the boxplots of the European countries considered following an egalitarian centralized path (median direction). Bottom panels show the boxplots obtained by selecting autonomous path (individual directions).*

Conclusions

- All MODELS for Quality are WRONG, but some are USEFUL (Box, 1979).
- The issue of Quality and its relationship with productivity and performance is a crucial operational issue in many fields of study including production economics, operations research, engineering and business management.
- In this paper we provide
  - a methodology for identifying latent quality factors,
  - estimate their statistical significance and
  - analyze their impact on the performance of the production process.
- This methodology is based on up-to-date computational methods and statistical tests for directional distances.
- We illustrated the approach on real data about the performance of European Universities.
- The approach seems promising and could be interesting to extend its application beyond the efficiency analysis stream of literature.

Main References

- **Daraio C., Simar L. and Wilson P. W. (2021), Quality as a Latent Heterogeneity Factor in the Efficiency of Universities, Economic Modelling, 99 (2021) 105485.**
- Daraio C. (2017a), A framework for the assessment of Research and its Impacts, *Journal of Data and Information Science*, Vol. 2 No. 4, 2017 pp 7-42.
- Daraio C. (2017b), Assessing research and its impacts: The generalized implementation problem and a doubly-conditional performance evaluation model, ISSI 2017 - 16th International Conference on Scientometrics and Informetrics, Conference Proceedings, pp. 1546-1557.
- Daraio C. (2019), Econometric approaches to the measurement of research productivity, in *Springer Handbook of Science and Technology Indicators* edited by Glänzel W., Moed H.F., Schmoch H. and Thelwall M. 633–666, ISBN 978-3-030-02511-3.
- Daraio C. (2018), The Democratization of Evaluation and Altmetrics, Technical Report n.1, 2018 Diag, Sapienza University of Rome.
- Daraio C., Simar L. (2007), *Advanced Robust and Nonparametric Methods in Efficiency Analysis. Methodology and Applications*, Springer, New York (USA).
- 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. (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.
- Daraio, C., Simar L., Wilson P. (2020), Fast and Efficient Computation of Directional Distance Estimators, *Annals of Operations Research*, 288, 805–835.