

# Material Cost Estimation of Newspaper Printing Process by using Multiple Linear Regression Models

J.M.T.N.Jayasinghe  
Department of Statistics and Computer Science  
University of Kelaniya  
Sri Lanka  
thiliniyayi@gmail.com

R.A.B.Abeygunawardana  
Department of Statistics  
University of Colombo  
Sri Lanka  
rab\_abey@stat.cmb.ac.lk

**Abstract**— Newspaper printing industry is one of the popular industry in developing countries like Sri Lanka. Newspapers become one of the major sources of disseminating news among people. When printing a newspaper, the company has to bear a huge cost since most of the newspapers are printed on a daily basis. Labour cost, material cost, machinery cost and electricity cost are major costs involved in newspaper printing cost. Among those costs, material cost is a combination of the various types of costs. According to the industrial investigations, material cost is a combination of the amount of paper, plate and ink usage in printing, number of printed newspapers and number of pages in a newspaper. The main objective of this research is to identify the contribution of the investigated factors on the material cost. Data were taken from a popular newspaper company in Sri Lanka for a popular daily newspaper. To analyse the weekly data for the material cost the data on the selected factors of; total number of pages, amount of paper, the amount of plate and ink, total printed newspapers were taken from the period of 06<sup>th</sup> January 2013 to 05<sup>th</sup> September 2014. Multiple linear regression models were fitted to identify the factors associated with the material cost. Among them, most suitable variables for the models were identified and the model was fitted. A number of pages per newspaper, the amount of printed newspapers, paper amount (kg) and ink amount (kg) becomes the significant factors associated with the material cost of newspaper printing. The amount of the variance explained by the model is about 78%. All the assumptions in multiple linear regression models were held in this model. Therefore we can use this model to estimate the material cost for a newspaper.

**Key Words:** Multiple Linear Regression, Stepwise method, Backward Elimination, Forward Selection

## I. INTRODUCTION

Newspaper printing industry is a popular industry in developing countries like Sri Lanka because newspapers are a major source of disseminating

news among people. In newspaper printing industry printing process is the heart of the newspaper processing. Therefore, there should be a proper management of the process of printing. When printing a newspaper, the company has to bear a huge cost since most of the newspapers are printed on a daily basis. Labor cost, material cost, machinery cost and electricity cost are major costs involved in newspaper printing cost. Among those costs, material cost is a combination of the various types of costs. According to the industrial investigations, material cost is a combination of the amount of paper, plate and ink usage in printing, number of printed newspapers and number of pages in a newspaper. The main objective of this research is to identify the contribution of the investigated factors on the material cost.

## II. MATERIALS AND METHODS

Paper, Plate, and ink costs are the major material costs in newspaper production. The Study conducted by [2] Gensis Analytics (Pvt) Ltd printing cost of books are highly affected by the print amount and a number of pages. Therefore those two factors were taken as well as the material amounts. Data were taken from a popular newspaper company in Sri Lanka for a popular daily newspaper. For analysis, weekly data for material cost, the total number of pages, amount of paper, plate and ink, total printed newspapers of newspaper data were taken from 06<sup>th</sup> January 2013 to 05<sup>th</sup> September 2014.

Correlation between every variable was checked and multiple linear regression models were fitted by taking material cost as dependent variable and paper amount, plate amount, ink amount, the number of pages per newspaper, the total number of newspapers printed were taken as independent variables.

The best fitted model was selected based on the multiple linear regression models fitted. Among them, most suitable variables for the models were identified and their effect on the response variable was estimated. The necessary model diagnostic test was done to check the model adequacy.

III. RESULTS

It was found that the amounts of four colors of ink (Cyan, Magenta, Yellow and Black) were highly correlated. [1] Ismail and Yahya suggested that applying stepwise regression removes the multicollinearity. But for these data, it does not work. Therefore rather than omitting these highly correlated variables total ink amount was taken for the analysis rather than taking individual ink amount of cyan, magenta, yellow and black. Correlation matrix of the material cost with the paper amount, plate amount, the number of pages per newspaper, the number of printed newspapers and ink amount was obtained. The results were mentioned in Table I.

TABLE I. CORRELATION COEFFICIENTS OF INDEPENDENT VARIABLES WITH MATERIAL COST

Variable	Correlation	P-value
Number of pages	0.589	0.00
Plate amount	0.343	0.00
Paper amount(kg)	0.576	0.00
Ink amount(kg)	0.534	0.00
Number of printed newspapers	0.759	0.00

According to the correlation coefficient, the number of printed newspapers has highly positive correlation with the material cost. A number of plates have the least correlation with material cost and number of pages, paper amount and plate amount has average correlations. Regression models which are obtained by using the stepwise method, forward selection method and backward elimination method result by taking the probability of 0.05 to enter the variables while the probability of 0.10 to remove the variable. Among the obtained results the model with highest R square value and adjusted R square value was selected as the best model in each method. Results for each method mentioned in Table II.

All of the three methods of the stepwise method, backward elimination method, and forward selection method excluded the variable "number of plates". Therefore, the number of printed

newspapers, the number of pages, paper amount and ink amount were taken into the account in model fitting.

TABLE II. COMPARISON OF THREE METHODS TO SELECT VARIABLES

Method	Included variables	R <sup>2</sup>	Adjusted R <sup>2</sup>	Significance of variables	Significance of the model
Stepwise	Number of printed newspapers , number of pages, paper amount , Ink amount	0.778	0.766	yes	yes
Backward Elimination	Number of printed newspapers , number of pages, paper amount , Ink amount	0.778	0.766	yes	yes
Forward Selection	Number of printed newspapers , number of pages, paper amount , Ink amount	0.778	0.766	yes	yes

TABLE III. ESTIMATED MODEL FOR MATERIAL COST

Variable	Coefficients	p-value	VIF
Constant	-3072458.8	0.000	
Number of printed newspapers	4.863	0.000	2.081
Number of pages per newspaper	36601.355	0.000	1.056
paper amount(kg)	291.581	0.000	2.085
Ink amount(kg)	-907.529	0.003	3.29

All the coefficients in the model are significant. As VIF values of all independent variables in the model are less than 10 there is no multicollinearity problem in the model. Therefore, we can write the final model as

$$\text{Material Cost} = -3072458.82 + 4.86 \text{ number of printed newspapers} + 36601.36 \text{ number of pages per newspaper} + 291.58 \text{ paper amount} - 907.529 \text{ ink amount}$$

TABLE IV. OVERALL FITNESS OF THE MODEL

R	R Square	Adjusted R Square
.882 <sup>a</sup>	0.778	0.763

Table IV shows R,  $R^2$  and Adjusted  $R^2$  of the final model.

The fitted model explains about 78% of the variability of the observed data and 77% after adjusted for the covariates. Figure 1 reveals that normality assumption is tenable as many points nearly lie on the straight line. Furthermore, it can be concluded that adequacy of the assumption which made the normality of the residuals (P-value of the Anderson darling test is 0.414).

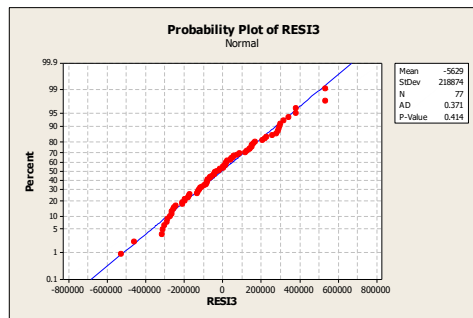


Figure 1. Normal Probability Plot of Residuals

#### IV. CONCLUSION

Different types of methods in multiple linear regression models were applied to find out which variables are affecting to material cost in newspaper printing. Total ink amount used was considered instead of four different colour amount used to avoid the problem of multicollinearity. A number of pages per newspaper, a number of newspapers printed, the paper amount (kg) and ink amount (kg) becomes the significant factors associated with newspaper printing. 78% of the variability in the data is represented by the model. All the assumptions in multiple linear regression models were satisfied by the fitted model. This model can adequately be used to predict the material cost for a newspaper.

#### REFERENCES

- [1] Z. Ismail, A. Yahya, and Shabri, A, "Forecasting Gold Price using multiple linear regression methods," *American Journal of Applied Science* 6, pp. 1509-1514, 2009.
- [2] G. A. Ltd, "Factors influencing the cost of books in South Africa," South African Book Development Council, 2007.