OPERATIONS RESEARCH AND DECISIONS
No. 1 2012
DOI: 10.5277/ord120103

Maciej MALARA*
Mariusz MAZURKIEWICZ*

MODELLING THE DETERMINANTS OF WINNING IN PUBLIC TENDERING PROCEDURES BASED ON THE ACTIVITY OF A SELECTED COMPANY

The purpose of this article is to identify the factors influencing the probability of winning in public procurement procedures and to assess the strength of their impact from the perspective of both: the bidder and procurer. The research was conducted with the use of series of quantitative methods: binary logistic regression, discriminant analysis and cluster analysis. It was based on a sample consisting of public tenders, in which the examined company performed the role of a bidder. Thus, the research process was aimed at both identifying the factors of success and estimating the probability of achieving it, where it was possible to obtain probabilities. The main idea of this research is to answer questions about the utility of various methods of quantitative analysis in the case of analyzing determinants of success. Results of the research are presented in the following sequence of sections: characteristics of the examined material, the process of modelling the probability of winning, evaluation of the quality of the results obtained.

Keywords: *binary logistic regression, discriminant analysis, public procurement law, public procurement procedures*

1. Introduction

In the era of the modern, global economy [10], [13], a major factor determining the survival of a company, as well as its further development, is its ability to constantly assess, improve and confront its capabilities in all areas of its activity. One of the forms allowing such action is participation in a competitive market and thus facing the necessity of undergoing evaluation. The situation is no different in the construction

*Institute of Organization and Management, Wrocław University of Technology, ul. Smoluchowskiego 25, 50-372 Wrocław; e-mail addresses: maciej.malara@pwr.wroc.pl, mariusz.mazurkiewicz@pwr.wroc.pl

sector, in which the most basic and common form of competition due to Polish legal regulations is to be involved in tendering procedures.

Given the premises above, the authors' aim was to investigate the factors affecting the probability of winning a tender and to assess the strength of their impact from the perspective of both bidder and procurer. Therefore, an attempt was made to join the stream of numerous [12], quantitative studies on tendering procedures being also widely studied in the field of game theory [11], however – with the use of a slightly different approach. The research was conducted using three quantitative methods: logistic regression, discriminant analysis and cluster analysis. The selection of these methods was based on the characteristics of the statistical data available. The sample comprised public tenders, in which the examined company performed the role of a bidder. The authors intend that their results allow fuller understanding of both the mechanisms which govern public procurement procedures and the detection of regularities associated with them, so that the bidding procedure (making an offer, but also the preceding such action decision) could be performed more efficiently, while the chances of victory could be significantly increased.

2. Characteristics of the research material

2.1. Legal framework of tendering procedures in Poland

The forms and fundamental legal framework for public procurements are specified by Act [14][1]. This act contains the current regulations concerning, in particular (Article 1): the principles and procedures of public procurement, forms of legal protection, control regarding the awarding of contracts, and the authorities competent in the matters covered by the act. Rules of awarding public contracts are, however, also contained in Prime Minister's regulations (such as *The kind of documents, which the purchaser may require from the contractor, and the forms in which these documents may be submitted*) and the relevant European Union directives, as well as many minor regulations and ministerial decrees[2].

---

[1]Act of 29 January 2004 Public Procurement Law. The consolidated text of the Act takes into account the changes introduced by the Act of 4 September 2008 amending the Act Public Procurement Law and some other acts (Journal of Laws No. 171, pos. 1058), the Law of 17 October 2008 amending the Law Public Procurement Law (Journal of Laws No. 220 item. 1420), the Act of 9 January 2009 on concessions for construction works or services (Journal of Laws of 2009 No. 19 pos. 101) and the Act of 21 November 2008 on the Civil Service (Journal of Laws No. 227, pos. 1505), the Act of 2 April 2009 amending the law on guarantees granted by the Treasury and certain legal entities, the Law on the Bank of the National Economy and certain other acts (Journal of Laws No. 65, item. 545).

[2]For e.g. regulations issued by the Ministry of Infrastructure.

2.2. Characteristics of the research sample

The research material consists of the data based on all the public tenders, in which the examined company participated over a calendar year, as a bidder. The research perspective assumed enables a more comprehensive look at the company's participation in market competition but also poses specific difficulties in the form of having to organize observations according to a changing set of characteristics, individual and appropriate to each case. In total, 113 tenders have been characterized by 14 variables which were acquired from the data set. This information exists in various forms, both: quantity (different rules for selecting contractors in particular tenders) and configurations. The consequence of these actions was the selection of attributes representative of all the analyzed cases.

With exception of the variables, *declared duration of investment*, *number of competitors* and *price offered* which were expressed as the corresponding numerical values, all the attributes are encoded by successive and consecutive arabic numerals (1, 2, 3, ..., *n*). It should be noted that the attributes *win*/*loss*, *presence of co-operators, declared warranty*, *future operating costs*, *investment duration considered*, *rejection of the bid at the preliminary stage*, taking into account their qualitative nature, are encoded in binary form according to the scheme: 1 – win , 0 – loss (in the case of the first attribute); otherwise: 1 – factor was taken into account during the assessment by the evaluation committee; 0 – factor was not taken into account, for the rest of the variables mentioned above. In addition, some qualitative variables had more than two categories (*type of construction*, *type of the tender*, *size of the procurer's city of origin*).

The variables are defined as follows:

Dependent variable:
Y – binary variable win (1)/loss (0).

Quantitative variables:
*X*1 – price offered by the company studied (in PLN),
*X*2 – declared duration of investment (in days),
*X*3 – number of competitors,
*X*4 – lowest bid submitted in the tender (in PLN),
*X*5 – highest offer made in the tender (in PLN).

Qualitative variables:
*J*1 – nature of the work,
*J*2 – type of tender,
*J*3 – size of procurer's city of origin,
*J*4 – presence of partner contractors (co-operators),

$J5$ – indicator of whether warranty was taken into account when assessing the bid,

$J6$ – indicator of whether operating cost was taken into account when assessing the bid,

$J7$ – indicator of whether duration of the investment (as in the cases of variables $J5$ and $J6$),

$Y1$ – indicator of rejection of an offer at the preliminary stage of tender (if occurred).

3. Modelling the probability of winning the tender

3.1. Overview of the research tools

3.1.1. Binary logistic regression

Of the three methods used, the primary research instrument was binary logistic regression [7]. It is a statistical tool used to describe the influence of several variables (quantitative or qualitative) on one dependent binary variable. In addition, binary logistic regression allows us to estimate the probability of the occurrence of an event (choosing one of several possibilities). Logistic regression is based on the method of expressing a probability in terms of *odds.* Instead of determining a probability classically (using the ratio of the number of successes to the number of all trials), the odds are calculated as the ratio of the number of successes to the number of failures [2, 9]. The odds can, therefore, be calculated based on classical probability ([6], p. 3)

$$\text{logit}(p) = \ln(\text{odds}) = \ln\left(\frac{p}{1-p}\right) = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \quad (1)$$

where p denotes the likelihood of the occurrence of an event such that $p \in (0, 1)$, $\boldsymbol{\beta}$ is a vector of unknown parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)^T$.

3.1.2. Discriminant analysis

Discriminant analysis ([1], p. 882) is a method used to determine the best way of dividing a given set of cases into groups based on a set of variables. It helps to decide whether the groups differ according to the average of a certain variable, and then uses this as one variable to predict group membership. From a numerical point of view, analysis of the discriminant function is very similar to analysis of variance (ANOVA). Using this method, the following two main steps are taken:

• *construction and learning* in which a classification rule based on the data set used for learning (statistical sample) is found,

• *classification/using the model* where the classification of a basic set of objects whose affiliation is unknown to us occurs based on the classification rule found above.

Forecasting based on the classification model takes place by building a function described by the equation ([9], p. 321):

$$f = \sum_{i=1}^{k} b_i X_i \tag{2}$$

where b_i, $i = 1, 2, ..., k$ is the weight of variable I in classification function, X_1 ..., X_k – explanatory variables.

In the process of calculating the weights (b_i) for each of explanatory variables in the classification function, the data set is divided into two separate groups according to the dependent variable ([3], p. 281). In the case of the research described in this paper, the first group of tenders will contain those awarded (won), in the second – those lost (not won).

Finally, the separating value is determined by calculating the f_0 function, which allows the separation of the two subspaces:

$$f_0 = \frac{\overline{f_1} + \overline{f_2}}{2} \tag{3}$$

where the symbols $\overline{f_1}$ and $\overline{f_2}$ denote the arithmetic means of the theoretical values of the classification function for observations belonging to subspaces 1 and 2, respectively, which in the case at hand, are expected to enable the classification of tenders according to the selected criterion (win, loss).

3.1.3. Cluster analysis

Cluster analysis [8] is an essential analytical tool in classification. Its purpose is to separate groups such that observations within each group are maximally similar and the observations belonging to different groups differ maximally. The forms of cluster analysis are divided onto *hierarchical*, where objects grouped together at an earlier stage of the analysis remain grouped together throughout the subsequent steps, and *non-hierarchical*, in which objects belonging to the same group at one stage of the analysis may belong to different groups according to the final solution (e.g. *k*-means). Cluster analysis is a method belonging to a group called methods of unsupervised learning. This method groups elements into relatively homogeneous classes. The basis of most clustering algorithms is the similarity between elements. This is expressed using the similarity function (metric). Basically, the grouping is made in order to obtain homogeneous classes to help distinguish the main features of each class and to reduce the large raw data set to a few basic categories that can be treated as objects for further analysis. The form of cluster analysis used in these studies is the nearest-neighbor method [5].

4.2. Presentation of the results obtained and assessment of their quality

4.2.1. Binary logistic regression

The final model was obtained by comparing numerous models obtained using various methods of selecting the explanatory variables for the logit regression equation [4]. These models were compared in terms of fit to the real data. Goodness of fit and the significance of the parameters were adopted as the decisive factors. The model which achieved the best fit was chosen. In this particular case, the *backward selection* method with the *likelihood ratio* criterion was adopted. In the analysis conducted, 96 of the 113 available tenders were used: 17 of them were omitted from the process, due to incomplete information for these cases (i.e. missing values). The coefficients in the model are statistically significant as a system at any level of significance (p-value = 0.000). The quality of the resulting model is shown in the classification table below (Table 1). The values on the leading diagonal of this table indicate the correctly classified cases. Classification is based on a standard split point equal to 0.5. A case is classified as 1 (winning the tender) if the probability of winning according to the model exceeds the value of 0.5 (split point). In this case, the split point has not been altered due to the high goodness of fit of the model chosen.

Table 1. Classification table for logistic regression

Observed			Expected		
			Y		Percentage of correct classification
			0	1	
Step 9	*Y*	0	87	1	98.9
		1	2	6	75.0
	Total percentage				96.9

Source: SPSS 17.0 processor.

The logit model for the probability of winning in the tendering procedure contains four variables and a constant. The estimates of the parameters for this model are presented in Table 2.

Table 2. Variables in the model

Variable	*B*	Standard error	Wald	Significance	exp*B*
*J*4	3.045	1.434	4.508	0.034	21.019
*X*1	–0.002	0.001	7.835	0.005	0.998
*X*4	0.002	0.001	7.884	0.005	1.002
*X*5	0.000	0.000	7.278	0.007	1.000
*Y*1	–25.594	7853.192	0.000	0.999	0.000
Constant	–2.528	0.920	7.550	0.997	0.080

Source: SPSS 17.0 processor.

The Wald statistics in Table 2, along with the corresponding significances, test the hypothesis regarding the significance of the impact of each variable contained in the model on the probability of winning the tender. The coefficients of variables *J*4, *X*1, *X*4, *X*5, and the constant are statistically significant. The coefficient of the *Y*1 variable is not statistically significant. However, removing this variable from the model leads to a loss of quality and this is a change so significant that it has been decided to leave this variable. Evaluation of the exp*B* indicators from the last column, which are interpreted in this case as describing the dynamics of the odds of victory in the tender, results in the finding that the variable *J*4 has a very strong influence on winning a tender. This effect should be interpreted as follows: bidding for a tender jointly with another contractor increases the odds of winning the bid by 2019%. Thus, we may assume that it is almost impossible to win a tender single-handedly.

In the cases of variables *X*4 and *X*5, the marginal impact on the probability of winning is relatively small. However, given that in the case of quantitative variables, the size of exp*B* indicates the change in probability, which is caused by a change in the value of the explanatory variable by one unit, a change of 0.2% should be considered as significant (in the case of variable *X*4). Thus, according to the model obtained, an increase in the maximum price occurring in the tender and an increase in the minimum price significantly increases the chances of victory. Variable *X*1 denotes the price proposed by the company studied – the model indicates that an increase in the value of this variable by one unit implies a reduction in the chances of winning the tender by 0.2%. Moreover, it seems reasonable to conclude that apart from the intuitive variables such as the prices offered by competitors (minimum, maximum), or the price offered by the company studied, the most important factor when applying for a contract is the presence (and thus help) of cooperating contractors.

4.2.2. Discriminant analysis

Another method used is discriminant analysis. This method was used to determine the attributes that affect the chances of winning a tender. In the first step of the analysis, the assumption that the explanatory variables have a normal distribution was examined. Qualitative variables were excluded from this analysis by definition. In some cases, it was necessary to use a transformation of the explanatory variable to obtain a positive result in the normality test[3].

Table 3. Normality tests

Variables	Kolmogorow–Smirnov			Shapiro–Wilk		
	Statistics	df	*p*-value	Statistics	df	*p*-value
lg*X*1	0.048	107	0.200*	0.991	107	0.715
*X*2	0.095	107	0.018	0.972	107	0.014
*X*3	0.139	107	0.000	0.954	107	0.001
lg*X*4	0.064	107	0.200*	0.991	107	0.662
lg*X*5	0.074	107	0.185	0.988	107	0.484

Source: SPSS 17.0 processor.

From the group of quantitative explanatory variables, only variables lg*X*1[4], *X*2, lg*X*4 and lg*X*5 were used in the discriminant analysis (the significance level for the normality tests was 0.01). The selection of significant explanatory variables was based on Wilks' lambda statistic and, as a result of this procedure, a discriminant model with only one explanatory variable (lg*X*5) was obtained.

Table 4. Functions of the centers of gravity of groups

Y	Function
	1
0	–0,044
1	0,345

Source: SPSS 17.0 processor.

The results obtained (see Table 4) allow us to determine the separating value on the discrimination scale. Applying the values from this table into Eq. (3), a separating value of 0.15 was obtained. If the calculated value of the discriminant function is

[3]Shapiro–Wilk and Kolmogorow–Smirnov normality tests.
[4]lg*X*1 means the decimal logarithm of *X*1, similarly for lg*X*4, lg*X*5.

greater than the separating value, then the company should win the tender. In contrast, when the calculated value of the discriminant function is smaller, then the contract is not likely to be granted.

Table 5. Classification results. Detailed statistics for observation

Y			Predicted belonging to group		Total
			0	1	
Original	Number of observations	0	85	6	91
		1	8	3	11
		ungrouped observations	12	0	12
	%	0	93.4	6.6	100
		1	72.7	27.3	100
		ungrouped observations	100	0	100
86.3% of originally grouped observations were correctly classified					

Source: SPSS 17.0 processor.

In the course of the calculations, 8 out of 11 tenders classified as group 1 (tender won) were wrongly classified as group 0 (tender lost), which represents 72.7% of the total. Only 27,3% of these observations were classified correctly within the group of winning tenders. In the case of tenders from the group 0 (not won) 85 were correctly classified, which represents 93.4% of the number of tenders not won. Overall, 86.3% of the originally grouped observations were classified correctly. The results should therefore be regarded as not satisfactory enough, however, given the necessary assumptions of discriminant analysis and results of the stepwise selection of explanatory variables, the model is the best model built.

The final form of the discriminatory equation obtained is as follows:

$$D = -5.32 + 1.20 \lg X5 \tag{4}$$

By analysing this equation, it is easy to make the following conclusion: when the highest offer presented in a tender increases, the chance of winning the tender from the participant's point of view also increases.

4.2.3. Cluster analysis

The third method used in the research was cluster analysis. From the forms outlined in this paper (Sect. 4.1.3), it was decided to apply the method of the Nearest Neighbor Clustering. The disadvantages of other methods such as a lack of capacity to analyze discrete data (in the case of the *k*-means method) or the possibility of analyzing only one type of variable (quantitative or qualitative) using a hierarchical method undoubtedly had an impact on this decision. An important advantage of the method used is the relative ease of acquiring estimates of the probabilities of winning a tender. The spatial distribution of observations is shown in Fig. 1.

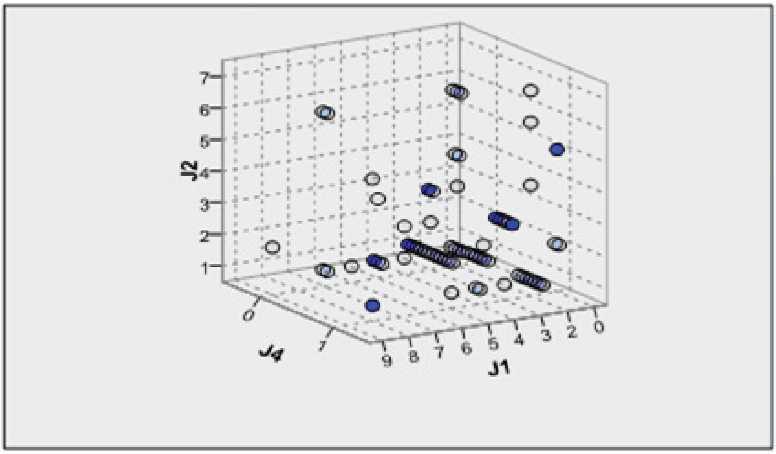

Fig. 1. Spatial distribution of observations. Cluster analysis. Source: SPSS 17.0 processor

The *Manhattan metric* (as an alternative to the Euclidean distance) was used to calculate the distance between points in the graphical representation of the variables. Moreover, the biasing of the variables according to their importance in the division was also performed. The probability distribution for $Y = 0$ (loss), obtained by cluster analysis is shown in Fig. 2.

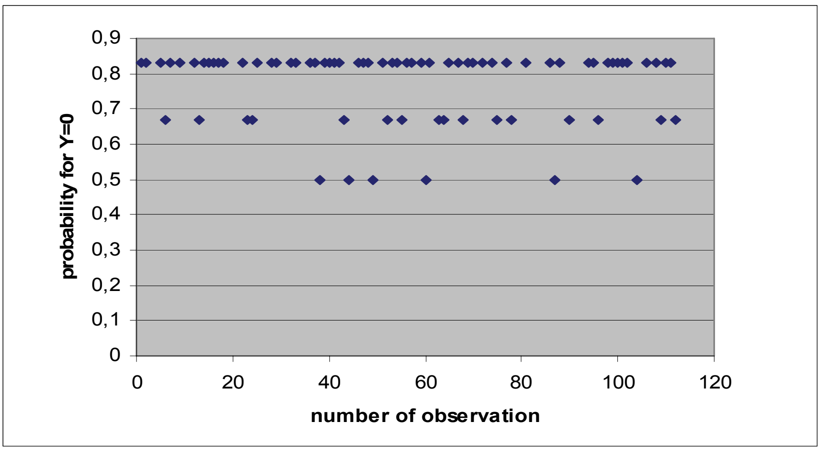

Fig. 2. Probability distribution for $Y = 0$ cluster analysis.
Source: own elaboration based on SPSS 17.0 computations

It should be stressed that in the course of the calculations, not all the tenders were classified. Tenders with missing data were excluded. Under this assumption, the results given in Table 6 were obtained:

Table 6. Indicators of fit. Cluster analysis

Observed		Expected		
		Y		Percentage of observations correctly classification [%]
		0	1	
Y	0	50	18	73.53
	1	4	4	50.00
Total percentage		71.05%		

Source: own elaboration, based on SPSS 17 computations.

This result was obtained using a split point of 0.7. This means that if the estimated probability of the event "loss" exceeds 0.7, the model classifies a case as "0" (loss). Changing the split point was necessary due to the fact that using the classical value of 0.5, the model did not classify any case as winning. We may conclude that the method of cluster analysis (assuming the most intuitive tool), in this particular case, gives the lowest quality predictions out of all the methods used, measured by the total goodness of fit measure. Furthermore, the measure of the impact of the attributes examined on the chances of winning the tender cannot be obtained by using this method. This fact also limits the usefulness of this technique for the purpose of accomplishing the goals of the research conducted.

5. Conclusions

The purpose of the paper was to identify factors influencing the probability of winning a tender and to evaluate the strength of their impact from the perspective of both the bidder and procurer. The research was conducted using three quantitative methods: *binary logistic regression*, *discriminant analysis* and *cluster analysis* conducted on a sample comprised of public tenders in which the company studied performed the role of a bidder. All the calculations presented in the paper were obtained using the SPSS 17.0 statistical package [4]. The research resulted in finding both: factors affecting the probability of winning and estimates of the probability of achieving it. The objectives of the study have therefore been achieved and results obtained allow a fuller understanding of both: the mechanisms which govern public tendering procedures and the detection of associated patterns. This allows establishing of norms, helpful even at the stage of taking the decision of whether to apply for contract, so the chances of victory may be significantly increased. Furthermore, the results obtained may constitute a contribution to further research, related to the possibility of elaborating a tendering strategy for the company studied. For example, the ability to estimate changes in the probability of winning when the value of an explanatory variable changes by one unit (as in the case of logistic regression) allows us to assess the impact of a potential decision to change the level of a bid – and thus the impact on the chances of winning a tender. Taking into account the results obtained, the next stage of work could therefore be a scenario analysis for future tenders, which could enable the assessment of chances of winning a tender according to the different strategies used by competitors. But for now, this action significantly goes beyond the objectives of this study.

References

[1] ACZEL A.D., *Complete Business Statistics*, Richard D. Irwin, Inc., Sydney 1993.
[2] AGRESTI A., *Categorical Data Analysis*, Wiley, New York 2002.
[3] AMEMIYA T., *Advanced Econometrics*, Harvard University Press, 1985.
[4] Documentation of the SPSS 17.0 statistical package.
[5] EVERITT B. S., LANDAU S., LEESE M., *Cluster analysis*, Oxford University Press, London 2001.
[6] GUTIERREZ P.A., SALCEDO-SANZ S., SEGOVIA-VARGAS M.J., SANCHIS A., PORTILLA-FIGUERAS J.A., et al., *Generalized Logistic Regression Models Using Neural Network Basis Functions Applied to the Detection of Banking Crises*, *Lecture Notes in Computer Science*, Trends in Applied Intelligent Systems, 2010, 6098, 1–10.
[7] HOSMER D.W., LAMESHOW S., *Applied Logistic Regression*, 2nd Ed., Wiley, Chichester 2000.
[8] JAIN A.K., MURTY M.N., FLYNN P.J., *Data Clustering: A Review*, ACM Computing Surveys, 31, 3, 1999, 264–323.
[9] MADDALA G.S., *Introduction to Econometrics,* 3rd Ed., Wiley, Chichester 2001.
[10] MALARA Z., *Przedsiębiorstwo w globalnej gospodarce*, PWN, Warszawa 2006.
[11] MALAWSKI M., *O rozwiązaniu średnich wypłat zagadnienia przetargowego Nasha*, Prace Instytutu Podstaw Informatyki Polskiej Akademii Nauk, 2001 (935), 1–10.
[12] MINASOWICZ A., KOSTRZEWA M., *Analiza wielokryterialnego wyboru najkorzystniejszej oferty w postępowaniach przetargowych z wykorzystaniem systemu eksperckiego opartego na przesłankach rozmytych*, http://suw.biblos.pk.edu.pl/downloadResource&mId=154505 state on: 25.02.2012.
[13] PRAHALAD C.K., HAMEL G., *Competing for the future*, Harvard Business Press, 1994.
[14] *Public Procurement Law*, Act of 29 January 2004 (together with subsequent amendments).