Predicting Audit Opinion by a new Metaheuristic Algorithm: Water Cycle Algorithm

Hoda Eskandar*

*Assistant Prof., Department of Accounting, Management and Accounting faculty, University of Allameh Tabataba'i, Tehran, Iran, *Corresponding author: heskandar@ut.ac.ir*

Mohammad Moradi

Associate Prof., Department of Accounting, Management faculty, University of Tehran, Tehran, Iran, moradimt@ut.ac.ir

Hassan Yazdifar

Prof. of Accounting & Management, Derby University, UK,

H.Yazdifar@derby.ac.uk

Aziz Seyedi

PHD., Department of Accounting, Management faculty, University of Tehran, Tehran, Iran, saccounting1980@yahoo.com

Hadi Eskandar

M.S., Faculty of mechanical Engineering, University of semnan, semnan, iran,

hadi.eskandar@yahoo.com

Abstract

An auditor evaluates if financial statements which the firms issue in public, present fairly and are free from material misstatement. The audit report is a written letter containing independent verification of the quality of financial statements used for making economic decisions. Hence, the issuance of such a report can lead to the transmission news and information about the firm and to enhance the degree of confidence in the financial statements.

This study predicts audit opinion of the firms listed in Tehran Stock Exchange during 2018-2020 by a new metaheuristic algorithm named Water Cycle Algorithm (WCA) and compares its results with one of the most popular methods called logistic regression (LG). 24 variables were extracted from the literature and used for this prediction. 4 evaluating criteria were used to compare the predictions of two methods.

According to findings, the superiority of the criteria in the WCA was confirmed in comparison to LG. Since WCA was more appropriate, users of financial reports can use it to predict the type of audit opinion in the unaudited interim financial statements, and also, auditors can use it while evaluating and accepting clients and achieving an acceptable level of audit risk, as a quality control tool.

Key Words: Audit Opinion, Water Cycle Algorithm, Logistic Regression.

Introduction

Companies produce financial statements that provide information about their financial position and performance. This information is used by a wide range of stakeholders in making economic decisions. Therefore, reliability of such statements and information is a vital issue because if they are prepared accurately, users are highly likely to make inappropriate decisions. To enhance the degree of confidence in the financial statements, a qualified external party (an auditor) is engaged to examine this information, to give their professional opinion on whether they fairly reflect, in all material respects, the company's financial performance and financial position. In other words, the auditor verifies that the financial reports and records offer an accurate picture of the company. It means that audited financial statements can be perceived as reliable and readily available sources of information, because such financial statements are provided to their users along with opinions from auditors or other independent professionals; such opinions may add to the reliability of the financial statements, so that the users can be more confident about the decisions they make on the basis of such financial statements (Karami, Karimiyan, Saba Salati, 2017).

Audit report refers to the written document of the audit opinion issued by the certified public accountant (CPA) on the financial statements of the audited entity based on performing the audit work in accordance with the provisions of the audit standards. Two types of opinions are obtained from the accounting audit process: unqualified opinion and qualified opinions. The first is the one in which the auditor does not detect significant differences between the figures presented by the company's administrators and those verified by the auditor. The second opinion shows significant differences that were not corrected by the company's administrators and that the auditor reveals in his audit report (Sánchez-Serrano, et al, 2020).

Predicting audit opinion has information content. Such a prediction gives a helping hand to audit firms, aiming to make some decisions such as assessing audit risk, accepting clients and determining audit fee based on their risk. In addition, it helps auditors to plan revision procedures and control their performances (Sánchez-Serrano, et, al. (2020). Hence, recently, researchers have shown strong tendency to do some research, hoping to predict audit opinion type. Some authors have contributed to the development of models that help to predict the opinion on audit reports. In this regard, previous studies have applied various methodologies in the search to adjust a model with better predictions. One of the most popular methods to predict binary variables (variables with only two values) is logistic regression (LG) (e.g., Moalla, et al, 2017, Yasar, Yakut, Gutnu, 2015, Spathis, Doumpos, Zopounidis, 2003, Laitinen and Laitinen, 1998).

The purpose of this study is to predict audit opinion by some variables extracted from the literature and at last, evaluate the efficiency of such a prediction. This study seeks to predict audit opinion by a new metaheuristic algorithm named water cycle algorithm (WCA) and compare its results with one of the most popular methods named logistic algorithm (LG).

Literature Review

As it was mentioned before, recently some researchers have tried to predict audit opinion and so, there are a few papers related to such prediction. Audit opinion prediction models in general, the role played by the auditor and its effect on the markets gives it a dual role: on the one hand, an informative role and, on the other hand, a role for information security and reliability. The auditor provides independent verification of the financial statements (DeAngelo, 1981).

To predict the audit opinion, previous studies have applied various methodologies in the search to adjust a model with better predictions. Early scholars usually use statistical analysis methods to study audit risk early warning of companies. One of the most popular of such methods is logistic algorithm (LG).

Laitinen and Laitinen (1998) applied a logistic model based on investigated financial ratios to identify the audit opinion. They analysed 37 firms listed on the Helsinki Stock Exchange using a set of 17 explanatory variables. They confirmed that qualified opinions show a greater association with low profitability, low growth, and high indebtedness, but the accuracy of their model was only 62%. Spathis, Doumpos, Zopounidis (2003) and Moalla, et al, (2017) considered a set of economic variables, financial ratios, and non-financial variables and used LG for predicting audit opinion among 100 Greek firms. The results suggested that the most predictive variables in the model were collection/sales, sales/total assets, net profit/total assets, and working capital/total assets. Saaydah (2019), Susanto and Pradipta (2017) foundout the relationship between corporate governance mechanisms and audit opinion by LG. Dopuch, W. Holthausen and W. Leftwich (1987) predicted audit qualifications with some financial and market variables.

Other authors utilized other methods in such a prediction and found the following evidence:

Yasar, Yakut, Gutnu, (2015) used discriminant analysis, logit, and decision trees to predict the audit opinion of a sample of companies from the Istanbul Stock Exchange. They found that some retios like profitability and bebt ones are strong predictors of audit opinion.

Pourheydari and Azami (2011) predicted audit opinion by a neural networks approach during 2003 to 2009. The input variables composed of financial ones such as financial distress and non-financial ones such as firm litigation.

Setayesh, et, al. (2015) forecasted audit opinions by data mining during 2001 to 2010. Predicting variables included liquidity, profitability, leverage, efficiency, size, cash flow.

Heng-Shu (2017) used some the financial indicators as variables and introduced Takagi-Sugeno fuzzy neural network to construct the prediction model of audit opinions.

Sánchez-Serrano, et, al. (2020) predicted audit opinion in consolidated financial statements with artificial neural networks in spain. They found that besides some financial ratios (current and quick ratio, operating and investing cash flow), the variables referring to size, auditor, and board members were converted into the main explanatory parameters of the prediction.

Zeng, Li and Li (2022) predicted audit opinion by Sparse Principal Component Analysis and Kernel Fuzzy Clustering Algorithm.

Since traditional research methods are limited by strict assumptions and have poor fault tolerance, other methods especially metaheuristic ones are used. Recently, metaheuristic algorithms especially one of the new ones named water cycle algorithm (WCA) has been validated and implemented for solving financial problems (Moradi, et al, 2017). The WCA is based on the observation of water cycle process and how rivers and streams flow into downhill towards the sea in nature (Eskandar et.al, 2012). It was first introduced by Eskandar et al. (2012) for solving engineering optimization problems. Recently, Moradi et al. (2017) used this method in financial field. They utilized it for optimizing portfolio selection. Their findings showed that this method is more efficient than genetic algorithm and particular swarm algorithm.

Since the efficiency of such algorithm has approved in engineering problems and solving portfolio selection problem, this study seeks to examine its performance in predicting audit opinion in comparison with LG.

According to above, the hyposesis are as the following:

1. *“WCA is appropriate for predicting audit opinion.”*
2. *“WCA is more efficient than LG regression in predicting audit opinion.”*

Methodology

The population consists of all of the firms listed on the TSE. The sample was also selected through a systematic removal method from the statistical population with considering the following criteria:

Firms listed in TSE from 2018 to 2020, excluding financial ones and those ones whose data is not accessible.

At last, the sample includes 237 firms during 3 years (711 observations). We collected their data from annual reports and from TSE reports obtained from electronic data and the Internet.

Dependent variable of this study is audit opinion. It is a dummy variable that is 1 when audit opinion issues an unqualified report and otherwise, it is 0. Moreover, independent variables include 24 predicting variables (recognized based on prior literature). They are shown in the following table:

Table 1. Research Variables

Title	**Variable**	**Measurement**	**Symbol**	**Reference**
Liqidity	Current Ratio	Current assets/current liabilities	x1	Sánchez-Serrano, et, al. (2020) Pourheydari and Azami (2011)
	Quick Ratio	Current assets (excluding inventory and prepaids)/ current liabilities	x2	
Asset management	Inventory Turnover	Net sale/inventory average	x3	Heng-Shu (2017) Pourheydari and Azami (2011)
	Asset Turnover	Net sale/ assets average	x4	Spathis, Doumpos, Zopounidis (2003) Pourheydari and Azami (2011)
	Receivables Turnover	Net sale/ receivables average	x5	Heng-Shu (2017) Pourheydari and Azami (2011)
	Return on Asset	Net income/assets	x6	Yasar,

profitability	Return on Investment	Net income / investment	x7	Yakut, Gutnu, (2015) Laitinen and Laitinen (1998) Pourheydari and Azami (2011)
	Return on Shareholders Equity	Net income/ stockholders' equity	x8	
	Net Income Ratio	Net income/ sale	x9	
Cash flows	Operating Cash ratio	Operating cash flow/sale	x10	Sánchez-Serrano, et, al. (2020) Pourheydari and Azami (2011)
	Investing Cash Ratio	Investing cash flow/sale	x11	
Debt management	Debt Ratio	Liabilities/ assets	x12	Yasar, Yakut, Gutnu, (2015) Laitinen and Laitinen (1998) Pourheydari and Azami (2011)
Market value	Market to Book Value	Market value/book value	x13	Dopuch, W. Holthausen and W. Leftwich (1987)
stock	Stock Return	Dividend/ stock price	x14	
	Price to Revenue	Stock price/ EPS	x15	
growth	Firm Growth	(Assetst-Assets$_{t-1}$)/Assets $_{t-1}$	x16	Laitinen and Laitinen (1998) Setayesh, et, al. (2015)
size	Log Net Sales	Log net sale	x17	
Corporate governance	Audit Firm Type	If audit organization is the firm auditor, 1 otherwise 0	x18	Saaydah (2019) Susanto and Pradipta (2017) Sánchez-Serrano, et, al. (2020)
	Prior AuditOpinion	If prior audit opinion is unqualified 1, otherwise 0	x19	
	Auditor Switch	If auditor switched to audit organization or the reverse, 1 otherwise 0	x20	
	Management Switch	If CEO of the members of director board are switched 1, otherwise 0	x21	
	Auditor Tenure	If auditor is not switched during 2 period 1, otherwise 0	x22	
	Nomber of Board of Directors Members	Log the number of director board	x23	
others	Firm Age	Firm age	x24	Zeng, Li and Li (2022) Setayesh, et, al. (2015)

In this study, variables are computed by Excel and WCA is run by Matlab software.

WCA

The WCA mimics the flow of rivers and streams toward the sea and derived by the observation of water cycle process. Let us assume that there are some rain or precipitation phenomena. An initial population of design variables (i.e., population of streams) is randomly generated after raining process. The best individual (i.e., the best stream), classified in terms of having the minimum cost function (for minimization problem), is chosen as the sea.

Then, a number of good streams (i.e., cost function values close to the current best record) are chosen as rivers, while the other streams flow into the rivers and sea. In an D dimensional optimization problem, a stream is an array of $1\times D$. Starting an optimization algorithm, an initial population representing a matrix of streams of size $N_{pop}\times D$ is generated. Hence, the matrix of initial population, which is generated randomly, is given as (rows and column are the number of population and the number of design variables, respectively):

$$Total \quad Population = \begin{bmatrix} Sea \\ River_1 \\ River_2 \\ River_3 \\ \vdots \\ Stream_{Nsr+1} \\ Stream_{Nsr+2} \\ Stream_{Nsr+3} \\ \vdots \\ Stream_{N_{pop}} \end{bmatrix} = \begin{bmatrix} x_1^1 & x_2^1 & x_3^1 & \cdots & x_D^1 \\ x_1^2 & x_2^2 & x_3^2 & \cdots & x_D^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_1^{N_{pop}} & x_2^{N_{pop}} & x_3^{N_{pop}} & \cdots & x_D^{N_{pop}} \end{bmatrix},$$

Where N_{pop} and D are population size and the number of design variables, respectively. Each of the decision variable values ($x_1, x_2, \ldots, x_D$) can be represented as floating point number (real values) or as a predefined set for continuous and discrete problems, respectively. The cost of a stream is obtained by the evaluation of cost function (fitness function).

At the first step, N_{pop} streams are created. A number of N_{sr} from the best individuals (minimum values) are selected as a sea and rivers. The stream which has the minimum value among others is considered as the sea. In fact, N_{sr} is the summation of number of rivers (which is defined by user) and a single sea. The rest of the population (i.e., streams flow into the rivers or may directly flow to the sea) are considered as streams.

Depending on magnitude of flow, each river absorbs water from streams. The amount of water entering a river and/or the sea, hence, varies from stream to stream. In addition, rivers flow to the sea which is the most downhill location. The designated streams for each rivers and sea are calculated using the following equation (Eskandar et al. 2012):

$$C_n = Cost_n - Cost_{Nsr+1} \qquad n = 1, 2, 3, \ldots, N_{sr} \ ,$$

$$NS_n = round\left\{\left|\frac{C_n}{\sum_{n=1}^{N_{sr}} C_n}\right| \times N_{Streams}\right\} \quad , \quad n = 1, 2, \ldots, N_{sr} \ ,$$

where NS_n is the number of streams which flow to the specific rivers and sea. As it happens in nature, streams are created from the raindrops and join each other to generate new rivers. Some stream may even flow directly to the sea. All rivers and streams end up in the sea that corresponds to the current best solution. Let us assume that there are N_{pop} streams of which N_{sr}-1 are selected as rivers and one is selected as the sea.

At the exploitation phase in the WCA, new positions for streams and rivers have been suggested as follows (Eskandar et al. 2012):

$$\vec{X}^{t+1}_{Stream} = \vec{X}^{t}_{Stream} + rand \times C \times (\vec{X}^{t}_{Sea} - \vec{X}^{t}_{Stream}) \qquad \vec{X}^{t+1}_{Stream} = \vec{X}^{t}_{Stream} + rand \times C \times (\vec{X}^{t}_{River} - \vec{X}^{t}_{Stream})$$

$$\vec{X}^{t+1}_{River} = \vec{X}^{t}_{River} + rand \times C \times (\vec{X}^{t}_{Sea} - \vec{X}^{t}_{River})$$

Where $1 < C < 2$ and the best value for C may be chosen as 2 and *rand* is an uniformly distributed random number between zero and one. Eqs. (16) and (17) are for streams which flow into the sea and their corresponding rivers, respectively. Notations having vector sign correspond to vector values, otherwise the rest of notations and parameters are considered as scalar values. If the solution given by a stream is better than its connecting river, the positions of river and stream are exchanged (i.e., the stream becomes a river and the river becomes a stream). A similar exchange can be performed for a river and the sea.
The evaporation process operator also is introduced to avoid premature (immature) convergence to local optima (exploitation phase) (Sadollah et al. 2015). Basically, evaporation causes sea water to evaporate as rivers/streams flow into the sea. This leads to new precipitations. Therefore, we have to check if the river/stream is sufficiently close to the sea to make the evaporation process occur. For that purpose, the following criterion is utilized for evaporation condition (Eskandar et al. 2012):

$$if \ \left\| \vec{X}^{t}_{Sea} - \vec{X}^{t}_{River_j} \right\| < d_{\max} \quad or \quad rand < 0.1 \qquad j = 1, 2, 3, ..., N_{sr} - 1$$

$$\quad Perform \ \ raining \ \ process \ \ by \ \ unifrom \ \ random \ \ search \ ,$$

$$end$$

where d_{max} is a small number close to zero. After evaporation, the raining process is applied and new streams are formed in the different locations (similar to mutation in the GAs). Hence, in the new generated sub-population, the best stream will act as a new river and other streams move toward their new river. This condition will also apply for streams that directly flow to the sea.

Similarly, the best newly formed stream is considered as a river flowing to the sea. The rest of new streams are assumed to flow into the rivers or may directly flow into the sea. The following equation is used only for the streams which directly flow to the sea. It encourages the creation of streams which directly flow to the sea in order to improve the exploration near the sea (the optimum solution) in the feasible region for constrained problems (Eskandar et al, 2012):

$$\vec{X}^{t+1}_{Stream} = \vec{X}^{t}_{Sea} + \sqrt{\mu} \times randn\,(1, D)$$

Where μ is a coefficient which shows the range of searching region near the sea, *randn* is the normally distributed random number. The larger μ increases the possibility to exit from feasible region. The smaller μ leads the algorithm to search in smaller region near the sea. Its suitable value is set to 0.1. Indeed, term $\sqrt{\mu}$ represents the standard deviation. The generated individuals with variance μ are distributed around the best obtained optimum point (sea).

Therefore, the evaporation operator is responsible for the exploration phase in the WCA.

A large value for d_{max} prevents extra searches and small values encourage the search intensity near the sea. Therefore, d_{max} controls the search intensity near the sea (i.e., best obtained solution). The value of d_{max} adaptively decreases as follows:

$$d_{\max}^{t+1} = d_{\max}^{t} - \frac{d_{\max}^{t}}{Max\_Iteration} \qquad t = 1,2,3,\ldots, Max.\ \ Iteration$$

Where t is an iteration index.

The detailed steps of WCA are described as follows (Sadollah et al. 2015):

Step 1: Choose the initial parameters of the MOWCA: N_{sr}, d_{max}, N_{pop}, maximum iteration number, and Pareto archive size.

Step 2: Create random initial population and form the initial streams, rivers, and sea.

Step 3: Create the value of multi-objective functions for each stream.

Step 4: Determine the non-dominated solutions in the initial population and save them in the Pareto archive.

Step 5: Determine the non-dominated solutions among the feasible solutions and save them in the Pareto archive.

Step 6: Calculate the crowding-distance for each Pareto archive member.

Step 7: Select a sea and rivers based on the crowding-distance value.

Step 8: Determine flow intensity of rivers and sea based on the crowding distance values.

Step 9: Some streams may directly flow into the sea.

Step 10: Exchange positions of sea with a stream which gives the best solution.

Step 11: Streams flow into the rivers.

Step 12: Exchange positions of river with a stream which gives the best solution.

Step 13: Rivers flow into the sea.

Step 14: Exchange positions of sea with a river which gives the best solution.

Step 15: Check the evaporation condition using the pseudo-code.

Step 16: The raining process will occur if the evaporation condition is satisfied.

Step 17: Reduce the value of d_{max} which is a user defined parameter.

Step 18: Determine the new feasible solutions in the population.

Step 19: Determine the new non-dominated solutions among the feasible solutions and save them in the Pareto archive.

Step 20: Eliminate any dominated solutions in the Pareto archive.

Step 21: Go to the Step 22 if the number of member in the Pareto archive is more than the determined Pareto archive sizes, other-wise, go to the Step 23.

Step 22: Calculate the crowding-distance value for each Pareto archive member and remove as many members as necessary with the lowest crowding-distance value.

Step 23: Calculate the crowding-distance value for each Pareto archive member to select new sea and rivers.

Step 24: Check the convergence criterion. The WCA will be stopped if the stopping criterion is satisfied, otherwise return to the Step 9.

Table 2 provides the pseudocode of WCA algorithm.

Table 2. Pseudo-code of the WCA

• Set WCA user parameter: N_{pop}, N_{sr}, and the maximum number of iterations.
• Determine the number of streams (individuals) which flow to the rivers and sea.

```
• Randomly create initial population in upper and lower bounds of a given
problem.
• Choose the sea, rivers, and streams within the current initial population.
• Define the intensity of flow
while (t ≤ Maximum_Iteration) or (any defined stopping condition)
    for i = 1 : Population Size (N_pop)
        Streams directly flow to the sea
        Calculate the objective function of the generated stream
            if Cost (New_Stream) < Cost (Sea)
                Sea = New_Stream;
            end if
        Streams flow to their corresponding rivers using
        Calculate the objective function of the generated stream
            if Cost (New_Stream) < Cost (River)
                River = New_Stream;
                    if Cost (New_Stream) < Cost (Sea)
                        Sea = New_Stream;
                    end if
            end if
        Rivers flow to the sea
        Calculate the objective function of the generated river
            if Cost (New_River) < Cost (Sea)
                Sea = New_River;
            end if
    Check the evaporation condition
end while
Post process results and visualization
```

Logistic Regression (LG)

logistic model (or logit model) is a statistical model that models the probability of an event taking place by having the log-odds for the event be a linear combination of one or more independent variables. In regression analysis, logistic regression[1] (or logit regression) is estimating the parameters of a logistic model (the coefficients in the linear combination). Formally, in binary logistic regression there is a single binary dependent variable, coded by an indicator variable, where the two values are labeled "0" and "1", while the independent variables can each be a binary variable (two classes, coded by an indicator variable) or a continuous variable (any real value). The corresponding probability of the value labeled "1" can vary between 0 (certainly the value "0") and 1 (certainly the value "1"), hence the labeling;[2] the function that converts log-odds to probability is the logistic function, hence the name. The unit of measurement for the log-odds scale is called a *logit*, from *logistic unit*, hence the alternative names. See § Background and § Definition for formal mathematics, and § Example for a worked example.

LG is used for predicting some events in various fields, including machine learning, most medical fields, and social science.

Evaluating Criteria

For evaluating the results of the methos, real outcome is compared with predicted outcome as the following table. This table is a confusion matrix:

Prediction group

	negative	positive	Real group
Sensitivity $\frac{TP}{(TP + FN)}$	False Negative (FN) β error	True Positive (TP)	positive
Specifity $\frac{TN}{(TN + FP)}$	(TN) True Negative	False Positive (FP) α error	negative
Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$	Negative Prediction Value $\frac{TN}{(TN + FN)}$	Percision $\frac{TP}{(TP + FP)}$	

Table3. Confusion Matrix

There are many ways to measure how well a statistical model predicts a binary outcome. Four very common measures are precision, accuracy, sensitivity, and specificity.

One simple way of measuring Accuracy is simply the proportion of individuals who were correctly classified–the proportions of True Positives and True Negatives. In other words, Accuracy means that how well the model predicts the output.

$$Accuracy= (TP+TN)/(TP+TN+FP+FN)$$

Precision means that when the model predicts a positive outcome, how much this outcome can be right and appropriate.

$$Precision= TP/(TP+ FP)$$

Sentivity is the rate of accurate positive outcome and specifity is the rate of accurate negative outcome.

$$Sensitivity= TP/(TP+ FN)$$
$$Specifity= TN/(TN+FP)$$

Fig. 1 demonstrates the steps of this study:

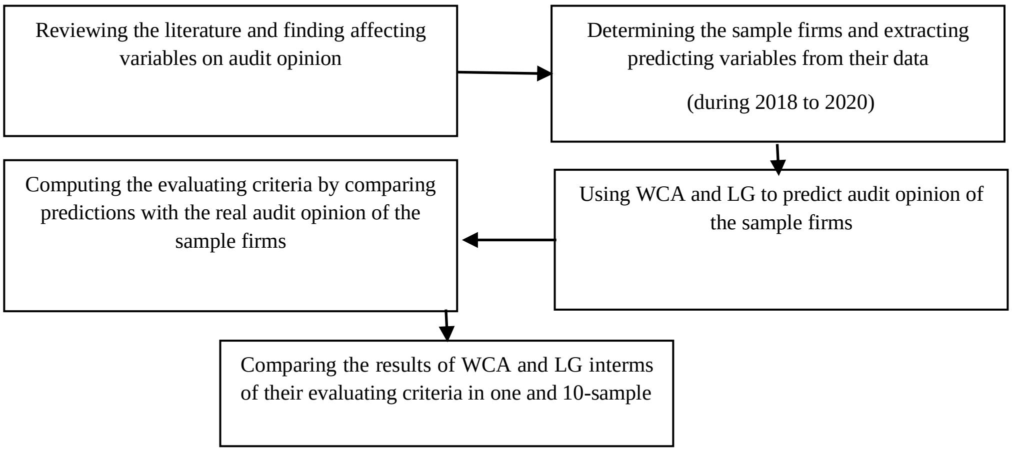

Fig1. The steps of the study

Results

Descriptive Statics

Table 4 demonstrates desciptive statistics of research variables.

Table 4. Descriptive Statistics

Variable	Symbol	Max	Min	Mean	S.D
Current Ratio	x1	6/17	0/39	1/53	1/01
Quick Ratio	x2	5/55	0/18	1/01	0/78
Inventory Turnover	x3	1730/70	3/09	167/38	210/72
Asset Turnover	x4	3/97	0/05	0/87	0/63
Receivables Turnover	x5	27/06	0/02	4/34	4/26
Return on Asset	x6	421/62	-41/37	35/64	72/05
Return on Investment	x7	326/83	-57/11	45/03	62/52
Return on Shareholders Equity	x8	70/55	-46/90	19/88	22/26
Net Income Ratio	x9	87/66	-33/15	9/03	17/59
Operating Cash ratio	x10	1/57	-5/35	0/06	0/34
Investing Cash Ratio	x11	0/14	-0/69	-0/07	0/12
Debt Ratio	x12	1/13	0/13	0/57	0/20
Market to Book Value	x13	61/39	1/01	7/77	9/99
Stock Return	x14	0/24	-0/33	0/05	0/08
Price to Revenue	x15	2105/45	-33/08	67/40	234/51
Firm Growth	x16	3/51	-0/20	0/19	0/47
Log Net Sales	x17	8/23	3/98	6/24	0/76
Audit Firm Type	x18	1/00	0/00	0/18	0/38
Prior Audit Opinion	x19	1/00	0/00	0/48	0/50
Auditor Switch	x20	1/00	0/00	0/25	0/43
Management Switch	x21	1/00	0/00	0/33	0/47
Auditor Tenure	x22	1/00	0/00	0/58	0/49
Nomber of Board of Directors Members	x23	0/85	0/48	0/70	0/02
Firm Age	x24	68/94	10/79	40/68	14/35

According to table 5, by using WCA, 711 observations (years-firms) are classified in 4 models (R1toR4).

In model R1, prior audit opinion has changed from maximum and minimum initial renge and so, is recognized as only independent variable and other are not recognized as independent variable. In model R2, inventory turnover, assets turnover, return on investment, net income ratio, stock return, prior audit opinion and the number of the board of directors have changed and they also are recognized as independent variable. If these variables are in the reange of (3/09-1100/47), (0/07-3/73), (-6/36-421), (33/15-87/66), (0/24-0/31), (1-1) and (0/70-0/85), it is possible to predict audit opinion with 76% precision and 70% accuracy. In this way, the number of independent variables in model R3 and R4 are 20 and 19 independent variables and the precision and accuracy of prediction are 28%, 78% and 100% and 78% respectively.

Table 5

Variable	R4		R3		R2		R1		symbol
	18		7		322		364		No
Current Ratio	1/70	0/39	2/59	0/51	6/17	0/39	6/17	0/39	x1
Quick Ratio	1/67	0/19	1/67	0/20	5/55	0/18	5/55	0/18	x2
Inventory Turnover	383/53	3/09	383/53	3/09	1100/47	3/09	1730/70	3/09	x3
Asset Turnover	3/97	0/18	3/97	0/18	3/73	0/07	3/97	0/05	x4
Receivables Turnover	27/06	0/63	27/06	0/63	27/06	0/02	27/06	0/02	x5
Return on Asset	31/27	-16/75	40/40	-16/75	46/39	-16/75	46/39	-16/75	x6
Return on Investment	200/75	-46/90	421/62	-30/79	421/62	-6/36	421/62	-57/	x7

								11	
Return on Shareholders Equity	78/76	-37/40	78/76	-37/40	78/76	-37/40	78/76	-37/40	x8
Net Income Ratio	6/19	-33/15	9/88	-33/15	87/66	-33/15	87/66	-46/90	x9
Operating Cash ratio	0/44	-0/12	0/77	-0/12	0/77	-0/31	0/77	-0/31	x10
Investing Cash Ratio	0/00	-0/04	0/09	-0/09	0/14	-0/69	0/14	-0/69	x11
Debt Ratio	1/13	0/22	1/13	0/22	1/13	0/13	1/13	0/13	x12
Market to Book Value	26/18	1/01	42/01	1/01	61/39	1/01	61/39	1/01	x13
Stock Return	0/05	-0/31	0/11	-0/29	0/24	-0/31	0/24	-0/33	x14
Price to Revenue	109/55	-33/08	109/16	-33/08	205/45	-33/08	2105/45	-33/08	x15
Firm Growth	1/11	-0/17	3/51	-0/17	3/51	-0/20	3/51	-0/20	x16
Log Net Sales	7/67	4/41	8/23	4/41	8/23	3/98	8/23	3/98	x17
Audit Firm Type	1/00	0/00	1/00	0/00	1/00	0/00	1/00	0/00	x18
Prior Audit Opinion	1/00	1/00	1/00	1/00	1/00	1/00	0/00	0/00	x19
Auditor Switch	1/00	0/00	1/00	0/00	1/00	0/00	1/00	0/00	x20
Management Switch	1/00	0/00	1/00	0/00	1/00	0/00	1/00	0/00	x21
Auditor Tenure	1/00	0/00	1/00	0/06	1/00	0/00	1/00	0/00	x22
Nomber of Board of Directors Members	0/70	0/70	0/70	0/70	0/85	0/70	0/85	0/48	x23
Firm Age	67/12	10/84	67/04	10/84	68/94	10/84	68/94	10/84	x24
group	0		1		1		0		
Audit opinion	qualified	unqualified	qualified	unqualified	qualified	unqualified	qualified	unqualified	
Prediction group	4	14	0	7	78	244	109	255	
Accuracy	0/78		0/28		0/70		0/36		
Percision	0/78		1/00		0/76		0/70		

Based on table 6, the number of the firms with qualified report which predicted accurately is 244 plus 7 (251 firms). The number of the firms with unqualified report but predicted unaccurately equals 109 plus 4 (113 firms). The number of the firms with qualified report and predicted accurately equals 255 plus 14 (269firms). At last, 78 firms with unqualified report predicted unaccurately (78+0). Aacording to tha table, precision, accuracy, sentivity and specifity criteria equals 76, 73, 69 and 78 percent, respectively.

Table 6

Prediction group

	0	1

Sensitivity 0/690	FN 113	TP 251	1	Real group
Specifity 0/775	TN 269	FP 78	0	
Accuracy 0/731	Negative Prediction Value 0/704	Percision 0/763		

In this section, audit opinion is predicted by WCA and GA for 10-times, aimimg to compare the results of these method. Table 7 shows the results in 10 samples.

Table 7

Specifity	Sensitivity	Percision	Accuracy	FN	TN	FP	TP	Sample
0/749	0/695	0/744	0/722	111	260	87	253	1
0/746	0/692	0/741	0/719	112	259	88	252	2
0/755	0/690	0/747	0/722	113	262	85	251	3
0/738	0/701	0/737	0/719	109	256	91	255	4
0/735	0/701	0/735	0/717	109	255	92	255	5
0/746	0/695	0/742	0/720	111	259	88	253	6
0/744	0/687	0/737	0/714	114	258	89	250	7
0/749	0/692	0/743	0/720	112	260	87	252	8
0/778	0/654	0/756	0/714	126	270	77	238	9
0/735	0/701	0/735	0/717	109	255	92	255	10
0/778	0/701	0/756	0/722			**best**		
0/735	0/654	0/735	0/714			**worst**		
0/013	0/014	0/006	0/003			**S.D**		
0/748	0/691	0/742	0/718			**Mean**		

According to table 8, current ratio, inventory turnover, prior audit opinion have changed from maximum and minimum initial range and so, are recognized as independent variables and other are not recognized as independent variable.

Table8

Variable	**range**		**Symbol**
Current Ratio	6/17	0/39	x1
Quick Ratio	5/55	0/18	x2
Inventory Turnover	1730/70	3/10	x3
Asset Turnover	3/97	0/05	x4
Receivables Turnover	27/06	0/02	x5
Return on Asset	46/39	-16/75	x6

Return on Investment	421/62	-57/11	x7
Return on Shareholders Equity	78/76	-37/40	x8
Net Income Ratio	87/66	-46/90	x9
Operating Cash ratio	0/77	-0/31	x10
Investing Cash Ratio	0/14	-0/69	x11
Debt Ratio	1/13	0/13	x12
Market to Book Value	61/39	1/01	x13
Stock Return	0/24	-0/33	x14
Price to Revenue	2105/45	-33/08	x15
Firm Growth	3/51	-0/20	x16
Log Net Sales	8/23	3/98	x17
Audit Firm Type	1/00	0/00	x18
Prior Audit Opinion	1/00	1/00	x19
Auditor Switch	1/00	0/00	x20
Management Switch	1/00	0/00	x21
Auditor Tenure	1/00	0/00	x22
Nomber of Board of Directors Members	0/85	0/48	x23
Firm Age	68/94	10/84	x24

According to table 9, the number of the firms which received unqualified audit opinion (1) and were predicted accurately are 253. Audit opinion of 111 firms was unqualified (1) but were not predicted accurately (0). The audit opinion of 260 firms was qualified (0) and they were predicted accurately. The audit opinion of 87 firms was qualified, However, their prediction was unaccrate.

Precision, accuracy, sentivity and specifity of the final model is 74, 72, 69 and 75% respectively. Therefore, if current ratio, inventory turnover and prior audit opinion are in the range of (0/39-6/17), (3/10, 1730/70) and (1-1), it is possible to predict audit opinion with 72% accracy and 74% precision.

Table 9

Prediction group

	0	1		
Sensitivity 0/695	FN 111	TP 253	1	Real group
Specifity 0/749	TN 260	FP 87	0	
Accuracy 0/722	Negative Prediction Value 0/701	Percision 0/744		

LG Regression Model

The results of appluing LG regression for 10-times (10 samples) are shown in table 10.

Table 10

Specifity	Sensitivity	Percision	Accuracy	FN	TN	FP	TP	sample
0/695	0/706	0/708	0/700	107	241	106	257	1
0/712	0/692	0/716	0/702	112	247	100	252	2
0/709	0/698	0/715	0/703	110	246	101	254	3
0/703	0/692	0/710	0/698	112	244	103	252	4
0/683	0/703	0/699	0/693	108	237	110	256	5
0/697	0/687	0/704	0/692	114	242	105	250	6
0/706	0/687	0/710	0/696	114	245	102	250	7
0/715	0/692	0/718	0/703	112	248	99	252	8
0/697	0/695	0/707	0/696	111	242	105	253	9
0/706	0/690	0/711	0/698	113	245	102	251	10
0/715	0/706	0/718	0/703			**best**		
0/683	0/687	0/699	0/692			**worst**		
0/009	0/006	0/006	0/004			**S.D**		
0/702	0/694	0/710	0/698			**Mean**		

The result of a sample of 10 samples in LG model are shown in tables 10 and 11. Based on table 11, only quick ration, return on assets, stock return and prior audit opinion are recognized as independent variables (p-value<5%). The coefficients of variables are as the following table.

Table 11

variable	P-Value	Z	coefficient	symbol
C	0/822	0/225	0/721	C
Current Ratio	0/148	1/448	0/350	x1
Quick Ratio	0/020	-2/327	-0/702	x2
Inventory Turnover	0/516	-0/649	0/000	x3
Asset Turnover	0/457	0/743	0/174	x4
Receivables Turnover	0/446	-0/762	-0/024	x5
Return on Asset	0/019	2/338	0/054	x6
Return on Investment	0/964	-0/045	0/000	x7
Return on Shareholders Equity	0/252	-1/146	-0/008	x8
Net Income Ratio	0/430	-0/789	-0/007	x9
Operating Cash ratio	0/835	-0/209	-0/138	x10
Investing Cash Ratio	0/212	-1/249	-1/118	x11

Debt Ratio	0/123	1/543	1/268	x12
Market to Book Value	0/235	1/187	0/012	x13
Stock Return	0/033	2/126	3/496	x14
Price to Revenue	0/783	0/275	0/000	x15
Firm Growth	0/427	-0/795	-0/154	x16
Log Net Sales	0/051	-1/948	-0/292	x17
Audit Firm Type	0/967	0/041	0/010	x18
Prior Audit Opinion	0/000	10/007	1/750	x19
Auditor Switch	0/456	-0/745	-0/207	x20
Management Switch	0/171	1/368	0/256	x21
Auditor Tenure	0/583	0/549	0/138	x22
Nomber of Board of Directors Members	0/803	-0/250	-1/057	x23
Firm Age	0/486	-0/696	-0/004	x24

Table 12 shows that accuracy of this model is 70%, its presision is 71% and its sentivity and specifity are 69% and 71% respectively.

Table 12

	Prediction group			
	0	1		
Sensitivity 0/695	FN 111	TP 253	1	Real group
Specifity 0/712	TN 247	FP 100	0	
Accuracy 0/703	Negative Prediction Value 0/690	Percision 0/717		

Comparing WCA and LG results

According to table 13, all criteria in WCA are better than LG. Therefore, in applying these 2 methods in one sample, WCA has more efficient performance.

Table 13

Method	Specifity	Sensitivity	Percision	Accuracy	FN	TN	FP	TP

WCA	0/749	0/695	0/744	0/722	111	260	87	253
LG	0/712	0/695	0/717	0/703	111	247	100	253

For the better comparision, the results obtained 2 methods are compared in 10 samples. Based on table 14, the best situation of WCA is better than GA. Accuracy and Precision criteria are 72 and 76 percent in WCA. Such criteria for LG are 70 and 71. Sentivity of both methods are approximately the same. Hence, gerenral performance of WCA are better than LG. The worst criteria of WCA (exept sentivity) are better than LG ones. The wost accuracy and precision in WCA are 71 and 74 percent (in comparision with 69 and 70 in LG). The mean of all criteria (exept sensitivity) is better in WCA. Sentivity of mthods are approximately the same.

To sum up, similar to 1 run of methods, in 10 samples, WCA has better performance.

Table14

	Method	Specifity	Sensitivity	Percision	Accuracy
best	WCA	0/778	0/701	0/756	0/722
	LG	0/715	0/706	0/718	0/703
worst	WCA	0/735	0/654	0/735	0/714
	LG	0/683	0/687	0/699	0/692
S.D	WCA	0/013	0/014	0/006	0/003
	LG	0/009	0/006	0/006	0/004
Mean	WCA	0/748	0/691	0/742	0/718
	LG	0/702	0/694	0/710	0/698

Conclusion

Audit report has informative content and importance for the users of the firms' financial statement, aiming to making the best economic decisions. This study seeks to predict audit opinion by using a new metaheuristic algorithm named water cycle algorithm (WCA) and comparising its results with one of the most popular algorithms i.e., logistic algorithm (LG).

This study reviewed the literature and selected 24 independent variables in predicting audit opinion of 237 firms listed in Tehran Stock Exchange during 2018 to 2020. Audit opinion was measured as a binary variable (1 if audit opinion was unqualified and otherwise (includes qualified, disclaimer and advers opinion) 0. Predictions obtained by algorithms are compared with real audit opinion and then their results are compared interms of some evaluating criteria.

Findings showed that the precision, accuracy, sentivity and specifity of WCA are 76, 73, 69 and 78%, respectivity and so, WCA is an appropriate method for redicting audit opinion. Moreover, the criteria obtained in WCA are better than LG and so, WCA is more efficient than LG.

According to the findings, among 24 independent variables, extracted from the literature, the most affecting ones were inventory turnover (consistent with Heng-Shu (2017) and Pourheydari and Azami (2011)) and assets turnover (consistent with Spathis, Doumpos, Zopounidis(2003) and Pourheydari and Azami (2011)), ROI, net income ratio (consistent with Yasar, Yakut, Gutnu, (2015) and Laitinen and Laitinen (1998) and Pourheydari and Azami (2011)), the nomber of board of directors' members and prior audit opinion (consistent with Saaydah (2019) and Susanto and Pradipta (2017) and Lu(2020) and Sánchez-Serrano, et, al. (2020)).

Given obtained results, this research has some implications. Since there are significant relationships between audit opinion with some variables (inventory and assets turnover, ROI, net income ratio and the nomber of board of directors' members and prior audit opinion), it is proposed that users should put more emphasis on such variables, hoping to predict audit opinion efficiently.

Moreover, it is proposed to the users of interim period financial statements, some of which are not audited, to use WCA for predicting audit opinion on such statements. Moreover, auditors can use this algorithm in developing audit plans and evaluating and making decision about accepting the clients. In addition, such algorithm can be used as a control tool in the stage of review of a commitment, and for the analysis of the variables that affect the probability of obtaining a qualified opinion. Morover, it is useful in estimating acceptable audit risk and determining an appropriate audit fee.

At last, it is proposed as a research opportunity to examine other metaheuristic algorithms and use other predicting variables, which were missed in this study, in predicting audit opinion and compare their results with each other.

References

DeAngelo, L. Auditor size and audit quality. J. Account. Econ. 1981, 3, 183–199.

Dopuch, Nicholas W. Holthausen Robert and W. Leftwich Richard. (1987). Predicting Audit Qualifications with Financial and Market Variables. *The Accounting Review*, 62(3): 431-454.

Eskandar, H., Sadollah, A., Bahreininejad, A., & Hamdi, M. (2012). Water Cycle Algorithm – A Novel Metaheuristic Optimization Method for Solving Constrained Engineering Optimization Problems. Computers & Structures, (110/111): 151-166.

Heng-Shu, Tang, "Audit opinion of listed companies: a Takagi-Sugeno fuzzy neural network-based study," *Journal of Discrete Mathematical Sciences and Cryptography*, 20(4): 899–912.

Karami, Gholamreza, Karimiyan, Tahere, Salati, Saba. (2017). Auditor Tenure, Auditor Industry Expertise, and Audit Report Lag: Evidences of Iran. *Iranian Journal of Management Studies*, 10 (3): 641-666.

Laitinen, E. K., & Laitinen, T. (1998). Qualified Audit Reports in Finland: Evidence from Large Companies. European Accounting Review, 7(4): 639-653.

Lu, Bin. (2020). Literature Review of Audit Opinion. *Modern Economy,* 11: 28-36.

Moalla, H. Audit report qualification/modification: Impact of financial variables in Tunisia. J. Account. Emerg. Econ. 2017, 7, 468–485.

Moradi, M., Ali, S., Eskandar, H. & Eskandar, H., (2017). The application of water cycle algorithm to portfolio selection. Economic Research-Ekonomska Istraživanja, 30(1):1277–1299.

Pourheydari, Omid and Azami, Zeynab. (2011). Predicting Auditor's Opinions: A Neural Networks Approach. Journal od accounting knowledge, 1(3): 77-97.

Sánchez-Serrano, J. R., Alaminos, D., García-Lagos, F., & Callejón-Gi, A. M. (2020). Predicting Audit Opinion in Consolidated Financial Statements with Artificial Neural Networks. Mathematics, 8(8): 1288.

Saaydah, M. (2019). Coeporate governance and the modification of audit opinion: a study in the Jordanian market. *International jounal of applied research in management and economics,* 2(2): 28-46.

Setayesh, M.H, Ebrahimi, F., Seyf, S.M., Sarikhani, M. (2015). Forecasting the Type of Audit Opinions: A Data Mining Approach. *Management accounting*, 5(4): 69-82.

Spathis, C., Doumpos, M., & Zopounidis, C. (2003). Using Client Performance Measures to Identify Pre-Engagement Factors Associated with Qualifed Audit Reports in Greece. The International Journal of Accounting, 38(3), 267–284.

Susanto, Y. K., Predipta, A. (2017). Coeporate governance and audit decision making. *Corporate ownership & control*, 15(1-2): 381-386.

Yaşar, A., Yakut, E., and Gutnu, M. M. (2015). Predicting Qualified Audit Opinions Using Financial Ratios: Evidence from the Istanbul Stock Exchange. International Journal of Business and Social Science, 6(8): 57- 67.

Zeng, Sen, Li, Yanru and Li, Yanru. (2022). Research on Audit Opinion Prediction of Listed Companies Based on Sparse Principal Component Analysis and Kernel Fuzzy Clustering Algorithm. *Mathematical Problems in Engineering,* Volume 2022: 1-13. https://doi.org/10.1155/2022/4053916