This is the author’s version of a work that was submitted/accepted for publication in the following source:


This file was downloaded from: https://www.escholar.manchester.ac.uk

© Copyright 2002 Taylor and Francis

Reproduced in accordance with the copyright policy of the publisher. 
**Notice:** Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source.
DATA MODELLING AND THE APPLICATION OF A NEURAL NETWORK APPROACH TO THE PREDICTION OF TOTAL CONSTRUCTION COSTS

Margaret W Emsley, Lecturer in the Manchester Centre for Civil and Construction Engineering, UMIST

David J Lowe, Lecturer in the Manchester Centre for Civil and Construction Engineering, UMIST

A Roy Duff, Senior Visiting Lecturer in the Manchester Centre for Civil and Construction Engineering, UMIST

Anthony Harding, former Research Assistant in the Manchester Centre for Civil and Construction Engineering, UMIST

Adam Hickson, former Research Assistant in the Manchester Centre for Civil and Construction Engineering, UMIST

Person responsible for correspondence:

Margaret Emsley,
Manchester Centre for Civil and Construction Engineering,
UMIST,
P. O. Box 88,
Manchester,
M60 1QD.
Tel: 0161 200 4234
Email: margaret.emsley@umist.ac.uk
DATA MODELLING AND THE APPLICATION OF A NEURAL NETWORK APPROACH TO THE PREDICTION OF TOTAL CONSTRUCTION COSTS

ABSTRACT

This paper describes the development of neural network cost models, which have been developed using data collected from nearly 300 building projects. Data were collected from predominantly primary sources using real life data contained in project files, with some data obtained from the Building Cost Information Service, supplemented with further information, and some from a questionnaire distributed nation-wide. The data collected included final account sums and, so that the model could evaluate the total cost to the client, clients’ external and internal costs, in addition to construction costs. Models based on linear regression techniques have been used as a benchmark for evaluation of the neural network models. The results showed that the major benefit of the neural network approach was the ability of neural networks to model the non-linearity in the data. The ‘best’ model obtained so far gives a mean absolute percentage error (MAPE) of 16.6%, which includes a percentage (unknown) for client changes. This compares favourably with traditional estimating where values of MAPE between 20.8% and 27.9% have been reported. However, it is anticipated that further analyses will result in the development of even more reliable models.

KEYWORDS:
INTRODUCTION

The importance of models to estimate the cost of buildings has been highlighted by Ferry et al. (1999). Newton (1991) reviewed over sixty cost models and classified the techniques used to develop each model under eight headings including regression techniques. However, in either case, no mention was made of the application of neural networks. Elhag and Boussabaine (1998) developed neural network models to predict the tender price of school buildings and later (Elhag and Boussabaine 1999a) developed two models to predict the tender price of office buildings, using linear regression and neural network techniques. They found that both techniques produced models which were able to map the underlying relationship between the cost factors and the tender price, but as the sample size was small (30 and 36 projects respectively), concluded that more projects were required for meaningful conclusions to be drawn.

This paper describes the development of neural network models of total construction project cost based on recent historical project data. The initial impetus for the research was the paucity of data available that can provide reliable information about the relative costs of using different procurement routes. However, in attempting to develop a model to address this strategic decision, it immediately became apparent that this variable cannot be isolated from the many other cost significant variables in a building project (Harding et al., 1999a) and a model which incorporates all the cost significant variables, the values of which are known at the early stage of the project, is required.
This work has been carried out in two stages:

- an initial pilot study, where potentially cost significant variables were identified, availability of data investigated and strategies for data collection established. In addition appropriate modelling strategies were examined and preliminary testing of these methods was carried out, using a relatively small number (46) of data sets (Duff et al., 1998).
- a full scale study, using data from nearly 300 projects, and hence addressing one of the deficiencies in the model developed by Elhag and Boussabaine (1999a), in which more sophisticated models were developed.

The data requirements and the data collection processes are both described and ways in which both the input variables (such as frame type) and output variables (cost) may be best represented in the model are discussed. For the purposes of comparison, linear regression models have also been developed and the results obtained are given before the development of the neural network modelling process is explained. The first two sets of models to be developed use first the 5 and then the 9 most significant variables identified by the linear regression modelling process and the final set incorporates all the variables.

**DATA REQUIREMENTS**

The model variables may be divided into two categories; input variables and output variables. Initially, 43 input variables were identified and this was subsequently reduced
to 41, as two variables were eliminated (sanitary installations and disposal installations) as almost no variation in their definition was found among the project data collected. Input variables were further categorised as project strategic, site related or design related, and are tabulated in Fig. 1.

The identification of potentially cost significant variables was achieved through a thorough literature search of over 60 publications, supplemented by discussion with the professional collaborators (see Acknowledgements). In addition, with the exception of quality of building, all other input variables are encapsulated in the cost analyses published by the Building Cost Information Service (BCIS). This is less than the 67 variables identified by Elhag and Boussabaine (1999b), but it should be noted that that list also includes factors affecting construction project duration and factors classified as contractor attributes, which would not be known at the stage when this model is intended to be used.

A criticism of previous cost models is that they used only the tender price to evaluate cost, whereas, in reality, the cost to the client of a building contract is the final contract sum. This is very rarely the same as the tender price and Corbett and Rowley (1999) suggested that the final account sum should be made available to cost planners, whereas the BCIS, for example, only provide the tender price. The model described here has been developed using final account figures as the output variable. In addition, the whole cost to the client includes not only the final contract sum but also the professional fees and whatever resources the client has provided to the project. Models were developed
separately for construction costs and client costs, but they may be summated to give the total project cost to the client.

The variables of time and geographic location were accommodated through the application of the BCIS cost indices to bring all projects to a common location and base date. The costs predicted, using the model, are then adjusted by the appropriate indices for the time and location in question.

Where projects included external works, demolition, fittings or specialist services, their associated costs were removed from the final account figure, and appropriate proportions were removed from the contract preliminaries and clients’ costs (Harding et al., 2000a). This was done because these costs are subject to wide variation, largely independent of the main variables defining the building. For example, for the projects where data were collected, the cost of external works varied from 1% to 30% of the total contract sum. Such variation makes these features impossible to model accurately and they are more reliably estimated independently.

**DATA COLLECTION**

In total, the data collection programme resulted in the collection of 288 full data-sets from predominantly primary sources, supplemented by some secondary data.
Primary sources

The professional collaborators provided a great deal of the data required and contact was also established with organizations, primarily Quantity Surveying and Project Management practices, that were willing to provide data. A data pro-forma was developed to assist both the researchers and, more importantly, those collaborators willing to carry out the data retrieval themselves.

This method of collection provided the great majority of building cost analyses. A total of 39 offices were visited from 20 different organisations across the United Kingdom.

Secondary sources

The BCIS publishes cost analyses for construction projects, which fulfilled the data requirements, except for the following information:

- final account
- actual duration
- quality of building
- clients’ external costs
- clients’ internal costs
In order to obtain this additional information, a questionnaire was administered and sent to BCIS subscribers, yielding 29 sets of data.

In addition to these questionnaires, data were obtained from a much more extensive mail-shot administered to 1239 practising quantity surveyors, all of whom had been canvassed by telephone, but this yielded only 6 additional projects.

**DATA REPRESENTATION**

Each of the variables has been analysed in order to determine the best way of representing that variable in the modelling process. The way these variables are represented fell into four distinct groups.

The first of these groups comprised variables which are real numbers, for example, *duration* and *no. lifts*. Where the range of these variables differed by more than one order of magnitude it was more appropriate to use the logarithm of that value, to ensure that the range of values was more evenly distributed.

The remaining variables are categorical variables that represent one of a choice of categories. As a general rule it is best that a single input is used for a variable only when that variable has some meaning as a single variable (Tarassenko, 1998). That is to say, if the value of the variable increases then it must represent an increase of some factor
which influences the outcome of the model. For some variables, obtaining such an order was simple. For example, with site access there is clearly an order between unrestricted, restricted and highly restricted, inasmuch as an increase in the restriction to access will be expected to cause an increase in cost. Therefore this variable can be represented by a single input.

There were a great many more variables for which no such order was immediately apparent, for example, internal wall finishes, where the variable represents the cost of different material combinations which will make up the finish. The value of the input was set to be the standard cost per m² of each finish, which provides an order proportional to how much each finish is expected to impinge upon the final building cost.

For some categorical variables a consistent order could not be identified, because the actual differences in cost between the possible choices are uncertain. For example, for frame type (where the choice was insitu, masonry, precast, steel or timber), there is a lack of consensus on the comparative costs and it is impossible to ascertain a consistent ordering, in terms of cost, which would apply in all circumstances. Therefore, a binary input coding (yes/no) was applied to each possible choice, thus treating each such categorical variable as a series of binary variables.
COMPARISON OF MODELLING TECHNIQUES

A comparison of linear regression analysis and neural networks has been made elsewhere (Harding et al., 1999b) and some preliminary analyses carried out (Harding et al., 2000b). However, in this situation, the main advantage that neural networks offer is their ability to capture the non-linearity that will inevitably exist between variables. Non-linear regression techniques can also be used to account for non-linearity, but have the disadvantage that the user must have detailed knowledge about the appropriate non-linear relationship between the predictor variables and the mean values of the observations (Christensen, 1996). However, when applying neural networks, these relationships are determined implicitly by the model and are not therefore required to be specified.

REPRESENTATION OF COST (OUTPUT VARIABLE)

Previously, cost models have often used the raw cost as the dependent variable. However, there are a number of assumptions implicit in this choice of variable. Firstly, it is assumed that the standard deviation of error remains constant. That is to say, the cost of a small project can vary by the same monetary amount as a large project. This is highly unlikely to be the case. Further, regression model fitting minimises the squares of the errors, so models developed using this technique will be inherently biased towards minimising the errors for very large projects, where the errors are greatest. It is
therefore unlikely to be a good predictor of the cost of smaller projects. Given that the costs of projects in the data collected vary between £36,000 and £15,800,000, the influence of errors on the cost of the largest projects is several orders of magnitude more than those of the smallest projects, so the effect will be pronounced.

The second inherent assumption which is questioned is that the effects of any variable are best expressed as a fixed cost change. If, for example, the specification of the floor finishes changed to one of a higher cost, the cost of the building would be expected to rise. However, the cost of a small building would not be expected to rise by the same amount as the cost of a very large one, but as a proportion of the building size or cost.

These criticisms raise serious questions as to the meaningfulness of models produced by using raw cost as the predictor for a linear regression model. Therefore, three other possible models were tested.

**Log of building cost**

In order to address the problem of the large cost differences, a common solution is to model the log of the cost. This assumes that the log of cost is normally distributed and that a change in any variable within the model will cause a proportionate change in cost. That distribution which corresponds to a normal distribution of the log of cost is one whose mean is the project cost, and whose standard deviation is a fixed proportion of project cost. When the normal distribution is converted to raw cost for any project, it is
a positively skewed distribution, such that the peak of the probability density function is less than the mean. The skewed nature of this function could, it might be argued, be a better representation of the possible variation in project cost than a true (unskewed) probability distribution as, generally, there is more scope for the project costs to be much higher than expected rather than much lower.

**Cost/m²**

The cost per m² is the cost predictor most used by quantity surveyors, as it provides a measure of cost that is essentially independent of building size. If this value were to be used in a regression model, then it assumes that any variation in project cost is proportionate to the size of the building, rather than the cost. This may seem to be an unrealistic solution, as projects which are of a higher specification (and hence a higher cost/m²) might be expected to show correspondingly higher variations in cost. However, it has the added advantage of removing the understood linear relationship between GIFA and project cost from the model. This should allow the modelling to focus on other, less understood, influences on project cost.
Log of cost/m$^2$

The log of cost/m$^2$ makes the same assumptions as the log of cost in that variations in project cost are proportionate to the expected cost. However, it also provides a variable which is devoid of the linear relationship between cost and size, in the same way as the cost/m$^2$ output. While this makes little difference to regression models, it could be useful in the neural network modelling, if the correlation between cost and GIFA creates difficulties for those models which use the log of cost in learning the relationship between cost and other variables, as this could stop the neural network learning only this relationship to the detriment of others.

FACTOR ANALYSIS

Factor analysis was used to establish the underlying dimensions of the input variables, confirming that all items should be retained. Principal component extraction with varimax rotation was also used and although this indicated that the true number of underlying dimensions lay between eight and nine factors, regression models created using the factor scores resulted in $R^2$ values less than those generated using the original input variables. A decision, therefore, was reached to proceed using the original input variables.
LINEAR REGRESSION ANALYSIS

The aims of the regression analysis were, firstly, to develop a robust regression cost model that would make a useful benchmark against which neural network models could be measured. Secondly, it was desirable to identify those variables that demonstrated a strong linear relationship with the cost, to assist in the management of neural network training. The software which was used was SPSS 7.5 for Windows.

In order to create a predictive regression model, two methods were attempted – forwards and backwards modelling. Models to predict cost/m², log of cost/m² and log of cost were generated for each method. The number of statistically significant variables in each model varied between 8, in the forward stepwise log of cost model, and 14, in both the log of cost and log of cost/m² backward models. Throughout the models a total of 19 different variables was used. A summary of the results is given in Fig. 2.

There were 5 variables that appeared in all six models: GIFA, function, duration, mechanical installations and piling. This suggests that these are the key linear cost drivers in the data. A further 4 variables appeared in five models: internal wall finishes, frame type, site access and protective installations.

The log of cost backwards models outperform the other models by most of the percentage error measures. However, the differences between all the models are small. One of the reasons the log of cost and log of cost/m² models perform so well is that they
incorporate more variables of whose inclusion in the model we can be confident, although they may not necessarily be the most useful variables in terms of which to model, because their values may be uncertain at the early estimating stage.

As the models exhibit similar performances, it might be more appropriate to consider the spread of error. Neural networks minimise error using the least squares approach. As this can be sensitive to non-uniformity of standard deviation and non-normality of error, the spread and normality of error was assessed for each model by considering the scatter plots of error against the value of the independent variable, all of which displayed the same tendency for the models to overestimate the cost of cheaper projects, and underestimate the cost of more expensive projects. The fact that as much as 30% of the error appeared to arise from this suggests that some key drivers of building cost were not being adequately represented. This either arises from their non-inclusion or, more likely, from non-linearities in the data, in which case it can be expected that a neural network will perform better.

**NEURAL NETWORKS**

Three sets of models have been developed, using:

- the five variables which were incorporated in all six of the linear regression models;
- the nine variables found in five of the six regression models; and
- all the variables for which data have been collected.

In addition, an optimum combination of variables was used, determined by using a combination of the forwards and backwards stepwise feature selection algorithm of a generalised regression neural network (GRNN) and a Genetic Algorithm (GA) global optimisation search technique. As a GA is theoretically better able to come up with an optimum combination of variables, and is a randomised process, this was repeated 4 times and the forwards and backwards algorithms were executed once each. By excluding all the variables that do not appear in the feature selection results, the model was reduced to a 25 variable model. However, as the performance of this model was poorer on both measures of performance than the all variable model, suggesting that one or more significant variables have been omitted from the analysis, this approach was, therefore, abandoned.

The software which was used was Trajan Neural Network Simulator Release 4.0E.

In order to assess the best approach to the modelling, a number of different networks were tried: three and four layer multi-layer perceptrons (MLPs), radial basis functions (RBFs) and generalised regression neural networks (GRNNs). Of these alternatives, three layer MLP networks (with one hidden layer) offered the best performance, in terms of the associated values of $R^2$ and mean absolute percentage error (MAPE).
To train the network, backpropagation, conjugate gradient descent, Levenberg Marquardt, quick propagation, delta-bar-delta and quasi-newton training algorithms were evaluated. Of these, the ordinary backpropagation algorithm was the most efficient, as it found solutions which were at least as good as the solutions found by other algorithms, and usually more quickly.

Once the structure and training method to be used had been established, three different function types were used as activation functions in the hidden and output layer: linear, logistic and hyperbolic. The hyperbolic function is approximately linear at 0, requiring the variables to be normalised and the weights set to very small values. This means that early in the training linear behaviour is assumed, while later the curved areas of the hyperbolic function are used to model the function. It was found that this technique yielded slightly better networks more quickly. The optimum number of nodes in the hidden layer was also investigated.

**Five variable model**

The results of the best networks for the five variable model gave $R^2$ values that are similar to, but not as good as the regression analyses. However, the regression models are tested on the same data set used to create the model, while the neural network models are tested using a small (only 45 cases) independent test data set, which could cause the accuracy of the network to be misrepresented. The verification set is similarly
prone to being unrepresentative. The process of training might, therefore, be terminated prematurely, when the model begins to lose its fit with the verification set, but before (or after) the model ceases to fit the real population. One solution to this problem is to use a “voting system”, involving the creation of a number of models, each of which uses different training, verification and test sets. The value of the output is then taken as the average of these models. Bias will exist within all the individual models, but provided all the projects in the data set are represented, other models will be biased in the opposite direction, producing an output which is closer to the mean of the data set, rather than just a small subset.

A voting system was trained for the cost/m$^2$ and log of cost/m$^2$ models. The values of R$^2$ were similar to those found when the voting system was not adopted, although the actual values for the individual models did vary significantly. This can be assessed by comparing the average R$^2$ values for each combination of training, verification and test sets using analysis of variance. The value of F obtained was 8.63, which is very highly significant, showing that the differences between the R$^2$ values have not arisen by chance, but from the fact that the test sets are not representative.

Examining the distribution of R$^2$ values, the best overall mean value is obtained by a network architecture containing 2 nodes in the hidden layer and, overall, the best values of both the mean and R$^2$ values found are for architectures with 8 nodes in the hidden layer or less. However, these values have been obtained with only 20 models, and the differences are small.
While determining the best architecture is difficult, because there is little significant difference between their performances, there are significant differences between the performances of different configurations of training, verification and testing sets. This validates the approach of the voting system. However, while it is possible to average the results of the entire set, it is not possible to obtain accurate estimates of the true error using the test set, as each project has not been included in the test set enough times.

In order to permit more effective representation of each project, the number of networks used in the voting system was increased to 50, but there was no improvement over the best regression analysis. Nevertheless, the cost/m² neural network model did improve in respect of its linear counterpart. This suggests that the model has encapsulated aspects of the relationships between the variables which the regression analysis has failed to do, suggesting that the neural network model is capable of improving upon the linear models where non-linear relationships do exist.

**Nine variable model**

Voting systems were trained in a similar way. These networks were created by 20 attempts at training for each of 15 different network architectures. The architectures were all three layer MLP networks with between 1 and 15 nodes in the hidden layer. However, the results show that with these models the neural networks are failing to model even some of the linear relationships. It would appear that there is simply not
enough data to model this problem effectively. This is partly because only a small proportion of the relationships are non-linear and most of the model can be explained using a linear model. Also, there are not sufficient data to permit good generalisation of networks with an architecture which is adequate to model the non-linearities. Generally, it is more difficult for neural networks to learn relationships which only explain a small part of the model’s performance. This is because the variation is not large enough in proportion to the amount of unique error in the problem (error which cannot be explained by the variables).

**All variables model**

This model was developed as a voting system trained to predict the cost/m$^2$. As the model involves a large number of variables, it was felt that the potential influence of bias would be larger than for the earlier models and the size of the voting system was increased to 100 networks. The model demonstrated significant improvement upon the equivalent regression model. From the results of the earlier networks, it is known that a network of this size will find it very hard to model some of the more subtle linear relationships. Thus, the increase in performance observed implies that there are significant non-linear relationships which the network is modelling.

The fact that earlier neural network models did not appear to find some of the non-linear relationships has implications for the accuracy of this model. The number of variables has gone up, making it harder for the neural network to model some of these weaker
relationships. Therefore, the model is almost certainly not modelling some of these relationships, which has a detrimental effect on its accuracy. In addition to this, it is also likely that the modelling of the non-linear relationships also lacks some accuracy for the same reasons. Therefore, although it already outperforms the regression model, it is clear that more data would produce an improvement in the model’s performance.

Results

The results obtained applying the methods outlined above are tabulated in Fig 3, which also shows (in parentheses), where appropriate, the results obtained for regression models developed using the same variables, so that a direct comparison between models obtained by both techniques can be made.

The best model obtained was a neural network model using all 41 variables and a voting system which used 100 networks; this gave an $R^2$ value of 0.789 and a MAPE of 16.6%, which is an improvement upon results obtained from regression analyses generally and the ‘best’ regression model specifically ($R^2$ of 0.661 and MAPE of 19.3% for backward log of cost model). Where linear regression and neural network models have been developed using the same variables, neural networks always outperform their regression counterparts and the best linear regression models always outperform those developed for direct comparison with neural networks.
These results compare favourably with past research that has shown that traditional methods of cost estimation are less accurate, as evidenced by reported values of MAPE between 20.8% (Skitmore et al., 1990) and 27.9% (Lowe, 1996).

CONCLUSIONS

The two approaches to modelling, predicting the cost/m$^2$ and the log of cost/m$^2$, yield similar performances but with subtle differences. Predicting the cost/m$^2$ tends to produce a model which has a higher R$^2$ value in cost/m$^2$ terms. However, the log model yields lower values of MAPE. This is a function of the fact that the log model explicitly minimises proportional differences, whereas the untransformed cost/m$^2$ model minimises the square of error on the cost/m$^2$. Thus the model selected should be based upon whether the user wants accuracy in proportional or cost/m$^2$ terms.

The overall results have significant implications for the assumptions on which the research is based. It was assumed that there were definite benefits to using a neural network approach, as it was capable of modelling the non-linear relationships in the data. While the models presented above may not be much more accurate than current cost estimation practice, they do show that there are non-linearities in the data and that neural networks are capable of modelling them. It is believed that further analyses will lead to better models and that there is evidence that the inclusion of more data would yield significant improvements in the accuracy that could be achieved by neural network
modelling; these analyses are currently being undertaken and the results will be reported subsequently.

ACKNOWLEDGEMENTS

The authors acknowledge, with great gratitude, the support of:

- the Engineering and Physical Science Research Council, who funded the project;
- our professional collaborators, E C Harris, Symonds and Tweeds;
- the BCIS for their contribution during the full scale study; and
- the many other Quantity Surveying and Project Management practices who provided data.

REFERENCES


Fig 1  Classification of input variables

<table>
<thead>
<tr>
<th>Project strategic variables:</th>
<th>Procurement strategy</th>
<th>Quality of building</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contract form</td>
<td>Purpose</td>
<td>Tendering strategy</td>
</tr>
<tr>
<td>Duration</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Site related variables:                      |                      |                     |
| Site access                                  | Type of location      |                     |
| Topography                                   | Type of site          |                     |

| Design related variables:                    |                      |                     |
| Air conditioning                             | Internal doors        | Roof profile        |
| Ceiling finishes                             | Internal walls        | Shape complexity    |
| Electrical installations                     | Internal wall finishes| Special installations|
| Envelope                                     | No. lifts             | Stair types         |
| External doors                               | No. storeys above ground| Substructure       |
| External walls                               | No. storeys below ground| Structural units   |
| Floor finishes                               | Mechanical installations| Upper floors  |
| Frame type                                   | Piling                | Wall to floor ratio|
| Function                                     | Protective installations| Windows          |
| GIFA                                         | Roof construction     |                     |
| Height                                       | Roof finishes         |                     |

Fig. 2  Linear regression model results

<table>
<thead>
<tr>
<th></th>
<th>Cost/m²</th>
<th>Log Cost/m²</th>
<th>Log Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R²</td>
<td>MAPE</td>
<td>R²</td>
</tr>
<tr>
<td>Forward</td>
<td>0.668</td>
<td>20.8%</td>
<td>0.648</td>
</tr>
<tr>
<td>Backward</td>
<td>0.666</td>
<td>21.7%</td>
<td>0.661</td>
</tr>
</tbody>
</table>

Fig. 3 Neural network model results

<table>
<thead>
<tr>
<th></th>
<th>Cost/m²</th>
<th>Log Cost/m²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R²</td>
<td>MAPE</td>
</tr>
<tr>
<td>5 variable model</td>
<td>0.564</td>
<td>20.3%</td>
</tr>
<tr>
<td>(without voting system)</td>
<td>(0.556)</td>
<td>(23.6%)</td>
</tr>
<tr>
<td>5 variable model</td>
<td>0.636</td>
<td>21.9%</td>
</tr>
<tr>
<td>(with voting system)</td>
<td>(0.556)</td>
<td>(23.6%)</td>
</tr>
<tr>
<td>9 variable model</td>
<td>0.688</td>
<td>20.1%</td>
</tr>
<tr>
<td>(with voting system)</td>
<td>(0.617)</td>
<td>(21.5%)</td>
</tr>
<tr>
<td>All variables model</td>
<td>0.789</td>
<td>16.6%</td>
</tr>
<tr>
<td>(with voting system)</td>
<td>(0.683)</td>
<td>(27.7%)</td>
</tr>
<tr>
<td>Reduced variables model</td>
<td>0.665</td>
<td>22.7%</td>
</tr>
</tbody>
</table>