Primljen / Received: 26.2.2015.
Ispravljen / Corrected: 11.7.2017.
Prihvaćen / Accepted: 22.9.2017.
Dostupno online / Available online: 10.2.2020.

Determination of residual value of construction machinery based on machine age

Authors:

Igor Milošević, MCE
Dabar D.O.O., Serbia
igornmilosevic@gmail.com
Corresponding author

Prof. **Predrag Petronijević**, PhD. CE
University of Belgrade, Serbia
Faculty of Civil Engineering
pecap@grf.bg.ac.rs

Prof. **Dragan Arizanović**, PhD. CE
University of Belgrade, Serbia
Faculty of Civil Engineering
ari@grf.bg.ac.rs

Subject review

Igor Milošević, Predrag Petronijević, Dragan Arizanović

Determination of residual value of construction machinery based on machine age

The relationship between the residual value and service life of machines is investigated in the paper. The study resulted in establishment of several models based on symbolic regression, where the dependent variable is the residual value, while independent variables are machine models, machine age, cumulative number of machine operating hours, and inflation index. The estimate of residual value is crucial for making investment decisions. The investigation has shown that development of a general prognostic mathematical model of residual value of construction machinery can be made using symbolic analysis of an available set of data formed of 61,153 machines.

Key words:
construction machinery, residual value, symbolic analysis

Pregledni rad

Igor Milošević, Predrag Petronijević, Dragan Arizanović

Određivanje rezidualne vrijednosti građevinske mehanizacije na temelju starosti strojeva

U radu je istražen odnos između rezidualne vrijednosti i vijeka trajanja strojeva. Rezultat istraživanja je kreiranje nekoliko modela primjenom simboličke regresije u kojoj je zavisna varijabla rezidualna vrijednost, nezavisne varijable su modeli strojeva, kalendarska starost stroja, kumulativni broj radnih sati stroja i indeks inflacije. Procjena rezidualne vrijednosti je ključna za donošenje investicijskih odluka. Istraživanje je pokazalo da postoji mogućnost razvijanja okvirnog prognostičkog matematičkog modela rezidualne vrijednosti građevinske mehanizacije koristeći simboličku analizu dostupnog seta podataka koji čine 61 153 stroja.

Ključne riječi:
građevinska mehanizacija, rezidualna vrijednost, simbolička analiza

Übersichtsarbeit

Igor Milošević, Predrag Petronijević, Dragan Arizanović

Ermittlung des Restwertes von Baumaschinen anhand des Alters der Maschinen

In der Abhandlung wird der Zusammenhang zwischen dem Restwert und der Nutzungsdauer der Maschinen dargelegt. Das Ergebnis der Untersuchung ist das Erstellen einiger Modelle unter Verwendung der symbolischen Regression, wobei die abhängige Variable der Restwert ist, die unabhängigen Variablen sind die Maschinenmodelle, das Kalenderalter der Maschinen, die kumulative Anzahl an Arbeitsstunden und der Inflationsindex. Die Schätzung des Restwertes ist der Schlüssel für Investitionsentscheidungen. Die Untersuchung hat gezeigt, dass es möglich ist, ein prognostisches mathematisches Rahmenmodell für den Restwert von Baumaschinen unter Verwendung einer symbolischen Analyse des verfügbaren Datensatzes zu entwickeln, der 61 153 Maschinen umfasst.

Schlüsselwörter:
Baumaschinen, Restwert, symbolische Analyse, Baumaschinen

1. Introduction

The purchase of construction machines is a serious investment for every construction company. Choosing the right moment for selling a used machine and determining its value are two important steps. In this research paper, a special attention will be paid to the concept and assessment of residual value. The relationship between the ownership and operating costs, residual value, and the problem of determining this price in practice, will be determined in the first step. A detailed analysis of the literature in which this concept is mentioned will be carried out in the following sections. The objective of this paper is to observe the relationship between the residual value and the age of machines. The result of this research is creation of several models through symbolic regression in which the dependent variable is the residual value, while independent variables are the construction machine model, machine age, and cumulative number of machine working hours.

This research aims to assist in the creation of regression models that can predict residual value of machines at auctions. The research has shown that there is a difference in the sensitivity of variables for machines with a total cumulative number of hours greater than 2000, compared to the machines with less than 2000 working hours. The residual value of machines has a higher sensitivity compared to the variable that describes calendar age of machines. This sensitivity is even more pronounced in the present model when the cumulative number of hours is greater than 2000. The machine value on the auction market declines sharply in the first years of exploitation for certain types of machines. This phenomenon can be explained by an unequal rate of decrease in residual value of machines of certain manufacturers during the first years, which is mainly dependent on the brand and manufacturer.

2. The concept of residual value

Residual value is defined as the price at which the used machinery can be sold on the market in a certain period. In the literature, there is a large number of terms that describe the concept of residual value. For example, the terms resale value, remaining value, regained value, saved value, partial value, final value, and fair market value are used to describe this concept [1, 2].

Machine costs accumulate over time, thus reducing the value of the machine. At a certain moment, it becomes more profitable for the construction company to sell the machine than to keep it. An optimum machine lifetime is also called the machine's economic life or useful life and can be significantly shorter than the physical life of the machine based on repairs. The machine's actual lifetime depends on machine wear during its service life and can be prolonged with adequate maintenance and repairs.

The residual value of a part of equipment is defined as "the amount of money for which the machine can be sold at a certain time" on the market [3]. When it comes to equipment assessment, this definition can be expanded by adding information about the exact circumstances of the sale; whether it is an equal sale between two equal parties, auction, liquidation, or resale. Additional confusion is caused by the use of the term "depreciation", which means the decrease in the machine's initial value. In the field of accounting, depreciation is the systematic process of allocating the purchase value of the fixed asset on expenditures during its service life. The residual value occurs with the loss of the equipment's value. Depreciation can be linked to the equipment itself, in terms of the physical condition, age, wear or obsolescence, or to the economic situation (offer and demand for the equipment or its products), on the basis of which the value is determined [4, 5]. This is different from the term "depreciation" as used in accounting. In accounting, it refers to a decrease in the book value of assets. This is calculated in such a way that the investment costs are deferred and periodically allocated to products for the purpose of administrative and tax liabilities. The concept of depreciation comes from the cost accounting, where it is used to assess "the value loss of a part of the equipment over time" [6] when collecting income.

For many machine owners, the potential resale value is the main factor when purchasing. A high resale value represents a lower depreciation, i.e. a lower overall cost, and it improves competitive position of the equipment. If the machine is resold long before it has been fully depreciated, its resale value is much higher. Although the actual resale value of the equipment at the moment of resale is determined by market value, there are ways to determine the resale value of the equipment using various depreciation methods. There are several common methods for calculating depreciation, such as the linear method, double declining balance method, and functional method. Double declining balance method is used to show how to calculate the ownership costs. However, any of these methods for calculating depreciation can be used.

The residual value may depend on various factors, such as the current machine condition, location and time of sale, and the state of technology and economy in that period. The mood at the moment of sale can also affect the sales price. Therefore, the assumption of a fixed residual value is unrealistic. It is much better to assess the residual value using previous data, such as the "evidences of value" [7] and, in particular, the data on the actual sale of machines.

When analysing Vorster's papers [8, 9] one encounters the concept of the current residual value RV_o (1), which is used for calculating loan repayment when renting construction machines. The current value of an annuity A (2) is equivalent to the normalized difference between the machine's purchase value and the current residual value.

$$RV_o = \frac{RV}{(1+i)^N} \tag{1}$$

$$A = \frac{(PP - RV_0)i}{1 - \left[\left(\frac{1}{1+i}\right)^{(N)}\right]} \tag{2}$$

$$AH = \frac{\frac{A}{h}}{N} \tag{3}$$

Where RV_o is the current residual value in dollars, RV is the residual value in dollars, *i* is the interest rate, N is the calendar year, A is the annuity of the current value and PP is the purchase price.

3. The concept of ownership and operating costs

Earthworks requiring the use of heavy construction equipment are usually necessary in heavy industry or in the construction of highways. However, large construction machines can also be found in other fields, depending on the projects in which they are used. The greater the share of earthworks in the project, the greater the share of equipment costs in the total project costs. Cost analysis of the utilization of construction equipment is an essential function for the owner of such equipment and is of vital importance for the success of the company. Focusing on individual machines, rather than on entire equipment, the costs related to some parts of the equipment can easily be divided into two categories: ownership costs and operating costs.
The individual cost elements for the parts of these categories have been described by Lucko and Vorster [1, 2]. The ownership costs consist of the price for which the machine has been purchased, insurance, licenses, ownership, delivery, assembly charge, and taxes. The ownership costs include all the costs of funding, such as interests on the principal amount and the like. Each of these elements is independent of the use of the machine. The sale of the machine on the market after the cessation of ownership creates a financial gain, which is the residual value.
The operating costs are incurred through the use of the machine. They consist of the costs of fuel, oil and lubricants, as well as the costs of regular servicing and repairs. These costs should include the costs of purchasing replacement parts, cost of the parts that are often replaced, such as tires or treads, cost of overhaul and complete reparation, and all workforce costs incurred during the machine repair, such as salaries and bonuses to repairmen.
When residual value of the equipment needs to be determined for the purposes of making relevant decisions such as the equipment repair, overhaul, removal or replacement, and the market price (e.g. auction) has not yet been determined for the equipment in question, then the price for a certain part of the equipment will be determined based on previous auction experiences for similar products. Since a large number of factors affect market value of construction equipment, such as age, manufacturer, model, intensity of use, and maintenance, as well as market demand and supply, and so on, it is not possible to establish industrial criteria for determining the price of used construction equipment.

4. Previous research in the field of residual value assessment

Lucko [1] developed a regression model for predicting residual value of different types of equipment. The variables for regression model he used in this research were the age, manufacturer, condition rating, geographic region, and some macroeconomic indicators. Vorster's work [11-15] contributed to the research on cumulative value of the cost of machines, while also providing better understanding of the concept of a decrease in residual value of equipment based on machine age.
The productivity of machinery changes during exploitation of machines and decreases over the years; therefore, the machine value on the market also decreases. Kannan [16] demonstrated some machine data collection methods and showed statistical indicators by which machine productivity is determined. Hildreth [17] demonstrated the use of a global positioning system (GPS) in the analysis of productivity of machines at a construction site. Martizne [18] discovered a general purpose simulator that can be used to determine productivity of equipment on complex projects.
The equation (4) was proposed by Vorster [11] in an attempt to empirically determine behaviour of residual value in such a way that the value of the equipment rapidly decreases at the beginning of use and slows down in later years:

$$RV = K \cdot PP \cdot \frac{1}{\sqrt{\frac{h_{rv}}{1000}}} \tag{4}$$

Where RV represents the residual value, K is the adjustment factor, PP is the purchase price and h_{rv} are the total hours of use for all machines that operated for more than 2000 hours. The K factor is the adjustment factor showing the parts of the residual value in order to demonstrate when non-standard machines have a lower (or higher) value compared to standard machines; it ranges from 0.0 to 1.0.
The formula (4) is used for machines [1, 2] with a total number of more than 2000 working hours. A modified formula (5) is used for machines with up to 2000 working hours:

$$\frac{h}{N} < 2000 \rightarrow h_{rv} = h + 1000 \cdot F \cdot \frac{2N - \frac{h}{1000}}{2} \tag{5}$$

where h is the number of hours, N is the machine age expressed in years, and hrv is the number of hours for calculating the residual value. The machine use coefficient F (with a value from 0 to 1), introduced by Kastens [18], is calculated on the basis of the internal database of the company itself. In this way, it is possible to make a difference in the calculation of the number of hours of a machine that has been used very little, compared to the one that has been used longer.
The determination and analysis of the residual value of agricultural equipment and logging equipment in forestry show the importance of residual value assessment in investment decision making [19]. The conclusion of this research is that the residual value is important because it affects the amount of investment that must be returned through the use of assets. The equipment assessed as having a high residual value will have lower depreciation costs compared to equipment with a lower residual value.
Cross and Perry [20] also wrote about the use of catalogue prices "which were used as a replacement for the actual machine sales prices, because the original prices were unavailable. The catalogue price was considered the closest value available to represent the sales price".
Cross and Perry [20] conducted one of the most comprehensive studies on the depreciation of agricultural equipment. The objective of their research was to find an appropriate mathematical formula (6) for calculating the relationship between the residual value of the equipment (RV) and known variables:

$$RV = f\begin{pmatrix} \text{age, use, maintenance,} \\ \text{manufacturer, type of auction,} \\ \text{region, microeconomic variables} \end{pmatrix} \qquad (6)$$

Statistical regression was used to create the model for predicting residual value of heavy construction equipment [2, 21].

5. Measuring machine age

The age of machine can be calculated in three ways [3]: machine age per productivity unit, machine calendar age, and machine age based on total number of operating hours.

Measuring machine age per productivity unit tells us how many tasks the machine has carried out. In this case, it is necessary to define the utilization unit for each machine. This is easy for some machines. The traction unit can indicate line movement of some volumetric measurement [3]. For some equipment, it is very difficult to define the productivity unit. A good example of this is the machine responsible for maintaining traction on the road.

There are three types of hours that can be monitored by construction companies. Construction hours are the hours for which the company charges for use of the machine at a particular workplace. These hours may or may not be an accurate measurement of how much the machine has been used at that workplace. For example, 40 hours' work on the machine may be charged, but it does not necessarily mean that the machine has been active for all of those 40 hours. In some cases, the construction hours are the hours reported by the inspector on the construction site. It can happen that the construction hours are deliberately decreased so that the business appears profitable.

Working hours are the hours during which the machine is operating. They indicate the measure of time. One way of monitoring working hours is to use a notebook filled in daily by machine operators. The time that they spend on the machine is the working time of that machine. The hours per time are the hours recorded by the meter installed on the machine. The machine's calendar age is calculated very simply as the difference between the current date and the date of purchase of the new machine. In the literature, this value is expressed in years, months or days. Due to the lack of new projects or due to bad weather conditions, construction firms often do not use the machinery for long periods. Therefore, the machine's calendar age is an unreliable information on the condition and amount of machine use. On the other hand, this value has a significant sensitivity in models that predict residual value of machines at auctions.

The machine age, measured by the total number of working hours, represents the value measured by the enterprise itself and is directly conditioned by the consistency and dedication of the owner. The sensitivity of the variables of the machine's calendar age, as well as the age measured by the total number of working hours in relation to the residual value, will specifically be analysed in this paper.

6. Research methodology

The residual value represents the machine price, which is very difficult to determine using a mathematical formula. Auctions are an excellent way to assess the value of used machines. Certain models of construction machines with a high reputation and excellent quality, despite a high number of working hours, can have a higher residual value than newer machines of the same capacity and rank. The influence of machine age on the used-machine sales price at auctions represents the initial basis of this research.

6.1. Hypothesis

Residual value estimates are essential for making investment decisions. It is possible to develop a predictive mathematical model for the residual value of heavy construction equipment using the symbolic regression analysis of a dataset composed of information from 61,153 machines, as a function of the age and working hours of the machine.

6.2. Research objectives

Although various algorithms use different methods to make conclusions about the models, they do have some common characteristics that make them better than the traditional statistical regression approaches for predicting residual value of equipment [22]. The assumption of statistical distribution or fundamental functional forms makes a statistical regression model subjective. Using the built-in models for predicting residual value of equipment, sellers can determine the best moment at which to sell their equipment, customers can determine the best moment to buy the equipment they need, and the owners of the equipment can analyse the operating life of equipment in order to make a decision about repairing, removing, or replacing the equipment. The application creates the model using genetic algorithms and enables the management team to gain insight into a large amount of data on the construction equipment that has been collected, and to make decisions at an appropriate time.

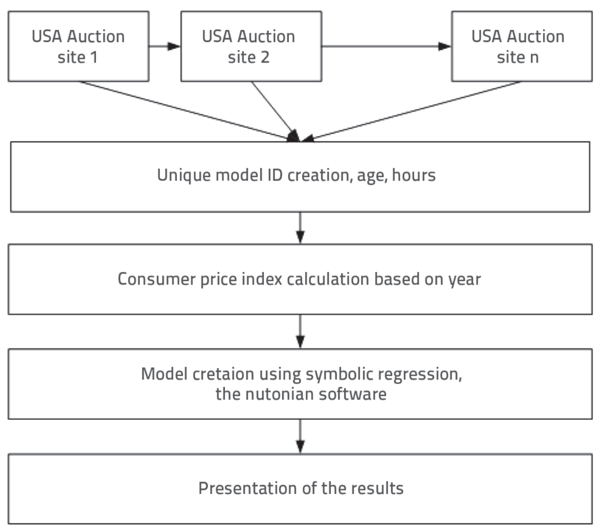

Figure 1. Activity workflow

The data used in this study were taken from America's leading auction sites that specialize in the buying and selling of heavy equipment. After data collection, each machine type was defined by its identification number (Model ID). In the next step, only the machines with a clearly defined number of working hours and age expressed in the number of days, were singled out. The Consumer Price index value was calculated in the next step (Chapter 7). In the last step, a model was created by applying symbolic regression, and a statistical interpretation of results was made ((Chapters 8 and 9).

6.3. Research scope and limitations

A machine database comprised of data from several North American auction sites was used for this research. The currency for the sale price is the US dollar. In this research, the influence of inflation was represented as the Consumer Price index value. One of the limitations in this study is uncertainty in the input variables. The data are collected from the auction sites and are dependent on human behaviour. Every data-mining process and model can be improved. The accuracy of the model can be increased by allocating additional software hours or by using a larger number of networked servers. This application model is designed for operation of the software within 48 hours using a dual core processor. The limitation is that a large amount of data and the small number of attributes result in a reduced accuracy of the models. The aim is to create models that can predict the preliminary (rough) residual value of a machine.

7. Research data

The data were collected in the period from 1989 to 2012. Only the data of the construction machines with a clearly defined machine model, total number of working hours, and machine age, were used in this research. The key that identifies the machines is Model ID; this is the attribute that connects multiple auction sites. Two of the biggest problems when sorting data are the problem of consolidating data from different sites and assigning the attribute of Model ID, as well as amending incorrectly written or abbreviated machine names. A total of 61,153 construction machines of different models were studied. The aim of the study was to create a model in which the dependent variable is the residual value (7), while the independent variables are the type of Model ID, machine calendar age, and the cumulative number of the machine operating hours:

$$\text{Sales price} = (\text{Model ID, age, hours, CPI}) \tag{7}$$

where the sales price represents the residual value; the age represents the machine's calendar age expressed in months; and the hours represent the cumulative number of the machine's operating hours, while CPI is the Consumer Price Index for the US dollar. The Consumer Price Index in the United States is used to calculate inflation. In this paper, CPI is value matched. The CPI value is equal to the average annual value at the date of sale.
For all machines, 1 January was taken as the starting date of the machine's calendar age, while the ending date was the sales date. The model sensitivity will be analysed in this research.

8. Symbolic regression

Regression refers to the process of finding coefficients in a predefined function so that it is fitted into the model as easily as possible. The problem with regression analysis is that it is necessary to find an appropriate model in a laborious way. In symbolic regression, the observer does not participate in the creation of hypotheses and theoretical models, but this task is left to the unconscious mechanism of evolution.
Symbolic regression is a type of regression which seeks the model and variables in order to find the model that best fits a given data set. Initial expressions [23] are formed randomly by combining mathematical blocks, such as basic mathematical operators, analytic functions, constants, variables, logical operators, etc. New equations are formed by combining previous equations, using a genetic programming process. In complex cases when scientists have a large number of data, the process of selecting mathematical formulas by using genetic algorithms is very slow [24].
The reliability of sellers, [25, 26] the appearance of the listing, and beginning and ending times of the listing, are all factors that make it difficult to predict the price at an auction. Even if all of the above variables are accounted for, there is still the uncertainty in human behaviour when bidding at auctions. Using symbolic regression, scientists let the data point to an appropriate model form, instead of imposing their a priori assumptions. A good model should be both accurate in terms of capturing the observed data behaviour and free of unnecessary structures and variables, as can be the case in predefined models.
The Nutonian Software created by Michael Schmidt and Hod Lipson [23] was used for the purposes of this experiment. When a researcher has a huge amount of data in the analysis and the domain of knowledge about the data generating system is limited, it is the task and responsibility of the researcher to prune the data variables to an uncorrelated subset and guess the right model form, or to try a machine-learning method. Hod Lipson [23, 27] tried to compare the Eureqa Nutonian software and symbolic regression method to other machine-learning methods, such as Neural Networks, Support Vector Machines, Decision Trees, and simple Linear Regression.

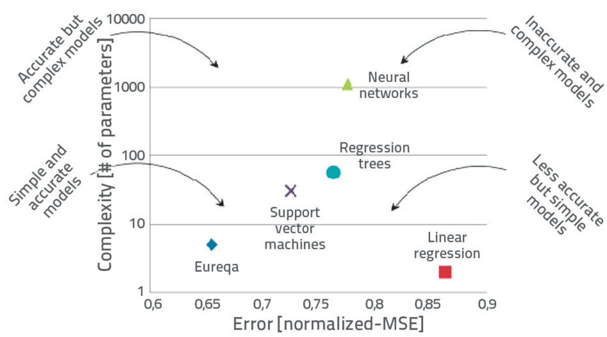

Figure 2. Hod Lipson comparison of Nutonian software to other machine learning methods

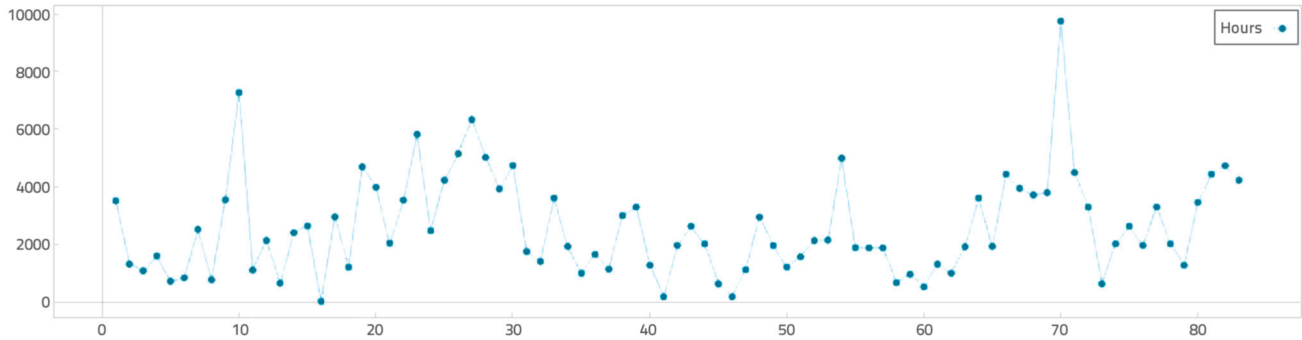

Figure 3. Display of data entered for dozer D39PX-21

In this comparison [27], one can see that Eureqa's use of symbolic regression apparently produces models that are both more accurate and simpler than other machine learning methods.

9. Example of model creation through symbolic regression

This example will demonstrate the process of creating model sensitivity using the data for only one construction machine model. The machine selected is Dozer D39PX-21 and 83 samples will be analysed. This model is most often repeated in the database. The data are presented in the Appendix. Model ID is the identifier of a unique machine model (for example Model ID=216 for all dozers D39PX-21). The age of the machine is calculated as the number of days between the sale date and the year of manufacture. The first of January was used as the starting date of the machine's calendar age (e.g., if the year of manufacture is 2007, then the age is the number of days from 1.1.2007 to the sale date).

Table 1. Display of statistics and sensitivity of model obtained based on data entered for dozer D39PX-21

Statistics	
R^2 goodness of fit	0.42
Correlation coefficient	0.66
Maximum error	23.092.78
Mean squared error	51.715.404.00
Mean absolute error	5.357.86
Variable sensitivity	
Variable hours sensitivity	0.99085
Variable age sensitivity	0.57814
Variable CPI sensitivity	0.013194

After normalization of the data, the next stage is the removal of extreme values (outliers). In this paper, the removal of outliers was performed automatically, through deletion of all the values above the double value of the difference between the third and first quartiles IQR (interquartile range). The data-mining model was built using the combination of formula building blocks as constant, integer, add, subtract, multiply, divide, negation, exponential, logarithm, factorial, power, and sqrt. Model-building software created 32 billion formula evaluations.

Various statistical parameters can be used to gain insight into the quality of the statistical model. The fit coefficient provides a display of how much data fitting can be expected from the regression model. This coefficient ranges from 0 (all data fit) to 1 (no data fit into the model). Based on Gunnar's [1] research on the small number of machines R^2 Goodness of Fit, less than 0.7 can be adopted as a good model. In this case, the model R^2 Goodness of Fit is 0.42 (Figure 4).

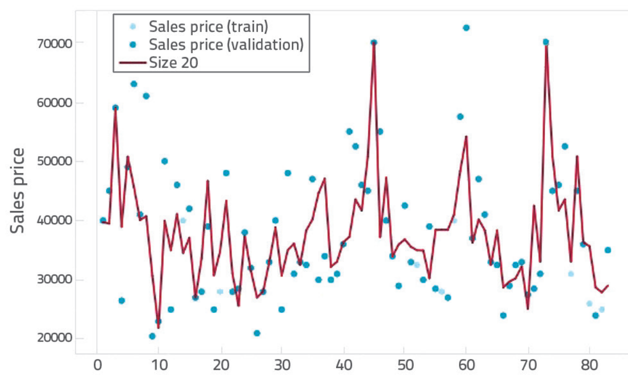

Figure 4. Display of fitting entered values in the model for D39PX-21

This dozer model was manufactured in the period from 2003 to 2007 and about half of the samples (45 machines) had a cumulative number of working hours greater than 2000. In the next step, a model will be made to observe only those D39PX-21 dozers that have worked for more than 2000 working hours, along with the model for those dozers that have worked for less than 2000 hours.

It can be seen in Table 3 and Table 2 that the prediction model with the Goodness of Fit of less than 0.7 can be created. The R^2 Goodness of Fit of dozer D39PX-21 machines that have less than 2000 working hours is 0.83 and the Fit for all D39PX-21 machines model is 0.42, as newer machine prices can be difficult to predict. In the first predictive model, where 82 most common machines were used for analysis, it can be observed that the dependent variable (DV) - residual value - and the independent variable - Retail

Table 2. UDisplay of input data for model creation using the example of the dozer D39PX-21

Year made	Age	Hours	CPI	Sale price
2007	899	3505	207.3	40000
2006	1516	1310	201.6	45000
2008	597	1078	215.303	59000
2006	1334	1593	201.6	26500
2007	798	715	207.3	49000
2003	1317	838	184	63000
2001	1417	2516	177.1	41000
2006	734	767	201.6	61000
2003	1925	3545	184	20500
2002	2480	7266	179.9	23000
2006	1019	1100	201.6	50000
2005	1412	2131	195.3	25000
2006	1082	648	201.6	46000
2005	1489	2399	195.3	40000
2006	1432	2637	201.6	42000
2005	1572	16	195.3	27000
2005	1572	2951	195.3	28000
2007	899	1203	207.3	39000
2005	1783	4685	195.3	25000
2006	1380	3977	201.6	28000
2007	1311	2039	207.3	48000
2004	2324	3533	188.9	28000
2003	2596	5815	184	28500
2006	1747	2470	201.6	38000
2005	2080	4223	195.3	32000
2003	2948	5141	184	21000
2005	2266	6327	195.3	28000
2006	1901	5012	201.6	33000
2007	1536	3921	207.3	40000
2005	2502	4736	195.3	25000
2004	1148	1749	188.9	48000
2004	2659	1401	188.9	31000
2005	1907	3605	195.3	33000
2006	1242	1923	201.6	32500
2006	962	993	201.6	47000
2007	920	1646	207.3	30000
2007	920	1135	207.3	34000
2003	2628	2994	184	30000
2005	2293	3290	195.3	31000
2005	2328	1268	195.3	36000
2005	1138	175	195.3	55000
2007	1360	1962	207.3	52500
2007	1360	2627	207.3	46000
2008	995	2012	215.303	45000
2008	1261	625	215.303	70000
2005	1401	175	195.3	55000
2007	660	1113	207.3	40000
2002	1957	2943	179.9	34000
2003	2002	1955	184	29000
2004	1825	1208	188.9	42500
2004	1825	1562	188.9	33000
2005	1825	2127	195.3	32500
2005	1095	2145	195.3	30000
2005	1460	4988	195.3	39000
2006	1825	1880	201.6	28500
2006	1460	1872	201.6	28000
2006	1095	1872	201.6	27000
2006	1095	664	201.6	40000
2007	730	953	207.3	57500
2007	730	517	207.3	72500
2005	1460	1308	195.3	37000
2006	730	993	201.6	47000
2006	1095	1911	201.6	41000
2005	1095	3605	195.3	33000
2006	1460	1923	201.6	32500
2002	1460	4433	179.9	24000
2003	2190	3939	184	29000
2003	2920	3710	184	32500
2005	2190	3790	195.3	33000
2006	1825	9748	201.6	27500
2008	730	4486	215.303	28500
2005	1825	3290	195.3	31000
2008	1095	625	215.303	70000
2008	1095	2012	215.303	45000
2007	1460	2627	207,3	46000
2007	1460	1962	207.3	52500
2005	2190	3290	195.3	31000
2008	730	2012	215.303	45000
2005	2190	1268	195.3	36000
2006	1460	3454	201.6	26000
2002	2920	4433	179.9	24000
2002	2920	4727	179.9	25000
2003	2920	4223	184	35000

price (model ID, the machine calendar age and cumulative number of hours of work of the machine) - have a strong and positive relationship with a value of r = 0.66, which is the simple correlation between the dependent variable (DV) and the independent variable (IV). This value indicates to what extent the value of these variables increases the value of the other variables. Conversely, if the value of these variables decreases, the value of other variables will decrease as well. Thus, the higher the selling price regarding the identification model, the machine calendar age, and the cumulative number of the machine's hours of work, the greater the residual value will be. The correlation coefficient suggests that there is a strong indication that the prediction model works, as attested by the value of R^2 = 0.42. The R^2 value is the amount of variance explained by dependent variables, which are predictors. Thus, it can be said for these data that a detachable percentage of 42% of the variance of dependent variables can be assigned to independent variables.

Table 3. Display of sensitivity of the models created on the basis of dozer D39PX-21 machines

Statistika dozera D39PX-21 koji imaju više od 2000 radnih sati	
R^2 goodness of fit	0.69
Correlation coefficient	0.83
Maximum error	9677.11
Mean squared error	14241549
Mean absolute error	2422.90
Variable sensitivity	
Variable hours sensitivity	0.47
Variable age sensitivity	0.09
Variable CPI sensitivity	0.59
Statistics dozer D39PX-21 machines which have less than 2000 working hours	
R^2 goodness of fit	0.83
Correlation coefficient	0.87
Maximum error	17133.78
Mean squared error	32246154
Mean absolute error	3587.12
Variable sensitivity	
Variable hours sensitivity	2.26
Variable age sensitivity	1.29
Variable CPI sensitivity	0.001

The machine age variable (0.99) showed the highest sensitivity, thus proving to be one of the most important factors in the residual value forecast. The second variable that showed higher sensitivity was the identification model (0.58), followed by the cumulative number of hours of machine work (0.01).

This predictive model was only used for the machines with more than 2000 hours of work, as selected among the 82 machines of the first model. The results of the statistical analysis show that the dependent variable (DV) - residual value - and the independent variable - retail price (identification model, machine calendar age, and cumulative number of machine operating hours) - have a strong positive relationship with a value of r = 0.83, which represents the simple correlation between the dependent variable (DV) and independent variable (IV). This value indicates the extent to which the value of these variables increases the value of other variables. Conversely, if the value of these variables decreases, the value of other variables will also decrease. Thus, the higher the selling price regarding the identification model, the machine calendar age and the cumulative number of hours of machine work, the greater the residual value will be. The correlation coefficient suggests that there is a strong indication that the prediction model works, as attested by the value of R^2 = 0.64. The R^2 value is the amount of variance explained by dependent variables, which are predictors. Thus, it can be said for this data that a detachable amount of 69% of the variance of dependent variables can be assigned to independent variables. Much of the variance of the VD is the result of VI, so it can be said that this model is an excellent predictor.

The variable hours of machine work (0.47) showed the highest sensitivity, proving to be one of the most important factors in the residual value forecast. The second variable that showed higher sensitivity was the identification model (0.59), followed by calendar age of the machine (0.09).

Compared to the first model and according to literature results, this predictive model was better than the first, considering that it included only machines with more than 2000 hours of work. Thus, the result of statistical analysis of these data showed that that the sale price (VD) can more accurately be predicted when the machines had more than 2000 hours of work.

In the predictive model, out of the total of 82 machines, only machines with less than 2000 hours of work were used in the first model and forecast. The results of the statistical analysis show that the dependent variable (RV) - residual value - and the independent variable - Retail price (identification model, machine calendar age and cumulative number of hours of machine work) - have a strong positive relationship with a value of r = 0.87, which represents the simple correlation between the dependent variable (DV) and independent variable (IV). This value indicates the extent to which the value of these variables increases the value of other variables. Conversely, if the value of these variables decreases, the value of other variables will also decrease. Thus, the higher the selling price regarding the identification model, machine calendar age, and cumulative number of hours of machine work, the greater the residual value will be. The correlation coefficient shows that there is a strong indication that the prediction model works, as attested by the value of R^2 = 0.83. The R^2 value is the amount of variance explained by dependent variables, which are predictors. Thus, it can be said for these data that a detachable amount of 83% of the variance of the dependent variable can be assigned to the independent variable.

Variable hours of machine work (2.26) showed the highest sensitivity, thus proving to be one of the most important factors in the residual value forecast. The second variable that showed higher sensitivity was the calendar age of machines (1.29), followed by the identification model (0.001).

For these data, it was the best predictive model, as compared to the previous two models. For machines that are more frequent, it is clear that the residual value can be predicted more accurately for machines with fewer than 2000 working hours, where the statistical forecasting model stood out as the best predictor.

10. Research results and discussion

The creation of three models was presented in previous example, where the dependent variable was the residual value of the dozer D39PX-21, and independent variables were the Model ID, dozer calendar age, and dozer cumulative number of working hours. By applying the same methodology, three models were created using a large sample of construction machines. The research results are presented in Table 4.

Table 4. Research results displaying sensitivity of the three models

Statistics 61.153 machines	
R^2 goodness of fit	0.21
Correlation coefficient	0.51
Maximum error	81756.60
Mean squared error	3.81616e8
Mean absolute error	14105.92
Variable sensitivity	
Variable age sensitivity	0.98
Variable model ID sensitivity	0.86
Variable hours sensitivity	0.85
Statistics dozer 38.182 machines with more than 2000 working hours	
R^2 goodness of fit	0.24
Correlation coefficient	0.52
Maximum error	88535.41
Mean squared error	4.56959e8
Mean absolute error	15847.12
Variable sensitivity	
Variable age sensitivity	1.12
Variable model ID sensitivity	0.79
Variable hours sensitivity	0.58
Statistics 22.861 machines with less than 2000 working hours	
R^2 goodness of fit	0.15
Correlation coefficient	0.84
Maximum error	60796.36
Mean squared error	1.67834e8
Mean absolute error	9027.90
Variable sensitivity	
Variable age sensitivity	3.29
Variable model ID sensitivity	0.24
Variable hours sensitivity	0

In the predictive model, where all 61,153 machines were used in the analysis, it can be seen that the dependent variable (RV) - residual value - and the independent variable - Retail price (model ID, calendar age of machine, and number of cumulative hours of machine work) - correlate strongly and positively with a value of r = 0.51, which is the simple correlation between the dependent variable (DV) and the independent variables (IV). This value points to the extent to which the value of these variables increases the value of other variables. Conversely, if the value of these variables decreases, the value of other variables decreases as well. Thus, the higher the selling price regarding the identification model, machine calendar age, and cumulative number of hours of machine work, the greater the residual value will be. The correlation coefficient shows that there is a strong indication that the prediction model works, as attested by the value of R^2 = 0.21. The R^2 value is the amount of variance explained by dependent variables, which are predictors. Thus, it can be said for these data that 21% of the variance of the dependent variable can be assigned to the independent variable. The machine age variable (0.98) showed the highest sensitivity, thus proving to be one of the most important factors in the residual value forecast. The second variable that showed higher sensitivity was the identification model (0.86), followed by the cumulative number of hours of machine work (0.85).

In the predictive model in which only 38,182 machines (with more than 2000 hours of work) were included in the analysis, the dependent variable (RV) - residual value - and the independent variable - retail price (identification model, machine calendar age, and cumulative number of hours of machine work) - correlate strongly and positively with the value of r = 0.52, which is the simple correlation between the dependent variable (DV) and the independent variable (IV). This value points to the extent to which the value of these variables increases the value of other variables. Conversely, if the value of these variables decreases, the value of other variables decreases as well. Thus, the higher the selling price regarding the identification model, machine calendar age, and cumulative number of hours of machine work, the greater the residual value will be. The correlation coefficient shows that there is a strong indication that the prediction model works, as attested by the value of R^2 = 0.24. The R^2 value is the amount of variance explained by DV through its predictors. Thus, it can be said for these data that 24% of the variance of the dependent variable can be assigned to the independent variable.

The machine age variable (1.12) showed the highest sensitivity, thus proving to be one of the most important factors in the residual value forecast. The second variable that showed higher sensitivity was the identification model (0.79), followed by the cumulative number of hours of machine work (0.58).

Compared to the first model and according to literature results, this predictive model was better than the first, considering that it includes only machines with more than 2000 hours of work. Thus, the result of the statistical analysis of these data showed that the sale price (DV) can be predicted more accurately when the machines have over 2000 hours of work.

In the predictive model in which only 22,861 machines (with less than 2000 hours of work) were included in the analysis, the dependent variable (RV) - residual value - and the independent variable - retail price (identification model, machine calendar age, and cumulative number of hours of machine work) - have a strong positive relationship, represented by a value of r = 0.84, which shows the simple correlation between the dependent variable (DV) and the independent variable (IV). This value points to the extent to which the value of these variables increases the value of other variables. Conversely, if the value of these variables decreases, the value of other variables will decrease as well. Thus, the higher the selling price regarding the identification model, machine calendar age, and cumulative number of hours of machine work, the greater the residual value will be. The correlation coefficient shows that there is a strong indication that the prediction model works, as attested by the value of $R^2 = 0.15$. The R^2 value is the amount of variance explained by DV through its predictors. Thus, for this data, it can be said that 15% of the variance of the dependent variable can be assigned to the independent variable.

The machine age variable (3.29) showed once again the highest sensitivity, thus proving to be one of the most important factors in the residual value forecast. The second variable that showed higher sensitivity was the identification model (0.24). Working hours did not participate in the sensitivity of the model. This result can be explained by the low number of working hours of the machines.

When only machines with less than 2000 hours were used, the forecasting model was worse; this being the case, the model showed lower performance in predicting the dependent variable.

11. Conclusion

This research has shown that predictive models for forecasting residual value of equipment at auctions can be made using symbolic regression, depending on the age of the machines and the number of working hours of the machines,.

This research shows that there is a difference in the results of regression models that take into account the price of used machines that operated for less than 2000 hours, as compared to the machines that operated for more than 2000 hours. The method the auction market uses to determine residual value of newer used machines is dependent on the manufacturer's brand and the machine type. Older machines have a slower depreciation and better fitting of variables (increased fitting) in the symbolic regression model.

Only a rough preliminary prediction of residual value can be made when thousands of machines with a limited number of criteria are included in the data set. In practice, construction companies can make a preliminary assessment of residual value of machines sold or purchased under circumstances when very little information is available about the state of such machines.

Research results tend to improve the decision-making process and help in areas of target costing, as well as in the process of determining the budget for future projects, as it will now be easier to predict the value of heavy equipment that the manager has to sell. This research points to the significance of analytics, predictive models, and innovations in the improvement of current business processes.

REFERENCES

[1] Lucko, G.: A statistical analysis and model of the residual value of different types of heavy construction equipment, Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA, 2003.

[2] Lucko, G.: Predicting the residual value of heavy construction equipment. Proc., 4th Joint Int. Symp. on Information Technology in Civil Engineering, pp. 1-11, 2003.

[3] Mitchell, Z.J.: A statistical analysis of construction equipment repair costs using field data and the cumulative cost model, Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA., 1998.

[4] Perry, G.M.: Durable asset depreciation: A reconciliation between hypotheses, American Journal of Agricultural Economics, 72 (1990), pp. 524-529.

[5] Grinyer, P.H.: The Effects of Technological Change on the Economic Life of Capital Equipment, AHE Transactions, 5 (1973) 3, pp. 203-213.

[6] Peurifoy, R.L.: Construction planning, equipment, and methods. McGraw-Hill, New York., 1996.

[7] Cowles, H.A.: Valuation of Industrial Property: A Proposed Model. The Engineering Economist, 23 (1978), pp. 141-161.

[8] Vorster, M.C.: Association of Construction Equipment Managers _ACEM_ development program at Virginia Tech, Virginia Polytechnic Institute and State University, Blacksburg, Va., 1995.

[9] DeGarmo, E., Sullivan, W., Bontadelli, J.: Engineering economy, 9th Ed., Macmillan, New York., 1993.

[10] Douglas, G., Popescu, C., Richard R.: Construction Equipment Management for Engineers, Estimators and Owners, 2005.

[11] Vorster, M.: Consequential equipment costs associated with lack of availability and downtime. Journal Construction Engineering Management., 116 (1996) 4, pp. 656-669.

[12] Vorster, M.: A systems approach to the management of civil engineering construction equipment, Ph.D. dissertation, Univ. of Stellenbosch, Stellenbosch, Western Cape, South Africa, 1980.

[13] Vorster, M.: Construction equipment economics, 1st Ed. Pen Publications, Organizing for equipment management, pp. 828, 2009.

[14] Vorster, M.: Construction equipment economics, Pen, Christiansburg, VA., 2009.

[15] Kannan, G.: A methodology for the development of a production experience database for earthmoving operations using automated data collection, Ph.D. dissertation, Virginia Polytechnic Institute and State Univ., Blacksburg, VA., 1999.

[16] Hildreth, J.C.: The use of short-interval GPS data in construction operations analysis, Ph.D. dissertation, Virginia Polytechnic Institute and State Univ., Blacksburg, VA., 2003.

[17] Martinez, C.: STROBOSCOPE: State and resource based simulation of construction processe, Ph.D. thesis,. Univ. of Michigan, Ann Arbor, MI., 2011.

[18] Kastens, T.H.: An owning and operating cost calculator model for construction equipment, Blacksburg, Va., 2002.

[19] Werblow, D.: Forest Harvesting Equipment Ownership and Operating Costs., Southern Journal of Applied Forestry ,10 (1986) 1, pp. 10-15.

[20] Cross, T., Perry, G.: Depreciation patterns for agricultural machinery, American Journal of Agricultural Economics, 77 (1995) 1, pp. 194-204.

[21] Lucko, G., Anderson-Cook, C., Vorster, M.: Unknown element of owning costs-Impact of residual value, Journal of Construction Engineering and Management, 133 (2007) 1, pp. 3-9.

[22] Lucko, G., Mitchell, Z.: Quantitative research: Preparation of incongruous economic data sets for archival data analysis, Journal of Construction Engineering and Management, 136 (2010) 1, pp. 49-57.

[23] Schmidt, M., Hod, L.: Distilling free-form natural laws from experimental data. Science (AAAS) 324 (2009) 5923, pp. 81-85.

[24] Milajić, A., Beljakovic, D., Petronijevic, P.: Genetic algorithms for assigning tasks to construction machine operators, GRADEVINAR, 63 (2011), 8, pp. 749-755.

[25] Rayid G., Hillery, S.: Predicting the End-Price of Online Auction, 2004.

[26] Bajari, P., Hortacsu, A.: Reserve Prices and Endogenous Entry: Empirical Insights from Ebay Auctions, The Rand Journal of Economics, pp. 329-355, 2002.

[27] Nutonian, http://blog.nutonian.com/bid/330675/How-does-Eureqa-Compare-to-Other-Machine-Learning-Methods