Productivity Estimation of Bulldozers in Earthwork Projects by Multiple Linear Regression Approach: A Case Study

Fatemeh Torfi[a,b]; Abbas Rashidi[c]; Najaf Hedayat[d]; Saeed Torfi[e]

[a]Department of Industrial Engineering, Islamic Azad University, Semnan, Iran

[b] PhD student, Department of Industrial Engineering, Mazandaran University of Science and Technology, Babol, Iran

f.torfi@semnaniau.ac.ir; f_torfi2000@yahoo.com

[c] PhD student, school of building construction, Georgia Institute of technology, Atlanta, GA
rashidi@gatech.edu

[d]Department of Civil Engineering, Islamic Azad University, Dezful, Iran

n.hedayat@yahoo.com

[e]MS student, Department of Psychology, Islamic Azad University, Tehran, Iran

s_torfi2008@yahoo.com

Abstract

The aim of this study is to propose a linear regression approach in order to estimate the productivity of bulldozers in earthwork projects, with due attention to parameters of estimation that are considered as real variables. Productivity Estimation of bulldozers is a crucial factor in earthwork projects. It is particularly important in some municipalities such as Tehran where a sound management calls for a realistic analysis of the earthwork system. The methodology involved critically investigating a least-square linear regression approach (LLR_s) to estimate the productivity of bulldozers in the earthwork problem (EP). A least-square linear regression estimator is applied to original data taken from a case study. The SSR values of the LLR estimator and real productivity are obtained and then compared to SSR values of the nominal productivity and real productivity. Empirical results showed that the proposed method can be viable in solving problems. The results further proved that application of the LLR was realistic and efficient estimator to estimate productivity in earthwork system management and solve relevant problems.

Keywords: Productivity Estimation, Bulldozer, Least-square regression

1. Introduction

Library investigation on soil-related projects in Iran indicated a significant difference between the real-time and estimated efficiency values which cause major challenges in project planning and financing. This seems to be not exclusive to the Iranian cases as there are reported discrepancies between the real and estimated efficiencies elsewhere. The soil-related contractors in Iran have in recent years shown interests to models which could estimate the performance efficiency of earth-moving machineries with a high degree of accuracy while having non-complex and easily applicable features. The first step in timing and estimating the project's operating costs is the necessary work/hours of various machines the achievement of which can be possible by two methods. One is to use the experience and expert's advice and the other use of the manufacturers' manuals. A series of modified coefficients for making appropriate corrections based on the particular operating conditions and environmental settings within which the project is implemented are provided. These include the class of soil to work on, skill of operators, site manager, and other factors which influence the performance efficiency of the machineries [1, 7].

This has resulted in the introduction of various models which the present paper focuses on linear regression type for estimating the hourly performance efficiency of a type of bulldozer that is most frequently-used machine in Iranian construction sector. By time efficiency we mean the volume of earth-displaced per hour. Recent years has seen the development and application of the artificial intelligence networks or the Fuzzy models in engineering domains and manufacturing sector.

The followings are the reasons for the use of linear regression techniques for modeling the estimated operational efficiency of the bulldozers in the Iranian construction industry. The artificial intelligent models require a variety of historical data for training and evaluation. In this case historical data include real-time efficiency of the bulldozer operation under various operating and project conditions. Given the geographical spread of the earth-moving projects, the collection of a vast amount of performance-related data is time-consuming, laborious and sometimes impossible task indeed. A linear regression model however can be applied with a comparatively less data requirement.

In this paper, attempt is made to apply an extension one of regression approaches with real numbers. The proposed methodology consist the least-squares method. The proposed method is assumed to be appropriate alternative approaches to productivity estimation of bulldozers in earthwork projects.

The remainder of this paper is outlined as follows: The next section introduce the method used to compute the productivity of bulldozers. The introduction of D-155 A1 Komatsu bulldozer and application of the algorithm to collected data set is presented in section 3. In section 4 illustrated this method in detail for the specifically-defined problem of this paper then Computational results are represented. Conclusions and future researches which are presented in section 5.

2. Application

The reasons for using the D-155 A1 Komatsu bulldozer compared with similar machineries for earth-moving projects in Iranian construction industry are as follows:

- The suitability of the machine for the operational conditions of the Iranian terrain and topographic features.
- Readily available accessories and equipments for maintenance and major overhaul.
- Availability and quality of the after-sale-services.
- Competitive price of the machine

Table 1. The approximate rate of this model compared with another type of resource

N0	Type of Resource	Approximate Rate (U.S.$ per Hour)*
1	Bulldozer D-155 A1	34-38
2	Loader W-120	15-18
3	Truck (6 m^3)	
4	Site Engineer	4-7
5	Unskilled Labor	1.5-2.5

The relatively high operating costs of this model of machinery necessitate an accurate estimation of its operational efficiency as a major requirement for project work-in-progress planning and financial consideration. It is shown in table 1.

2.1- Data collection

The first priority for developing the linear regression model is collection of data which are usually gathered from the expert's views on factors that influence the bulldozer's performance efficiency.

The second phase involved measurements of the real-time efficiency of 60 bulldozers engaged in 38 earth-moving projects during a one year period. Since some of the crucial factors of efficiency (such as the operator's skills and project) are measured qualitatively and as such being partially subjective and value-ridden, data collection was carried out by earth-moving experts to improve the reliability and validity components.

The earth-moving activities under various topographic and climatic conditions of Iran were selected randomly to include the variability requirements. The main performance criteria were the volume of earth moved during a working shift and the volume that was loaded by the loaders and transported by trucks. By dividing the total earth moved by the number of shift work hours, the performance efficiency was calculated as can be seen in table 2.

Table 2. The main performance criteria

NO	CRITERIA	STATUS
1	Total Service Life Time (Hour)	0-150000
2	Service and Maintenance Condition	Good-Average-Rather poor-Poor
3	Type of Blade	Straight tilt dozer , U-tilt dozer , Semi U-tilt dozer, Angle dozer
4	Max. Blade Capacity (m^3)	4.8 , 6.8 , 8.8 , 11.8
5	Blade Sharpness	Good-Average-Rather poor-Poor
6	Application of Ripper	Yes , No
7	Gear Shifting (Sec)	Less than 5, Between 5~10 , More than 10
8	Operator Technical Skills	Good-Average-Rather poor-Poor
9	General Condition of Operator During Performance	Good-Average-Rather poor-Poor
10	Site Management Condition	Good-Average-Rather poor-Poor
11	Number of Days Operation is in Progress	Between 0 ~100
12	Type of Earth	sand-sandy clay-clay-gravelly soil-Broken rocks
13	Existence of Big Pieces of Rocks	Not existed- Rarely - Commonly
14	Equipment Rotation Space	Easy-Average-Rather difficult-Difficult
15	Grade of the Ground (%)	%25- ~25%
16	Dosing Distance	0~150
17	Time of Operation	MORNING,AFTERNOON,NIGHT
18	Average Temperature During Operation (^0C)	-15 ~ 45
Actual Measured Productivity (Lm^3 Per Hour)		

3. Linear Least Squares Regression

Linear least squares regression is by far the most widely used modeling method. It is what most people mean when they say they have used "regression", "linear regression" or "least squares" to fit a model to their data. Not only is linear least squares regression the most widely used modeling method, but it has been adapted to a broad range of situations that are outside its direct scope. It plays a strong underlying role in many other modeling methods.

- *Definition of a Linear Least Squares Model*

Used directly, with an appropriate data set, linear least squares regression can be used to fit the data with any function of the form

$$f(\vec{x};\vec{\beta}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \quad (1)$$

In which

1. each explanatory variable in the function is multiplied by an unknown parameter,
2. there is at most one unknown parameter with no corresponding explanatory variable, and
3. All of the individual terms are summed to produce the final function value.

In statistical terms, any function that meets these criteria would be called a "linear function". The term "linear" is used, even though the function may not be a straight line, because if the unknown parameters are considered to be variables and the explanatory variables are considered to be known coefficients corresponding to those "variables", then the problem becomes a system (usually over determined) of linear equations that can be solved for the values of the unknown parameters. To differentiate the various meanings of the word "linear", the linear models being discussed here are often said to be "linear in the parameters" or "statistically linear".

- *Advantages of Linear Least Squares*

Linear least squares regression has earned its place as the primary tool for process modeling because of its effectiveness and completeness. Though there are types of data that are better described by functions that are nonlinear in the parameters, many processes in science and engineering are well-described by linear models. This is because either the processes are inherently linear or because, over short ranges, any process can be well-approximated by a linear model.

The estimates of the unknown parameters obtained from linear least squares regression are the optimal estimates from a broad class of possible parameter estimates under the usual assumptions used for process modeling. Practically speaking, linear least squares regression makes very efficient use of the data. Good results can be obtained with relatively small data sets. Finally, the theory associated with linear regression is well-understood and allows for construction of different types of easily-interpretable statistical intervals for predictions, calibrations, and optimizations. These statistical intervals can then be used to give clear answers to scientific and engineering questions.

3.1. Least Squares

In least squares (LS) estimation, the unknown values of the parameters, $\beta_0, \beta_1, \cdots$, in the regression function, $f(\vec{x}; \vec{\beta})$, are estimated by finding numerical values for the parameters that minimize the sum of the squared deviations between the observed responses and the functional portion of the model. Mathematically, the least (sum of) squares criterion that is minimized to obtain the parameter estimates is

$$Q = \sum_{i=1}^{n} [y_i - f(\vec{x}_i; \vec{\beta})]^2 \qquad (2)$$

As previously noted, $\beta_0, \beta_1, \cdots$ are treated as the variables in the optimization and the predictor variable values, $x_1, x_2, \cdots$ are treated as coefficients. To emphasize the fact that the estimates of the parameter values are not the same as the true values of the parameters, the estimates are denoted by $\hat{\beta}_0, \hat{\beta}_1, \cdots$. For linear models, the least squares minimization is usually done analytically using calculus. For nonlinear models, on the other hand, the minimization must almost always be done using iterative numerical algorithms.

- Significance testing for regression parameters

The following formula is used to obtain t- statistic for each β coefficients:

$$t = \frac{\hat{\beta}_i}{\sqrt{\mathrm{var}(\hat{\beta}_i)}}$$

Where

$$Var(\hat{\beta}_i) = \frac{\sigma^2}{\sum_j (x_{ij} - \bar{x}_i)^2}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}$$

If

$$T > t_{\frac{\alpha}{2}, n-1}$$

Then β coefficients are Significance.

- Total significance testing for regression model

Another aspect of regression analysis is accounting for the variation in the dependent variable as it relates to variation in the independent variable. A fundamental equation is

$$y - \bar{y} = (y - \hat{y}) + (\hat{y} - \bar{y}). \quad (3)$$

Equation (3) states that a deviation of a value of the dependent variable from the overall mean, $y - \bar{y}$, is equal to the sum of a deviation of the dependent variable from the predicted value, $y - \hat{y}$, plus the deviation of the predicted value from the overall mean, $\hat{y} - \bar{y}$. It can be shown That $\hat{y} - \bar{y} = 0$ when $x = \bar{x}$. Also, $\hat{y} - \bar{y}$ changes by an amount $\hat{\beta}$ for each unit of change in x. Thus, the deviation $\hat{y} - \bar{y}$ depends directly on the independent variable *x*. But the deviation $y - \hat{y}$ can be large or small, and positive or negative, for any value of *x*. So the deviation $y - \hat{y}$ does not depend directly on *x*.

It turns out that the sums of squares of the deviations in equation (7) obey a similar equation,

$$\sum (y - \bar{y})^2 = \sum (y - \hat{y})^2 + \sum (\hat{y} - \bar{y})^2 \quad (4)$$

The sums of squared deviations have names. The left side of the equation is call the total sum of squares, and is denoted SS(Total)=$\Sigma(y - \bar{y})^2$. The terms on the right side are the error and regression sums of squares, SS(Regression) =$\Sigma(\hat{y} - \bar{y})^2$ and SS(Error)=$\Sigma(y - \hat{y})^2$. For brevity, We write SSR=SS(Regression), SSE=SS(Error), and SST=SS(Total), and equation (4) takes the form

SST=SSR+SSE. (5)

The total sum of squares, SST, is a measure of the total variation in the values of the dependent variable y. The regression sum of squares, SSR, is a measure of the variation in y that is attributable to variation in the independent variable x. Finally, the error sum of squares measures the variation in y that is not attributable to changes in x. Thus equation (5) shows the fundamental partitioning of the total variation into the portion attributable to x and the portion not attributable to x. The coefficient of determination, usually denoted R^2, is SSR divided by SST, and thus measures the proportion of total variation in y that is attributable to variation in x, R^2 = SSR/SST.

- Total significance testing for regression model

The following formula is used to obtain F- statistic:

$$F = \frac{SSR / df(R)}{SSE / df(E)}$$

Where

Total degree of freedom (df(T))= (n-1)
Regression degree of freedom (df(R))=number of β coefficients
Error degree of freedom (df(E))= df(T)-df(R)

$$Var(\hat{\beta}_i) = \frac{\sigma^2}{\sum_j (x_{ij} - \bar{x}_i)^2}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}$$

If

$$\text{F} > F_{df(R), df(E), \alpha - 1}$$

Then regression model are Significance.

4. Numerical results

In Sections 3, least-squares method have been constructed for the estimation of productivity of bulldozers in earthwork projects with due attention to parameters of estimation that are considered as real variables.

In this section some numerical results are given. The first step for constructing the linear regression model involved collection of available data about the problems to be solved. For the purpose of this study, the most important and effective factors, which are perceived by the experts to influence the productivity made by a bulldozers, will be determined.

The second step involved measuring the real productivity for the 60 operating bulldozers in various projects in Iran. The following factors were taken into consideration:

Independent variables which have entered the model for estimating the performance efficiency of bulldozers included factors such as those described in table 2. The tine type variable(X3) has four different states such as, U, semi-u, angle dozer and straight. From these four varieties, the type "u" is considered to be the best state, followed by semi-u, angle dozer and straight.

Let us apply regression model to this case study. The algorithm is implemented on data set, with the results shown in Table 3.

Since the following function is obtained by least squares method at the same time measuring the real-time performance efficiencies of a fleet of 60 bulldozers with SPSS software:

$$\begin{aligned} y &= 78 + 0.000041X_1 - 11.8X_2 + 17.2X_4 - 19.9X_3(u) + 18.9X_3(angledozer) - 50.8X_3(straight) + 0.5X_5 \\ &+ 15.8X_6 + 9.2X_7 + 1.8X_8 + 28.6X_9 + 1.06X_{10} + 0.701X_{11} - 124X_{12}(brockenrocks) - 29.5X_{12}(sandy-clay) \\ &- 92.5X_{12}(gravelysoil) - 82.1X_{12}(clay) + 4.2X_{13} - 9.4X_{14} - 1.18X_{15} - 1.52X_{16} \\ &- 34.3X_{17}(evening) - 83.3X_{17}(night) - 0.702X_{18} \end{aligned}$$

Table.3.Results of linear regression model for the 60 bulldozers

Sig مقدار	t مقدار	مقدار ضریب	آماره
.520	.650	78	(Constant)
.664	.439	.000041	X_1
.294	-1.065	-11.8	X_2
.011	2.677	17.2	X_4
.278	-1.102	-19.9	$X_3(u)$

.540	.619	18.9	$X_3(angledozer)$
.075	-1.836	-50.8	$X_3(straight)$
.971	.037	0.5	X_5
.349	.950	15.8	X_6
.369	.910	9.2	X_7
.875	.158	1.8	X_8
.036	2.178	28.6	X_9
.913	.110	1.06	X_{10}
.310	1.031	0.701	X_{11}
.000	-5.559	-124	$X_{12}(brockenrocks)$
.180	-1.370	-29.5	$X_{12}(sandy-clay)$
.002	-3.354	-92.5	$X_{12}(gravelysoil)$
.000	-4.154	-82.1	$X_{12}(clay)$
.737	.339	4.2	X_{13}
.459	-.748	-9.4	X_{14}
.083	-1.787	-1.18	X_{15}
.000	-5.084	-1.52	X_{16}
.023	-2.375	-34.3	$X_{17}(evening)$
.148	-1.479	-83.3	$X_{17}(night)$
.341	-.966	-0.702	X_{18}

It is found that the F-statistic using algorithm is 6.365 greater than 1.96 (for greater than 99% confidence). Thus total significance testing is confirmed for regression model. A non-graphical test is the Shapiro-Wilk test for normality. It tests the hypothesis that the distribution is normal; in this case the test is that the distribution of the residuals is normal. The results are also shown in Table 4.

Table 4. Tests of Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
Standardized Residual for راندمان	.981	60	.462

A scatter plot is also very useful when we wish to see how two comparable data sets agree with each other. In this case, the more the two data sets agree, the more the scatters tend to concentrate in the vicinity of the identity line. A scatter plot reveals relationships or association between two variables. This sample plot reveals a linear relationship between the two variables indicating that a linear regression model might be appropriate. The t-values test the hypothesis that the coefficient is different from 0. To reject this, you need a t-value greater than 1.96 (for 95% confidence). You can get the t-values by dividing the coefficient by its standard error. The t-values also show the importance of a variable in the model. R-square shows the amount of variance of Y explained by X. In this case the model explains 81.43% of the variance.

At the same time measuring the real-time performance efficiencies of a fleet of 60 bulldozers, their efficiency was calculated from the manufacturers' manual data. The efficiencies of these machines were also calculated by the linear regression models. Calculation of the efficiencies of three methods including the manuals, regression models and their comparisons with the real-time performance of the machines, the deviation in three methods were obtained the results of which are described in the following figure 1.

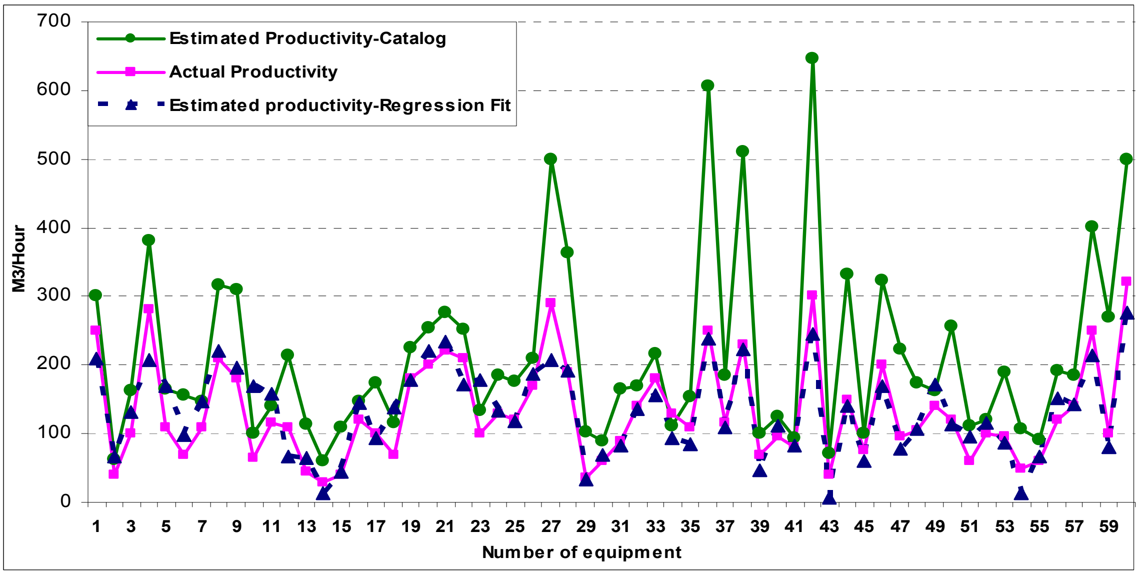

Figure 1. Difference between actual productivity and catalog, with actual productivity and regression model.

5. Conclusions

In the present study, a linear regression approach, least squares is suggested for solving the problem. The aim of this study is to propose a linear regression approach in order to estimate the productivity of bulldozers in earthwork projects, with due attention to parameters of estimation that are considered as real variables. The application of this method in this study provided a solution which seems to be closer to the real world. Results also indicated that the method used in this investigation presented better answer than the data on productivity estimation of bulldozers in earthwork projects which are obtained from the company documents.

References

[1] G.D. Anon, Caterpillar performance Handbook, Caterpillar, Illinois, USA, 1997.

[2] Specifications and Application Handbook, 24th Edition, Komatsu Publications and Training Group, Tokyo, 2003.

[3] A.M.Elazouni, I.M.Basha, Evaluating the Performance of Construction Equipment Operators in Egypt, ASCE Journal of Construction Engineering and Management 122 (2) (1996) 109-114.

[4] D.W.Halligan, L.A.Demsets, J.D.Brown and C.B.Pace, Action-response Model and Loss of Productivity in Construction, ASCE Journal of Construction Engineering and Management 120 (1) (1994) 47-64.

[5] S.D.Smith, Earthmoving Productivity Estimation Using Linear Regression Techniques, ASCE Journal of Construction Engineering and Management 125 (3) (1999) 133-141.

[6] S.D.Smith, G.S.Wood, M.Gould, A New Earthworks Estimating Methodology, Journal of Construction Management and Economics, 18 (2) (2000) 219-228.

[7] L.Chao, M.J.Skibniewski, Estimating Construction Productivity: Neural-Network-Based Approach, ASCE Journal of Computing in Civil Engineering, 8 (2) (1994) 234-251.

[8] C.M. Tam, T. K.L. Tong & S. L. Tse, Artificial neural networks model for predicting excavator productivity, Engineering, Construction and Architectural Management,9 (5) (2002) 446 - 452

[9] H.Adeli, Novel Computing for Intelligent Infrastructure, Proceedings of 12th International Conference on Computing in Civil and Building Engineering, October 2008, Beijing, China

[10] S.W.Nunnally, Managing Construction Equipment, 2nd Edition, Prentice-Hall, June 2002, N.J., USA

[11] Richard A. Berk, Regression Analysis: A Constructive Critique, Sage Publications (2004)

[12] David A. Freedman, Statistical Models: Theory and Practice, Cambridge University Press (2005)

[13] Rencher , C.A. and Bruce Schaalje , G.(2007). Linear Models in Statistics.