cepe
Centre for Energy Policy and Economics
Swiss Federal Institute of Technology Zurich

CANADIAN Building Energy End-Use DATA AND ANALYSIS CENTRE
CBEEDAC

SEEC
Surrey Energy Economics Centre

3rd International Workshop on Empirical Methods in Energy Economics (EMEE2010)

Surrey Energy Economics Centre (SEEC)
University of Surrey, UK
24th – 25th June 2010

NOTE:

The following Poster represents *Work in Progress* for presentation and discussion at the EMEE2010 workshop. It therefore must not be referred to without the consent of the author(s).

Sponsored by:

bp

BIEE
British Institute of Energy Economics

Department of Economics
UNIVERSITY OF SURREY

DIW BERLIN

Overcoming Data Limitations in Nonparametric Benchmarking: Applying PCA-DEA to Natural Gas Transmission

Maria Nieswand , Astrid Cullmann, Anne Neumann

Abstract

We empirically demonstrate a practical approach of efficiency evaluation with limited data availability in some regulated industries.

We apply PCA-DEA for radial efficiency measurement to US natural gas transmission companies in 2007.

PCA-DEA reduces dimensions of the optimization problem while maintaining most of the variation in the original data.

Our results suggest that the methodology reduces the probability of over-estimation of individual firm-specific performance.

Motivation

- Regulatory benchmarking often relies on limited data due to few firms and yearly conducted analyses.
- The need for a well represented technology yield in a benchmarking model that includes a high number of variables compared to the number of observations.
- In practice nonparametric benchmarking is widely applied, particularly Data Envelopment Analysis (DEA).
- However, DEA is sensitive to bad ratios of observations over variables.
- Thus, inefficiency is likely to be over-estimated in this setting and potential cost reductions due to regulation would remain uncovered.

Approach

- Create Principal Components (PCs) from original (output) data by means of Principal Components Analysis (PCA).
- Introduce a small number of PCs into the DEA procedure to lower dimensions while maintaining most of the original information: PCA-DEA.

Aim of paper

- Provide a pragmatic approach for European regulators who predominantly undertake efforts for national benchmarking .
- Apply PCA-DEA for the first time in the context of natural gas transmission regulation.

Methodology and Descriptive Statistics of Data

Linear Programming

Traditional DEA approach

$$\min_{\theta,\lambda} \theta$$
$$s.t.\ Y\lambda - s_Y = Y_j$$
$$-X\lambda - s_X = \theta X_j$$
$$e\lambda = 1$$
$$\theta, \lambda, s_Y, s_X \geq 0$$

PCA-DEA approach

$$\min_{\theta,\lambda} \theta$$
$$s.t.\ Y_{PC}\lambda - L_Y s_{PC} = Y_{PC,j}$$
$$-X\lambda - s_X = \theta X_j$$
$$L_Y^{-1} Y_{PC} \geq s_{PC}$$
$$e\lambda = 1$$
$$\theta, \lambda, s_{PC}, s_X \geq 0$$

where

θ	represents the relative efficiency of each firm contained in the set *J = {1,2,…,n}*
X_j, Y_j	represent column vectors of *k* inputs and *l* outputs of unit *j*
X	represents $k \times n$ matrix collecting column vectors X_j
Y	represents $l \times n$ matrix collecting column vectors Y_j
L_Y	represents a matrix collecting the weights obtained by PCA
Y_{PC}	represents a matrix containing the PCA-weighted original (output) data (Principal Components, PCs)
λ	represents the input and output weights
s_X, s_Y	represent the input and output slack variables
e	represents a unit vector

→ The programs are equivalent if all PCs are included in the PCA-DEA approach (i.e. all information is available)

Data

Since European natural gas companies are not easily comparable, we use the US natural gas market as our reference model. The US natural gas industry offers a comprehensive record of publicly available data and regulatory history, making it ideal for our analysis.

Table 1: Descriptive statistics of US natural gas transmission companies, onshore (2007), N=37

Variable	Opex	Total Deliveries	Transmission System	Peak Deliveries	Installed Horsepower	Transmission System Losses
Unit	mn USD	mn Dth	Miles	mn Dth	thou Hp	thou Dth
Sum	2,860.32	34,191.24	127,783.20	86.81	11,003.22	38,677.68
Min.	1.25	49.93	59.00	0.19	9.00	0.00
Max.	402.67	6,046.71	14,463.20	8.44	1,434.27	6,684
Mean	77.31	924.09	3,453.60	2.35	125.95	1,045.34
Median	31.50	403.89	1,680.40	1.68	297.38	615.66
Std. Dev.	99.61	1,255.53	3,703.33	2.12	371.72	1,399.32

Source: FERC Form No. 2

Model Specification

Figure 1: "OPEX-Benchmarking"

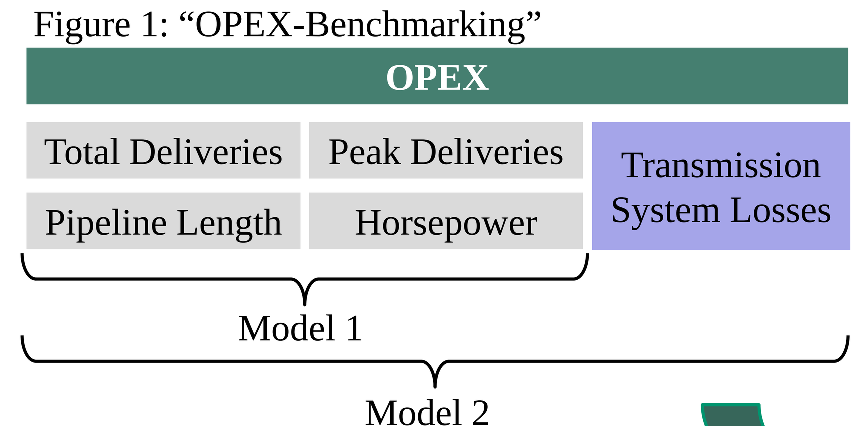

Table 2: Model specification

	Model 1		Model 2	
	DEA	PCA-DEA	DEA	PCA-DEA
DEA	x		x	
PCA-DEA		x		x
VRS	x	x	x	x

Note: x denotes the presence of the assumption in each model specification.

Use either four (Model1) or five (Model 2) outputs for Model Specification DEA or use two Principal Components out of four or five outputs for specification PCA-DEA.

Results and Conclusion

Selected Results

Table 3: Efficiency of US American natural gas transmission companies

	Model 1 (without TransLos)		Model 2 (with TransLos)	
Statistic	DEA	PCA-DEA	DEA	PCA-DEA
Minimum	27.02	19.23	30.65	19.10
25%-quantile	44.78	31.46	53.83	39.52
Mean	66.89	46.54	77.55	60.04
Median	63.86	39.51	93.45	48.39
75%-quantile	95.53	57.08	100	98.50
Maximum	100	100	100	100

Table 4: Peers of Northern Border Pipeline Company in PCA-DEA model specifications (FERC Form No. 2)

Variable	Opex	TotDeliv	TransSys	PeakDeliv	HorPow	TransLos
Unit	mn USD	mn Dth	Miles	mn Dth	thou Hp	thou Dth
NBPC	165.3	907.0	1,399	2.6	536.6	77.9
	Peers in Model 1					
I/I	9.3	420.6	414	1.4	78.3	489.4
TGPC	117.3	3,270.0	10,325	8.4	1,434.3	6,684.6
	Peers in Model 2					
DTI	70.7	1,360.1	3,344	4.0	350.2	398.5
EPNGC	373.4	6,046.7	10,240	5.1	1,136.4	3,038.8

Notes: NBPC = Northern Border Pipeline Company, I/I = IPOC as Agent/Iroquois Gas Trans Sys. LP., TGPC = Transcontinental Gas Pipeline Corporation, DTI = Dominion Transmission, Inc., EPNGC = El Paso Natural Gas Company

Technical Interpretation

- In both model specifications the first two Principal Components account for more than 90% of total data variation. Thus, both PCA-DEA specifications include two PCs while DEA specifications include four (Model 1) and five (Model 2) variables, respectively.
- Compared to the traditional DEA approach, PCA-DEA yields lower relative efficiency across both model specifications reflecting the *curse of dimensionality*.
- Comparing the particular specifications of Model 1 with their counterparts in Model 2 (including TransLos), we mostly observe higher efficiency in the latter model. We note that both PCA-DEA specifications exhibit the same dimensionality.
- PCA-DEA improves the identification of local conditions against which firms are compared to, i.e. the peers are structural more similar than under the traditional DEA approach.

Conclusions

- PCA-DEA improves the discriminatory power in nonparametric efficiency measurement and helps to overcome the practical obstacle of bad ratios of observations over variables.
- PCA-DEA notably reduces the probability of over-estimation of relative efficiency.
- Thus, our findings support current regulatory practice by mitigating the conflict between too few observations and the demand for many variables to produce an appropriate representa-tion of the relevant structures.

Selected References

Adler, N., and Yazhemsky, E. (2009) Improving discrimination in Data Envelopment Analysis: PCA-DEA or Variable Reduction. European Journal of Operational Research, doi: 10.1016/j.ejor.2009.03.050.

Dyson, R., Allen, R., Camanho, A., Podinovski, V., Sarrico, C., and Shale, E. (2001) Pitfalls and protocols in DEA. European Journal of Operational Research, 132, 245--259.

Haney, A. B., and Pollitt, M. G. (2009) Efficiency analysis of energy networks: An international survey of regulators (EPRG Working Paper 0915, Cambridge Working Paper in Economics 0926). University of Cambridge, Electricity Policy Research Group.

Jamasb, T., Pollitt, M. G., and Triebs, T. (2008) Productivity and efficiency of US gas transmission companies: A European regulatory perspective. Energy Policy, 36, 3398-3412.

Podinovski, V. V., and Thanassoulis, E. (2007) Improving discrimination in Data Envel-opment Analysis: Some practical suggestions. Journal of Productivity Analysis, 28, 117-126.