METODY ILOŚCIOWE W BADANIACH EKONOMICZNYCH
X, 2009, str. 20 - 29

COMPARISON OF THE DETERMINISTIC AND STOCHASTIC APPROACHES FOR ESTIMATING TECHNICAL EFFICIENCY ON THE EXAMPLE OF NON-PARAMETRIC DEA AND PARAMETRIC SFA METHODS

Agnieszka Bezat
Katedra Ekonomiki Rolnictwa i Międzynarodowych Stosunków Gospodarczych
SGGW
e-mail: agnieszka.bezat@ierigz.waw.pl

Streszczenie: The author considers in the article estimation of the technical efficiency, which measures the ability of the company to obtain the maximum output from given inputs or to use the minimum input to achieve given outputs. The comparison of two approaches: deterministic (on the example of Data Envelopment Analysis) and stochastic (on the example of Stochastic Frontier Approach) has been carried out, the advantages and disadvantages of both were also described. These methods were chosen because they have become popular in polish research. In the article the possible limitations and problems, which may influence results of studies conducted by using these methods, were considered.

Keywords: efficiency measuring, Stochastic Frontier Approach, Data Envelopment Analysis, Decision Making Units

INTRODUCTION

The empirical applications of efficiency analysis were conducted in such sectors like: accounting, advertising, auditing and law firms, airports, air transport, bank branches, bankruptcy prediction, community and rural health care, dentistry, education, electricity, environment, fishing, forestry, hospitals, hotels, macroeconomics, military, rail transport, sport, tax administration, water distribution etc. [Fried et al. 2008, p 16]. The DEA (Data Envelopment Analysis) and SFA (Stochastic Frontier Approach) are the main methods commonly used to estimate efficiency of a DMU – Decision Making Units (commercial entities that produce tangible goods and services that are sold in the market place, enterprise involved in delivering services or in the non-market sector, public bodies, national economic sector etc.).

The first method is a non-parametric, deterministic procedure for evaluating the frontier. Non-parametric procedures determine a frontier which "envelops" the observations. DEA employs flexible, nonparametric methods to construct the best-practice frontier and so allows the data to "speak for themselves" [Bates et. al 1996, p 1443]. For the deterministic approaches the frontier is defined by the maximum distances between input and output. Random error and characterises deviations from the frontier are interpreted as inefficiency [Coelli et al. 2005]. The DEA bases on a linear programming model which allows to build a piecewise linear frontier and assumes a convex production sets. A less constrained alternative to DEA is non-stochastic method known as free-disposal hull (FDH).

The second one widely uses stochastic procedure for parametric evaluating the frontier. The approach is stochastic – it considers additionally a random variable. The stochastic frontier approach treats deviations from production function as comprising both random error (white noise) and inefficiency [Mortimer&Peacock 2002, p 2]. This enables a distinction between a random symmetrical component which accounts for measurement errors and stochastic effects (e.g. due to weather influences) and a symmetric deviation component which represents the inefficiency. The SFA as a parametric approach requires assuming a specific function form a priori, the frontier is estimated econometrically by some variant of last squares or maximum likelihood [Coelli et al. 2005]. The SFA bases on econometric regression model, the frontier is smooth, appropriately and curved.

Both methods require all decision making units to have comparable inputs and outputs. Both methods can handle multiple input and multiple output models. These techniques should be used in conjunction with carefully complied data on input and output quantities and prices [Coelli et al. 2005, p 133].

DETERMINISTIC, NON-PARAMETRIC MODEL (ON THE EXAMPLE OF DEA)

The piece-wise-linear convex hull approach to frontier estimation, proposed by Farrell (1957), was considered by only a few authors in the two decades following Farrell's paper. Boles (1966), Shepard (1970) and Afriat (1972) suggested mathematical programming methods that could achieve the task, but the method had not received wide attention until the paper by Charnes, Cooper and Rhodes (1978), in which the term data envelopment analysis (DEA) was first used. Since then a large number of paper have appeared, which have extended and applied the DEA methodology.

For each DMU, to estimate efficiency, one would like to obtain a measure of the ratio of all outputs over all inputs, such as $u'q_i/v'x_i$, where u is a vector of output weights and *v* is a vector of input weights. The optimal weights are obtained by solving the mathematical programming problem [Coelli et al. 2005, p 162]:

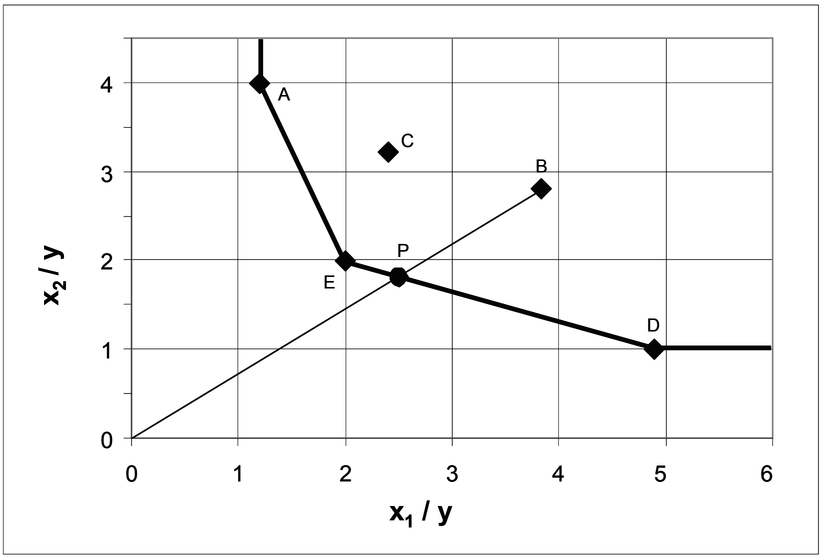
$$\max_{u,v}(u'q_i / v'x_i) \tag{1}$$

$$\text{st} \quad u'q_i / v'x_i \leq 1 \tag{2}$$

$$u, v \geq 0 \tag{3}$$

This involves finding values for u and v, such that the efficiency measure for the i-th DMU maximised, subject to the constrains that all efficiency measures must be less or equal to one. One considers the input-output combinations that are observed in the examined sample of decisions making units. An efficiency measurement is calculated basing on the distance between the input-output combination of the respective unit and the frontier [Farrell 1957].

Figure 1. A two input and one output case



Source: Cooper et al. 2007, p 57

The technical efficiency for DMU B is calculated as a ratio (s. Figure 1):

$$TE^B = \frac{0\text{P}}{0\text{B}} \tag{4}$$

Point P indicates a virtual DMU which is similar to B, but located on the frontier. Using virtual benchmarks such as P implies that linear combinations of real DMUs (in this case E and D) are considered.

STOCHASTIC, PARAMETRIC MODEL (ON THE EXAMPLE OF SFA)

Stochastic frontier analysis (SFA) is an alternative method to frontier estimation that assumes a given functional form for the relationship between inputs and an output [Coelli et al. 2005, p 209]. The stochastic production function model was proposed independently by Aigner, Lovell, Schmidt and Meeusen, van den Broeck in 1977. The model had following form:

$$\ln q_i = x_i'\beta + v_i - u_i \tag{5}$$

where q represents the output of the i-th DMU, x_i is a K×1 vector containing the logarithms of inputs, β is a vector of unknown parameters, and u_i is a non-negative variable associated with technical inefficiency, v_i is a symmetric random error, to account for statistical noise. The statistical noise arises from the inadvertent omission of relevant variables form vector x_i, as well as from measurement errors and approximation errors with the choice of functional form.

In the equation (5) the output values are bounded by the stochastic variable:

$$\exp(x_i'\beta + v_i) \tag{6}$$

The value v_i can be positive or negative and so the stochastic frontier outputs vary about the deterministic part of the model:

$$\exp(x_i'\beta) \tag{7}$$

The SFA requires choosing a production function model: Coob-Douglas, CES, translog, generalised Leontief, normalised quadratic and its variants. The translogarithmic and the Cobb-Douglas production functions are the two most common functional forms which have been used in empirical studies on production, including frontier analyses [Battese & Broca 1997, p 397]. Using the Cobb-Douglas stochastic frontier model[1] one can illustrate the variation of random error about the deterministic part. A Cobb-Douglas stochastic frontier model takes the form:

$$\ln q_i = \beta_0 + \sum_{j=1}^{k} \beta_j \ln x_{ij} + v_i - u_i \tag{8}$$

so

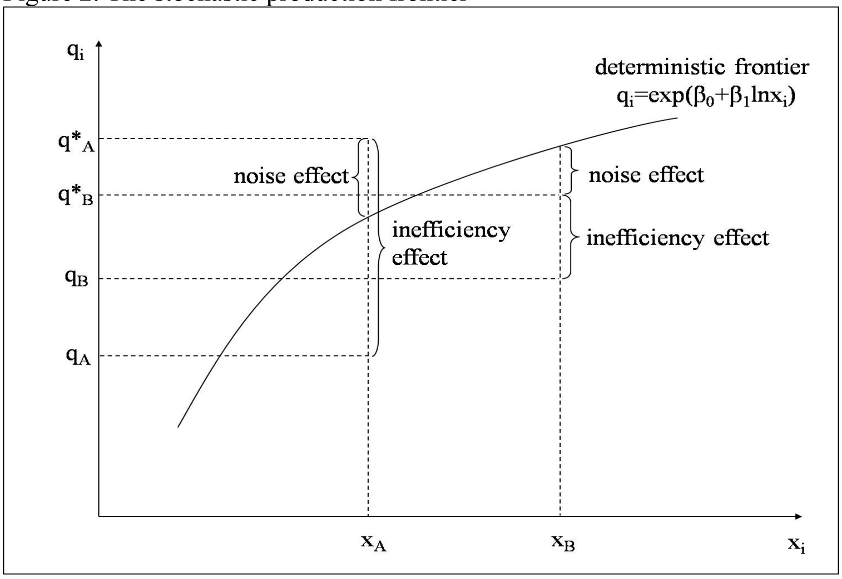
$$q_i = \exp(\beta_0 + \sum_{j=1}^{k} \beta_j \ln x_{ij}) + \exp(v_i) - \exp(u_i) \tag{9}$$

[1] The analysis is carried out for the DMU that produce the output q_i using only the input x_i.

The observed deviation from actual point of production to the frontier ($\exp(v_i)-\exp(u_i)$) is a composed error. Random component v refers to unsystematic deviations from the frontier – noise.

In the Figure 2 the deterministic model with the error noise and inefficiency effect was shown. The feature of frontier model was explained on the example of two DMUs: A and B, which are using the input x_A and x_B for producing the output y_A and y_B respectively.

Figure 2. The stochastic production frontier



Source: Own work on the basis of Coelli et al. 2005, p 244

In the figure the so-called frontier output is depicted. This measure assumes the case if there were no inefficiency effects (i.e., $u_A=0$ and $u_B=0$). The values q^* assume no inefficiency effects (i.e., if $u_A=0$ and $u_B=0$). For the DMU A the frontier output lies above the deterministic part of the production frontier (the value q^*_A) only because the noise effect is positive (i.e., $v_A>0$). The frontier output of the second DMU lies below the deterministic part of the frontier (the value q^*_B) because the noise effect is negative (i.e., $v_B<0$).

The frontier outputs tend to be evenly distributed above and below the deterministic part of the frontier. However, observed outputs tend to lie below the deterministic part of the frontier. Indeed, they can only lie above the deterministic part of the frontier when the noise effect is positive and larger than the inefficiency effect [Coelli et al. 2005, p 244].

Commonly using output-oriented measure of technical efficiency is the ratio of observed output to the corresponding stochastic frontier output:

$$TE_i = \frac{q_i}{\exp(x_i'\beta + v_i)} = \frac{\exp(x_i'\beta + v_i - u_i)}{\exp(x_i'\beta + v_i)} = \exp(-u_i) \tag{9}$$

Regarding to the equation (9) the TE takes a value between zero and one. It is important to understand that it measures the output of i-th DMU relative to the output that could be produced by a fully-efficient DMU using the same input vector. So using the SFA one can estimates a relatively efficiency of some DMU.

In series of studies authors have explored the implications of variety of distributional assumption an estimation of efficiency. Generally it is required to assume a distribution of *u* from:

half-normal distribution: $u_t \sim N^+(0,\sigma_u)$,

exponential distribution: $u_t \sim EXP(\lambda)$,

truncated-normal distribution: $u_t \sim N^+(\mu_u,\sigma_u)$,

gamma distribution: $u_t \sim \Gamma(m,\sigma_u)$.

The first two distributions have just a single parameter, are empirically traceable and easy to estimate. The last two distributions have two parameters, what makes them more flexible but also more difficult to estimate. The choice of distribution of *u* influences quite strongly a level of TE and less rankings of DMUs. Under fairly weak assumption it is usually possible and appropriate to estimate models using the method of last squares. Slightly stronger distributional assumption allows estimating the unknown parameters using maximum likelihood[2] or Bayesian techniques[3] [Coelli et. al 2005, p 240]. When decisions about function and distribution must be made, it is recommended to estimate a number of the alternative models and to select a preferred model using likelihood ratio test [Coelli 1996].

Commonly used method for estimation of stochastic frontier is a maximum likelihood (ML). ML estimations rest on the assumption that the distribution of the errors is actually known. Battese and Coelli (1992) propose a stochastic frontier production function which is assumed to be distributed as truncated normal random variables.

LIMITATIONS AND PROBLEMS WITH DEA APPLICATIONS

In conducting a DEA study some limitations and possible problems may encounter.

[2] Maximum likelihood estimators are popular because they have desirable large sample proprieties.

[3] Bayesian estimation is becoming increasingly popular, not least because it allows obtaining exact finite-sample results concerning nonlinear functions of the parameters.

- One of the most notable characteristics of DEA is its deterministic approach to efficiency measurement. DEA does not allow for estimation or measurement error. The full distance of a brand to the efficiency frontier is interpreted as inefficiency. But a measurement error or other noise and outliers may influence the shape and position of the frontier.
- The exclusion of an important input or output can result in biased results.
- Efficiency measurements can differ depending on the model specifications (input- vs. output-oriented models) and the variable specification (e.g. the degree of aggregation and the units used to measure inputs and outputs).
- The efficiency scores are only relative to the best DMU in the sample. The inclusion of extra DMU (e.g. from other countries) may reduce efficiency scores.
- It has been pointed out that the technical efficiency of any single enterprise or decision making unit (DMU) estimated using data envelopment analysis (DEA) will tend to decrease as the number of DMUs included in the DEA application increases [Zhang & Bartels 1998, p 187]. This is because, as the number of DMUs increases, the chance of encountering enterprises close to the true production frontier increases, and therefore the frontier constructed by DEA approaches the true frontier asymptotically as the number of enterprises in an industry increases [Banker 1989].
- DEA is good at estimating "relative" efficiency, thus the measurements are only valid in a sample. Units which have not been included in the sample can produce a shift of the frontier. The method's results say nothing about the efficiency of one sample relative to the other – they only reflect the dispersion of efficiencies within each sample. One should not compare the mean efficiency scores from two studies.
- The addition of an extra DMU in a DEA analysis cannot result in an increase in the TE scores of the existing enterprises.
- The addition of an extra input or output in DEA model cannot result in a reduction in the TE scores.
- When one has a few observations and many inputs and/or outputs, many of DMUs will appear on the DEA frontier. One could reduce the sample size and increase the number of inputs and/or outputs in order to increase the TE scores.
- Not accounting for environmental differences may give misleading indications of relative managerial competence.
- Standard DEA does not account neither for multi-period optimisation nor risk in management decision making.
- Since DEA is a nonparametric technique, statistical hypothesis tests are difficult.

LIMITATIONS AND PROBLEMS WITH SFA APPLICATIONS

In conducting a SFA study some limitations and possible problems may encounter. An analyst should keep these limitations in mind when choosing whether or not to use SFA.

- In SFA studies an assumption regarding to a specific functional form of stochastic frontier is required a priori. A wrong choice of production function may influence the results.
- The simple production frontier model does not permit the prediction of the technical efficiencies of companies that produce multiples outputs.
- The maximum likelihood does not allow assessing the reliability of inferences in small samples.
- Absolute level of TE is quite sensitive to distributional assumptions, rankings are less sensitive.
- The SFA requires using of large number of DMUs.
- It should also be stressed that all of issues described concerning to DEA method are also applicable (in varying degrees) to the stochastic frontier method.

CONCLUSIONS

The traditionally econometric belief in the presence of external forces contributing to random statistical noise is continuing maintained. Thus, it is desirable for the econometric approach to be relatively more successful than others, so as to provide the basis for a subsequent investigation into determinants of variations in the efficiency. On the other hand, a researcher has to choose the functional form of the frontier and make an assumption regarding to distribution of variation in inefficiency. A wrong choice may be corrected on the basis of statistical test (e.g. likelihood ratio test or alternatively Wald's test).

In conducting a DEA study, the production frontier "envelops" the observations, thus it does not require an assumption of a functional form relating inputs to outputs. Random deviations from the frontier are interpreted as inefficiency; the statistical noise is not included. A disadvantage of DEA method is that a frontier position and efficiency scores may be strongly influenced by a measurement error or outliers. In the same way the exclusion of an important input or output results. The DEA method is useful for the relatively to SFA smaller sample. But one should be carefully because a small number of DMUs may causes that many of DMUs will appear on the DEA frontier and an average level of efficiency scores will be very high.

REFERENCES

Afriat S.N. (1972) Efficiency Estimation of Production Functions. International Economic Review, 13, 568 – 598.
Banker R.D. (1989) Econometric Estimation and Data Envelopment Analysis. Research in Government and Non-profit Accounting 5, 231 – 243.
Bates J.M., Baines D., Whynes D.K. (1996) Measuring the efficiency of prescribing by general analysis. Journal of Operational Research Society, 47 (12), 1443 – 1451.
Battese G.E., Broca S.S. (1997) Functional Forms of Stochastic Frontier Production Functions and Models for Technical Inefficiency Effects: A Comparative Study for Wheat Farmers in Pakistan. Journal of Productivity Analysis. 8, 395 – 414.
Battese G.E., Coelli T.J (1992) Frontier Production Function, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India. Journal of Productivity Analysis. 3, 153 – 169.
Boles J.N. (1966) Efficiency Squared – Efficiency Computation of Efficiency Indexes. Proceedings of the 39th Annual Meeting of the Western Farm Economics Association, 137 – 142.
Charnes A., Cooper W.W., Rhodes E. (1978) Measuring the Efficiency of Decisions Making Units. European Journal of Operational Research, 2, 429 – 444.
Coelli T.J., Rao D.S.P, O'Donell Ch.J., Battese G.E. (2005) An introduction to efficiency and productivity analysis (2nd ed.). USA: Springer.
Coelli T.J. (1996) A Guide to Frontier Version 4.1: A Computer Program for Stochastic Frontier Production and Cost Function Estimation. Armidale: University of New England, CEPA Working Papers, p 6.
Cooper W.W., Seiford L.M., Tone K. (2007) Data Envelopment Analysis. A Comprehensive Text with Models, Applications, References and DEA-Solver Software (2nd ed.). USA: Springer.
Farrell M.J. (1957) The Measurement of Productive Efficiency. Journal of the Royal Statistical Society, Series A (General) 120 (3), 253 – 290.
Fried H.O., Lovell C.A.K., Schmidt S.S. (2008) The Measurement of Productive Efficiency and Productivity Growth. USA: Oxford University Press US, 16 – 20.
Mortimer D., Peacock S. (2002) Hospital Efficiency Measurement: Simple Ratios vs Frontier Methods. Australia: Centre of Health Program Evaluation. (Working Paper 135).
Shepard R.W. (1970) Theory of Cost and Production Functions. Princeton: Princeton University Press.
Zhang Y., Bartels R. (1998) The Effect of Sample Size on the Mean Efficiency in DEA with an Application to Electricity Distribution in Australia, Sweden and New Zealand. Journal of Productivity Analysis. 9, 187 – 204.

Porównanie deterministycznych i stochastycznych podejść w wyznaczeniu efektywności technicznej na przykładzie metod: nieparametrycznej DEA i parametrycznej SFA

Streszczenie: Autorka analizuje w artykule efektywność techniczną, która pozwala na mierzenie zdolności przedsiębiorstwa do uzyskiwania maksimum efektów przy danym poziomie nakładów lub do wykorzystywania minimum nakładów do osiągnięcia danego poziomu wyników. Przeprowadzono porównanie dwóch podejść: deterministycznego (na przykładzie Data Envelopment Analysis) oraz stochastycznego (na przykładzie Stochastic Frontier Analysis), a mianowicie przedstawiono zalety i wady tych dwóch metod. Wybrano te metody ze względu na ich coraz szersze zastosowanie w polskich badaniach. W artykule uwzględniono możliwe ograniczenia i problemy, które mogą wpływać na wynik bada dań przeprowadzonych za pomocą omawianych metod.

Słowa kluczowe: mierzenie efektywności, Stochastic Frontier Approach, Data Envelopment Analysis, Decision Making Units