OPERATIONAL EXPENDITURE BENCHMARKING OF REGIONAL DISTRIBUTION UNITS AS A TOOL FOR EFFICIENCY EVALUATION AND DESIRED COST LEVEL ESTIMATION

INTRODUCTION

Evaluation of activity efficiency of regional power distribution units is nowadays one of the most important management tasks faced by a power distribution company. Regional power distribution units are responsible for performing basic functions of a distribution system operator as well as for correct spending of operational funds for efficient performance of these functions. One of the ways of approaching the above mentioned problem of funds allocation is the application of operational cost benchmarking methods. This approach was recently investigated and reported in numerous studies. Moreover, a regulatory practice for allowable revenue estimation was developed using various benchmarking models of distribution company operational expenditure (OPEX). The above mentioned methodology of operational cost allocation is applied in the presented paper to regional power distribution units of Enea S.A. This company was created in the beginning of 2003 in result of a merging process of 5 smaller companies situated in the north-west of Poland. The biggest company after merging included 10 basic regional distribution units. After consolidation this number increased to 32. The necessity to separate the distribution system operator functions from sales activity (resulting from 2003/54/EC directive) requires unification of the organization structure containing such units. The new structure is facing the budget allocation problem which implies the need for performing the benchmarking study. The aim of such a study is to develop a budget formation support system. It was undertaken by Power Distribution Department of Enea S.A. in cooperation with Computing Science Institute of the Poznań University of Technology.

CHOICE OF OPERATIONAL COST CARRIERS

The determination of operational efficiency of regional distribution units was performed using a special computing model tying observations of parameters characterizing the operational activity of individual units with the observed overall cost of this activity. The construction of such a model involves the choice of parameters characterizing a regional unit and having an important influence on the cost level (variable to be explained) as cost carriers (or cost explaining variables).

The initial choice of cost carriers given below was made according to suggestions obtained from Power Distribution Department of Enea S.A.:

- electric energy delivered to customers,
- number of distribution substations,
- surface density of electric energy delivered.

A verification of the above choice was made by checking correlation of particular carriers with the operational cost [1]. First, linear correlation coefficients of the explaining variables were calculated for every regional unit. In result, a correlation matrix \( R \) was obtained with the number of rows and columns equal to the number of explaining variables. A relatively high correlation between explaining variables indicates that they are carrying similar information. Then, the correlation coefficients between the explaining variables and the cost to be explained were calculated giving the vector \( R_0 \). A high value of the elements of \( R_0 \) indicates an important influence of explaining variables on the operational cost. The matrix \( R \) and the vector \( R_0 \) can be used to calculate the multiple correlation coefficient \( W_{bw} \). This coefficient gives a measure of linear dependence between the variable to be explained and the linear combination of explaining variables. The multiple correlation coefficient is calculated according to the formulae given below:

\[
W_{bw} = \sqrt{1 - \frac{\text{det } R^*}{\text{det } R}} \tag{1}
\]

where

\[
R^* = \begin{bmatrix} 1 & R_0^T \\ R_0 & R \end{bmatrix} \tag{2}
\]

\( \text{det} \) - the determinant of a matrix.

When the value of multiple correlation coefficient is equal to 0, there is no linear dependence between the variable to be explained and the linear combination of explaining variables; when the coefficient is equal to 1, such a dependence is linear. In our case, the value of the multiple correlation coefficient is equal to 0.96 which indicates a strong dependence between the combination of variables chosen as cost carriers and the operational cost itself.

DUMMY VARIABLE AS SUBSTITUTE TO COST CARRIERS

The simultaneous analysis of influence of the five variables on the operational cost can be simplified by a reduction of explaining variables. In the case of reduction to only one substitute variable, the efficiency of regional units can be illustrated on figures having only two coordinates – the substitute variable and the cost.
To calculate a substitute variable, the principal component analysis (PCA) was applied [2]. This method enables to transform an \( n \)-element vector of \( n \) explaining variables \( x^T = [x_1, x_2, \ldots, x_n] \) into another \( n \)-element vector \( t^T = [t_1, t_2, \ldots, t_n] \) having the useful property of carrying important information in its first components. Such a transformation enables reduction of variables to a small number of important substitute variables and to neglect higher index variables of vector \( t \) without an important loss of information. The base of the method is the assumption that the high content of information is reflected by a high variance. Let \( x^T = [x_1, x_2, \ldots, x_n] \) be a vector of values of explaining variables for regional distribution unit \( i \). So, if we want to reduce the dimension of vector \( x \) to a single variable, we should transform it to variable \( t_1 = \alpha^T x \) choosing \( \alpha \) in a manner to obtain \( t_1 \) with the highest variance possible. The value of substitute variable \( t_1 \) for regional distribution unit \( i \) is denoted by \( t_i \) and can be obtained from the following formula:

\[
t_i = e_1^T x_i,
\]

where:

\( e_1^T \) – the transposed eigenvector having the greatest eigenvalues.

The calculations of principal components \( t \) were performed using the chosen cost carriers for 32 regional distribution units. The necessary eigenvalues and eigenvectors were computed using function \texttt{eig()} \textit{from the Matlab system}. The first principal component \( t_1 \) was recognized to have a sufficient content of information comparing with the chosen cost carriers – the ratio of the first principal component to the sum of all components was grater than 0.75.

**DETERMINATION OF REGIONAL DISTRIBUTION UNITS EFFICIENCY BY REGRESSION METHOD**

Regression is one of few methods available to estimate efficiency of units in the case where the main activity of such units lies in the area of natural monopoly (power, gas or water distribution networks) and the market verification of unit efficiency cannot work.

In this study the regression method [1] was used to estimate the efficiency of regional distribution units conducting the operation and maintenance of power distribution network of Enea S.A. However, because of different structure of five companies before the merging process leading to the creation of Enea S.A., the costs of operational activity for the unit were known only for 10 of them being formerly a part of the company in Poznan area. In other companies merged, smaller in size that the one in Poznan, the separate cost budgets for regional distribution unit were not subdivided before consolidation. Poznan area regional units conducted only the distribution tasks without sales which were performed in sales and customer service departments. To create the budgets for the distribution units in the newly created company, the efficiency of regional units (active already in the desired shape in the Poznan area) was to be verified first because for these units the operational parameters and the costs were known and, moreover, cost regression function obtained in such an analysis could serve next to determine the cost for other units. The calculation of cost regression for Poznan area units was performed in function of one substitute variable characterizing the distribution unit size obtained by PCA method and in function of 5 original variables chosen as cost carriers. The multiple regression gives function in the multidimensional space where the dimensions are the chosen operational parameters and the cost. In general, the resulting regression cost function has the following form:

\[
K_{OLS} = \sum_{i=1}^{n} a_i x_i + b,
\]

The numerical calculations were conducted in Excel using function \texttt{reglinp()}. The resulting parameters are coefficients \( a_i \) to \( a_5 \) (only \( a_5 \) in the case of substitute variable \( t_1 \)) and the parameter \( b \) being the value of cost for zero values of variables characterizing the operational tasks. The function \( K_{OLS} \) obtained by ordinary least square (OLS) method is describing the linear correlation of cost and the operational parameters and enables to estimate the efficiency of any regional distribution unit in comparison to the average performance. Basing on obtained function, two methods of efficiency correction were proposed for inefficient units. The corrected efficient cost is the result of cost reduction by a value calculated as the difference of the cost spent by the unit and the corrected \( K_{OLS} \) function obtained by:

- shifting it in the manner to let it pass through the most efficient point and keep its direction (COLS),
- transforming it to the line having the same parameter \( b \) as OLS and bounding all real cost points above the \( x \) axis (CICOLS).

The linear functions OLS, COLS and CICOLS for the case of 10 Poznan distribution units are presented in Fig. 1. The performed calculations let to formulate the following conclusions:

- the results obtained show important differences in efficiency of operation of the units,
- the suggested improvements of efficiency following from the proposed methods lead to important cost reductions reaching even 50%,

![Fig.1 Efficiency estimation by regression method](image-url)
• it seems that COLS method is more convenient to evaluate the cost reduction for units big in size because the cost function can have negative values for some very small units,
• CICOLS proposes the corrections of efficiency in proportion to the scale of unit activity.

The functions OLS, COLS and CICOLS obtained by the regression method can serve as a base to propose an efficient cost level for regional units for which the real historical costs are unknown. Such cost levels can be obtained by substituting to the equation (4) the values of variables characterizing the operational parameters of a regional unit, in original or substitute form.

DEA APPROACH TO REGIONAL DISTRIBUTION EFFICIENCY

The DEA (Data Envelopment Analysis) method permits to estimate the unit efficiency relating its position to the efficient best practice frontier obtained by linear programming and being composed of linear pieces connecting the units being Pareto efficient. In DEA, there is no comparison of the unit results to average statistic values but individual linear programming tasks are resolved for every unit in order to estimate its performance within the set of other units.

In the case of considered regional power distribution units, the input variable enabling to conduct the operation and maintenance is the granted cost level. The size of tasks to be performed in the area of customer and network service, characterized by the chosen cost carriers, is reflected by the level of output variables. Application of the DEA method is presented below in its additive BCC version [3]; it involves the following form of the linear programming task:

$$\min_{\theta, \lambda, s^+, s^-} z_o = \theta - \varepsilon \cdot Is^+ - \varepsilon \cdot Is^-$$  \hspace{1cm} (5)

where:

- $Y$ – matrix of output variables with $u$ columns corresponding to the units and $s$ rows corresponding to the chosen output variables,
- $X$ – matrix of input variables with $u$ columns corresponding to the units and $m$ rows corresponding to the chosen input variables,
- $y_o$ – vector of $s$ output values of a particular unit considered,
- $x_o$ – vector of $m$ input values of a particular unit considered,
- $s^+$ – vectors of positive differences of output variables,
- $s^−$ – vectors of negative differences of input variables,
- $\lambda$ – vector of coefficients determining the combination of input and output values,
- $\theta$ – reduction coefficient applied to all inputs to improve the efficiency.

The calculation process is repeated once for every unit to be classified. The obtained values of objective function divide the set of distribution units into two subsets: these with value of $z^*_o=0$ that are efficient and determine the efficient frontier and those with objective function value $z^*_o > 0$ that are not efficient and located below the frontier. Scalar variable $\theta$ determines the proportional reduction of values of input variables to obtain an improvement up to full efficiency. Such reduction applied to input variable results in radial movement of the unit towards the frontier. The presence of constant $\varepsilon$ in the objective function let to solve the problem in a desired two stage sequence: first for reaching the efficient position in optimizing the value of $\theta$ and next, when the previous was not sufficient, to optimize the values of $s^+$ and $s^-$. The non zero values of differences $s^+$ and $s^-$ and the value of $\theta^* \leq 1$ identify the sources and the values of all possible inefficiencies. For one input and one output variable the efficient frontier is composed of linear pieces and can be presented on a plane.

The same calculations were also conducted for the CCR version of DEA [3]; in this version it is assumed that $1 \lambda > 1$ and thus the related constraint is removed from model (5). In the case when only one input and output variables are considered, the efficient frontier becomes the straight line. As for the case of the regression analysis, the efficiency of distribution units using DEA was estimated for two sets of variables. First one, composed of 5 chosen cost carriers as output variables and the operational cost as the input variable. The second one, composed of the substitute aggregate value $t_i$ as output variable and the operational cost as input variable. The linear programming task was formulated in form required by the function linprog() of the Matlab. For each given set of variables, the task was solved for every unit separately.

The results obtained for the units of the Poznan area are presented for BCC and CCR variants in Fig. 2. Using BCC, four efficient units were identified as able to draw the best practice piecewise frontier. Other units are characterized by different levels of inefficiency. The scale of the inefficiency can be estimated by the value of $\theta$, which determines the level of cost reduction related to real initial budget granted last year. The value of $\theta$ for the worst units is equal to 0.6 which indicates that the cost reduction should
reach 40% of initial budget. In Fig. 2 the cost reduction is illustrated by the length of the black line starting from the considered unit position and ending at the frontier line. Using CCR, the number of efficient points was reduced to one and the frontier became the straight line starting from the origin of the system of coordinates. The corrected efficiency according to CCR is independent of the scale of activity of the distribution unit and is equal to the ratio of output to input variable values which corresponds to the frontier line angle. The obtained frontiers (piecewise or straight lines) can be used to estimate the justified level of the operational cost for the unit with no granted budget. One way to conduct such estimation is the reading of cost value for the given value of substitute variable \( t_i \) from Fig. 2 making the necessary extrapolation of frontier lines in some cases.

The verification of such estimation can be made by the solution of linear programming tasks for 11 units including the 10 units of the Poznan area plus the unit considered with the value of cost equal to that obtained when calculating the efficient frontier line in function of \( t_i \). Such verification showed that for the value of cost obtained by using \( t_i \), the value of \( \theta \) is equal to 1 and it is also equal or very close to 1 when using multiple output variables. The conclusion is that the cost level estimated using substitute variable \( t_i \) and the frontier line gives a very good approximation of justified cost levels for the units considered. The verification of results obtained for \( t_i \) by joint consideration of the 10 units from the Poznan area and more then one new unit has not been performed because it changes the original data set too much.

**DATA ANALYSIS OF REGIONAL DISTRIBUTION UNITS USING DOMINANCE-BASED ROUGH SET APPROACH (DRSA)**

Rough set theory plays an important role in decision support systems. In particular, when risk and uncertainty are present, it can be a base for inferring decision rules which constitute a preference model discovered from data. The most important answers given by application of the DRSA method [4,5] to analysis of a data set given in form of a decision table describing the units by a set of condition attributes (characterizing the operational task to be performed) and by a decision attribute (cost of the operation) are the following:

- are the data in the decision table consistent?
- what types of strong decision rules can be discovered basing on decisions already made?
- how to use the discovered rules for decision support concerning budget granting for new units?

In our case the data describing distribution units were subdivided in two groups: the one with known condition and decision attribute values and the second with only condition attribute values known. The first group was used to discover the decision rules which could be applied to the second group for the assignment of its units to the classes of size associated with a specific cost range.

The first group was composed of the 10 units belonging to the Poznan area; they where described by condition attributes being the same explaining variables as before for regression and DEA analysis; the decision attribute was a discretized value of the substitute variable \( t_i \) obtained form the PCA. It corresponds to the size of units and, implicitly, to the cost range. The discretization of \( t_i \) was performed such as to obtain possibly high quality of approximation of the classification of the units. In result we obtained 5 classes of units coded as follows:

- \( d = BW \), units „very big” in size with \( t_i > 2.00 \),
- \( d = W \), units „big” in size with \( t_i > 0.825 \),
- \( d = S \), units „medium” in size with \( t_i > 0.75 \),
- \( d = N \), units „small” in size with \( t_i > 0.65 \),
- \( d = BN \), units „very small” in size with \( t_i > 0.53 \),

where \( d \) denotes the decision attribute.

To perform the DRSA analysis of the decision table constituted in this way, we applied the rough set analyzing tool 4eMka [6]. No „reduct” of attributes, able to approximate the classification with the same quality as using the whole set of condition attributes, was found. Using this tool, we discovered 4 decision rules of “at most” type and 4 decision rules of “at least” type. They constitute a preference model of the deciding entity granting operational budget to the units from the Poznan area. This model can be used to classify units which did not receive the budget yet. Such a classification enables to define the cost range of these particular units.

This classification, which implies the cost allocation, can be compared with the costs obtained from regression analysis and DEA. The discovered inconsistencies, i.e. the reduced cost lower or higher than the cost range of the class to which the unit has been assigned using DRSA, can be an additional indication supporting the final decision about the level of the efficient cost for a unit.

**Fig. 3. Comparison of cost reduction with the cost range corresponding to classification to class S using the DRSA method**

Fig. 3 presents an example of the discovered inconsistencies of CICOLS, BCC and CCR cost corrections with the cost range corresponding to classification of units to class S using the DRSA method. The reduced cost levels for CICOLS, BCC and CCR are lower then the lower limit of the cost range corresponding to class S. The limit is marked in Fig. 3 by the red lower horizontal line. The BCC corrections and actual costs are rather consistent with the cost range.
CONCLUSIONS

The main goals of the conducted analysis concerning the cost efficiency of power distribution units performing the operational task within the power distribution company were:

- the estimation of efficiency level for the distribution units with the known operational costs,
- the estimation of efficient cost levels of operational activity for the new budget units conducting the same activity.

The application of results obtained within the first goal to the second one adds new useful elements supporting the decision making process concerning the budget for new units with no history in this matter.

The analysis was performed using three groups of methods: regression, DEA and rough set analysis.

Two approaches were applied to the description of the unit operational task: multi-dimensional one, where cost carriers were the 5 real values characterizing the distribution unit, and substitute one, where the size of the distribution unit is represented by the substitute variable $t_i$. The reduction of the number of variables (from 5 to 1) permits to avoid the repetition of the treatment of the same information which could be contained in different variables and, moreover, it facilitates the analysis of results obtained.

The application of various methods for estimation of efficiency seems to be valuable in the supporting such important decision as annual operational cost allocation. The DEA identifying the best practice units using linear programming techniques identifies the necessary corrections in a completely different way than regression analysis where the artificial average unit performance is created forming the base for efficiency corrections.

The problem for DEA results may be the existence of several efficient units in comparison with the only one resulting from the regression analysis. The improvement of efficiency in such DEA cases should perhaps limit the operational cost of every efficient unit to the extent granting the same resulting efficiency taking into account the scale of operational activity.

The variety of methods used seems to yield less possible errors in decision making process because it can be supported by the efficiency correction estimates resulting from application of additional knowledge about the possible inconsistencies of results concerning especially the efficiency improvements basing on traditional methods.

Rough set method seems to be an interesting approach for the verification of results obtained using DEA and regression analysis. The classification of distribution units into classes of size associated with cost ranges gives the deciding entity the each of them. The additional justification for this variety is the limited number of units with complete data concerning the operational cost, as well as the risk associated with data uncertainty and small differences concerning operational conditions of various distribution units.

REFERENCES


