Multiple regression models for electronic product success prediction

Lo, Frank Cheong Wah; Foo, Say Wei; Bauly, John A.

2000

Lo, F.C.W., Foo, S.W. & Bauly, J.A. (2000). Multiple regression models for electronic product success prediction. IEEE International Conference on Management of Innovation and Technology 2000: (pp. 419-422). Singapore: Singapore Polytechnic.

<https://hdl.handle.net/10356/91312>

<https://doi.org/10.1109/ICMIT.2000.917374>

© 2000 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder. <http://www.ieee.org/portal/site> This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.

MULTIPLE REGRESSION MODELS FOR ELECTRONIC PRODUCT SUCCESS PREDICTION

Frank Cheong-Wah Lo
Singapore Polytechnic
Email: F.LoCW@sp.edu.sg

Say-Wei Foo
National University of Singapore
Email: EleFoosw@nus.edu.sg

John A Bauly
National University of Singapore
Email: EngJab@nus.edu.sg

ABSTRACT

As the cost of failure in new product development is very high, product developers are looking for good product success/failure prediction models. The general direction of search is towards Knowledge Based Systems (KBS) that incorporate the wisdom of experienced developers and extracts from data of past projects. In this paper, results of investigation using multiple regression models are reported. It is found that 90% accuracy may be achieved in success/failure prediction of electronic product development using a multiple regression model based on six critical factors.

Keywords: Multiple regression model, knowledge-based system, product development, success/failure prediction.

1. INTRODUCTION

The rapid advance of technology has sped up the rate of product obsolescence. Companies must continue to produce new products that take advantages of the new technology. However, the risk in new product development is high. Millions of dollars may be lost in a short time for a failed product. On the other hand, successful new products may bring in huge fortune. Faced with this dilemma, it is desirable to develop a model that enables accurate prediction of the outcome of a new product concept before heavy expenditures are incurred.

Research in this area tends towards the development of Knowledge-Based System (KBS) which incorporates the wisdom and knowledge of experts in new product development [1]. The KBS is usually formulated and tested using data of past projects. It attempts to replicate the intelligent activities of specific human experts. Experts of interest in this project may include managers in different functional

areas who need to make go/kill decision during different phases of development process. [2].

Major advances have been made in the application of KBS in engineering and other areas. The application of KBS becomes more attractive when computers become faster and cheaper. A KBS may incorporate techniques such as neural network [3], fuzzy logic [4] or plain old statistical analysis.

In a separate study [5], various KBS models on product success/failure prediction based on regression analysis, fuzzy logic and neural network were investigated. The models were tested for their accuracy, reliability, applications and limitations using data collected from local electronics industry. It was found that multiple regression model is a promising method. A multiple regression model based on a small set of critical variables is able to provide a high degree of accuracy in success/failure prediction.

In this paper, details of the development, formulation and testing of the multiple regression models for product success/failure prediction are presented.

In Section 2, a multiple regression model with ten input variables is presented. A multiple regression model based on a reduced set of input variables is described in Section 3. This is followed by concluding remarks in Section 4.

2. MULTIPLE REGRESSION MODELS

2.1 Choice of Input Variables

In a prior research [6], 10 product success factors were identified as critical factors that determine the success or failure of new product development. For ease of description, these ten factors are named F1 to F10. These 10 factors are: F1: definition of concept; F2: specifications of user requirements; F3: degree of

fit of product with company; F4:technical expertise; F5:marketing expertise; F6:management planning and control; F7:commitment of top management; F8:motivation of personnel; F9:availability of facilities and other resources and F10:quality of facilities and other resources.

The first regression model proposed makes use of all these ten factors as input variables.

2.2 Collection of Data

For the development and testing of the multiple regression models, a set of questions relating to the 10 factors was prepared. Major companies in Singapore that are involved in electronic product development were identified. Interviews were conducted with the willing companies on their past successful and/or failed products. There were not many such companies in Singapore and only 32 sets of data were obtained. Fortunately, this set of data includes all spectra of electronic products, from the very successful to the complete failure.

These 32 sets of data are divided into two groups. Group 1 consists of 22 sets of the data of different nature and these are used primarily in the formulation of the model. The second group, Group 2, includes the remaining 10 sets of data and are used for testing the model.

In our model, profit is the sole indicator of success of the product. If the product does not contribute to the bottom line of the company, it is not considered successful even though it may be technically excellent.

The questions, grouped into 6 categories, are presented in Appendix 1.

2.3 Formulation of the Model

The mathematics of multiple regression model can be found in many books [7,8] on this topic. Software packages written for computation of the coefficients of the model are also readily available [9]. Interested readers may consult the referenced books for details.

The proposed models are assumed to be linear. The set of coefficients for the models are computed using Group 1 data as mentioned in Section 2.2. The resulting equation representing the model with 10 input variables is given by:

$$Y = -3.99 + 0.48X_1 - 0.58X_2 - 0.38X_3 + 0.12X_4 + 0.20X_5 + 1.08X_6 - 0.55X_7 + 0.43X_8 - 0.01X_9 - 0.04X_{10} \quad (1)$$

where

- Y = values of the dependent variable denoting the success of the project. A value above 2.5 is considered success.
- X1 to X10 = values of the independent variables corresponding to Factors F1 to F10.
- Analysis of variance (ANOVA) indicates that the regression relation is very significant as chance error is small.
- The residual output is random. This supports the linear regression model.

2.4 Testing and Performance of the Model

The model is tested using the set of data used in the formulation, i.e., Group 1 with 22 sets of data and also the Group 2 data reserved for testing as mentioned in Section 2.2.

The results are tabulated in Table 1.

Table 1: Performance of Multiple Regression Model with 10 input variables

Accuracy of Prediction using Group 1 data	Accuracy of Prediction using Group 2 data
91%	70%

From Table 1, it can be seen that the accuracy of prediction using Group 2 data is very low compared with that using Group 1 data. One of the possible reasons is over-fitting. With too many input variables, the model is trained to memorize the set of data and this does not leave room for dealing with new data that does not follow the so-called norm established by the training data. Another possible reason is multi-collinearity of input variables.

3. MODEL WITH REDUCED NUMBER OF INPUT VARIABLES

3.1 Motivation

It would appear logical to construe that should the model include more product success factors, the power of prediction is increased as more pieces of information are taken into account. However, this may not necessarily be the case.

Research done [10] for predicting the nation's economic future indicates that having many economic

variables as inputs did produce a multiple regression model with high regression coefficient of 0.99 to GNP based on the data formulating the model. However, when new data were tested, the predictions from the model were worse than pure guesses of the GNP.

There may be multi-collinearity in the model. In other words, there may be strong correlation among the set of input variables and hence some redundancy may be removed by reducing the number of input variables.

The best multiple regression is one that explains the relationship among the data by accounting for the largest proportion of the variation in the dependent variable, with the fewest number of independent variables.

This leads us to investigate multiple regression models with reduced number of input variables.

3.2 Selection of Input Variables

The 10 factors are grouped into 6 categories. One way to reduce the number of factors is to retain the factor with the highest correlation in each of these categories. The correlation values of the ten factors with the success of the project are given in Table 2.

Table 2: Correlation values of the ten factors

Category	Factor	Correlation Value	Most significant factor in the category
I	F1	0.41	F2
	F2	0.51	
	F3	0.22	
II	F4	0.49	F4
III	F5	0.19	F5
IV	F6	0.74	F6
V	F7	0.42	F8
	F8	0.67	
VI	F9	0.46	F10
	F10	0.64	

Thus, only Factors F2, F4, F5, F6, F8 and F10 are retained.

3.3 Multiple Regression Model with 6 Input Variables

To formulate the multiple regression model based on the 6 chosen factors, the Group 1 survey data are used. The resulting model is depicted by the following equation

$$Y = -3.31 + 0.49X_1 - 0.02X_2 - 0.14X_3 - 0.91X_4 + 0.34X_5 + 0.08X_6 \quad (2)$$

where

- Y = values of the dependent variable denoting success of the project A value above 2.5 is considered success.
- X₁ to X₆ = values of the independent variables (F2, F4, F5, F6, F8 and F10)
- Analysis of variance (ANOVA) indicates the regression relation is very significant as chance error is small.
- The residual output is random. This supports the linear regression model.

3.4 Testing and Performance of the Model

Table 3: Performance of the Multiple Regression Model with 6 input variables

Accuracy of Prediction using Group 1 data	Accuracy of Prediction using Group 2 data
91%	90%

The findings show that with this reduced number of carefully chosen factors as input variables, the accuracy of prediction based on Group 2 data improves significantly.

4. CONCLUSION

The application of multiple regression models for prediction of success/failure of electronic product development is investigated. It is shown that even with 6 key success/failure factors used as input variables for the regress model, an accuracy of 90% can be achieved in predicting the success or failure of product.

It is also shown that it is not true in general that the more input factors is used, the better is the model. The model based on 10 established input factors does not give as high an accuracy of prediction using data not employed in the formulation of the model.

REFERENCES

- [1] Robert J. Mockler, "Developing Knowledge-Based Systems Using An Expert System Shell", Macmillan, 1992.
- [2] "New Product Process at Leading Industrial Firms", Robert G. Cooper and Elko J. Kleinschmidt, Industrial Marketing Management, Vol.20, No.2, May 1991, pp137-147.

- [3] Laurene Fausett, "Fundamental of Neural Networks: Architectures, Algorithms and Applications", Prentice Hall, 1994, pp. 1-192.
- [4] Earl Cox, "The Fuzzy Systems Handbook: A Practitioner's Guide to building, using and maintaining Fuzzy Systems", AP Professional, Academic Press, 1994.
- [5] Frank Cheong-Wah Lo, "Knowledge-based System for Electronic Product Success/Failure Prediction", National University of Singapore, 2000.
- [6] Chee Meng Yap, "A System Approach to New Product Development Management with a View of Optimizing Success Rates under various Technical and Market Contingencies ", University of Pittsburgh, USA, 1992.
- [7] Amir D. Aczel. "Complete Business Statistics". 3rd Edition, Richard D. Irwin, 1996.
- [8] John Neter et al, "Applied Linear Statistical Models", Richard D. Irwin Inc., 1987.
- [9] Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. "Numerical Recipes in C: The Art of Scientific Computing". 2nd Edition New York: Cambridge University Press, 1992.
- [10] Peter Clifton, Hai Nguyen and Susan Nutt, "Market Research: Using Forecasting in Business", Butterworth Heinemann, 1992, pp.470-473.

APPENDIX 1

QUESTIONNAIRE FOR PRODUCT DEVELOPMENT STUDY

Please mark a x on the 1 to 5 scale.

I. Product Requirements and Fit

(1) The product concept was:

(2) The user's requirement were:

(3) The perceived level of fit between the technology and the company was

II. Technical Expertise

(4) The perceived technical expertise was

III. Marketing Expertise

(5) The perceived marketing expertise was

IV. Management

(6) The degree of detailed planning and control was

V. Human Resource

(7) The level of top management involvement was

(8) The motivation of project-level personnel was"

VI. Other Resources

(9) The level of available resources was

(10) The quality of resources assigned was

(11) Please indicate whether the product is profitable?

Yes, it is profitable / No, it is not profitable.

End of Questionnaire