Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 44 No. 11
November 2023

An Optimization Model for Bank Efficiency Forecasting Using Operational Research and Data Driven Methods

[1]Niloufar FarhangDoust, [2]Saiedehgholami
[1]Department of Industrial Engineering, University of Technology K.N.Toosi, Tehran, Iran
[2]Department of Industrial Engineering,K.N.Toosi University of Technology,Tehran,Iran

Abstract- In today's world economy, the role of banks is essential and crucial for any economy, and they have got a pivotal place in the economic growth processes of society. Moreover, the expansion of banking services in the world through information and communication technology networks and the development of virtual and semi-virtual banks and financial institutions and the presence of the private banking system in the country has created high competition in the banking industry. Competition, durability, producing new services, continuous changes in the needs and demands of customers, have forced banks to editing their strategies in a way as to retain current customers and absorb new customers by providing better services. Therefore, it is necessary to measure the performance of banks by an efficiency evaluation system that provides managers with a comprehensive prospect of the business.

Many researchers have applied data envelopment analysis (DEA) to evaluate the efficiency of banks. However, according to the literature, less attention has been paid to the issue of decision making under uncertainty.

In this research, a flexible and reliable model will be proposed in order to optimize the prediction of the bank's efficiency to deal with uncertainty, and it will survey and analyze various performance and efficiency evaluation methods, and by investigating the process of change and evolution of evaluation methods, especially Combined operations research and data-driven models found their deficiencies and weaknesses so that by analyzing and comparing the efficiency of performance and efficiency evaluation methods, the most suitable model is introduced to optimize the bank's efficiency forecast.

In conclusion, among the types of efficiency evaluation methods with the purpose of increasing the accuracy of forecasting in conditions of uncertainty, the robust data-driven data envelopment analysis model was chosen.

Keywords- Performance evaluation, Efficiency evaluation, Banking industry, Data envelopment analysis method, Robust data-driven optimizatio

1. Introduction

In the highly competitive and unsustainable economy of the world, the role of banks is necessary and pivotal for economy because of the following four reasons: (a) presenting intermediation services, (b) creating a wide range of assets and liabilities, (c) providing financial services and (d) creating financial incentives [1]. Therefore, they play a vital role in the economic growth processes of society.

Moreover, evaluating organizational performance is not a simple phenomenon at all and it is a complicate and multidimensional concept. As a result, it is inevitable to apply performance evaluation systems to facilitate the implementation of organizational strategies and to improve organizational performance.

Actually, one of the most important issues of the management chain in the bank is performance measurement, and due to the importance of establishing a proper efficiency evaluation system in today's banks, a number of researchers pay attention to it and have addressed the issue and have presented different models and methods in this field.

Since performance evaluation in the banking industry through conventional and common indicators such as "capital adequacy ratio" and "return on assets" are not accurate enough to estimate efficiency, the use of previous methods to evaluate the efficiency of banks with greater reliability and accuracy, as well as considering innovative and data-driven features that can sometimes use imprecise or low-quality variables as input and provide an acceptable output is the most important justification and incentive for doing the current research [2].

Considering the importance of evaluating the performance and efficiency of banks, and also the importance of the reliability of the numbers extracted from these evaluations to monitor the current situation and find out the difference with the optimal situation in order to maintain or change the behavior patterns and activities of banks, and the problems of optimizing and predicting the efficiency of a bank in the real world have created some problems and they are sometimes a large standard deviation and they include many resources [3]. As a result, the purpose of the current paper is to find an optimal solution to solve such problems.

In addition, the issue of evaluating the efficiency of banks is a sensitive and difficult task due to the wide range of domains, the lack of full disclosure of financial information according to banking laws and sciences in the world, and the presence of some ambiguities in the search process. Furthermore, banks' databases were often formed to comply with traditional accounting methods and do not provide themselves with the mixed analysis of marketing, financial and operational data, as well as interested competing banks. Since they don't share their competitive data, it is critical to examine different methods of predicting performance and finding their merits and demerits, and as a result, the optimal method will be chosen.

In the next section, the variables of data envelopment analysis are determined and the multivariable linear programming model will be presented. The most important points in the definition of variables are their execution and operation. So, the definition of non-implementable criteria is avoided, on the one hand, and the defined variables can be measured and calculated, on the other hand. In addition, the appropriate approach to select input and output variables according to the policies and industry of the investigated organizations, considering banks as service intermediaries, is the cost approach.

After defining the variables, the amount of weight assigned to each of the investigated factors has been received from the stakeholders and experts of the organization in the form of designing and completing the questionnaire.

Pursuit of determining the variables, assigned weights and the data envelopment analysis model, the efficiency was measured. Then after obtaining the measurement results, the efficiency was evaluated. Finally, after providing the optimal solution, we proceeded to present the linear programming model by data envelopment analysis and data-driven robust optimization and analyze the results of two evaluations.

2. Preliminaries

Robust and data-driven optimization

Robust optimization was first proposed by Suister. Algaoui and Lebert and Ben Tal and Nimirevsky extended the robust optimization theory and proposed a new robust model based on the elliptic uncertainty set. Subsequently, Bertsimas and Sim and Bertsimas et al. developed a robust optimization approach based on multivariate uncertainty sets. But less than the robust DEA optimization method has been used to solve the problems of the banking system.

The findings and review of numerous articles show that the DEA method is the best method for data organization and analysis compared to other performance evaluation methods. In addition, since the banking system is affected in the event of an economic crisis, as a result, it is in a non-robust condition, and this uncertainty is caused by various factors such as economic shocks and policy adjustments, and this uncertainty in previous studies have not been widely studied by the researchers [4]. Hence, it ultimately leads to a decrease in accuracy in estimation and decision-making.

Therefore, in the current study, a flexible and reliable model is proposed to optimize the prediction of banks' efficiency to deal with uncertainty to ensure that the performance of the banking system be completely robust. Thus, the purpose is to optimize robustness in case of data disruption, and to ensure that the most appropriate model be introduced to optimize the prediction of banks' efficiency.

The concept of data-driven robust optimization

Robust and data-driven optimization is also known by other titles, such as data-based robust optimization, and is known as a modern approach to optimization and decision-making under conditions of uncertainty, which has attracted the attention of optimization and operational research specialists in the academic and industrial environment.

In developing the non-deterministic programming approaches, data-driven robust optimization has been evolved, which includes various methods and models and it is used in many problems and applications. It is necessary to explain that all data-driven optimization methods are not included in the robust optimization category because there is not necessarily a robustness property in the output of all of them. For example, methods such as data-driven optimization based on the extraction of point estimates of non-deterministic parameters or probability distribution estimation are never included in the category of robust data-driven optimization models, but are considered to be classical approaches of data-driven optimization. In other words, it can be said that data-driven robust optimization models are a subset of data-driven optimization models in which not only the available data related to non-deterministic parameters are used, but also the robustness of the model response is guaranteed and the extreme conservatism of classical robust optimization models is reduced.

Data-driven optimization uses observations of random variables as direct input to mathematical programming problems. Meanwhile, optimization for the worst value of parameters in a set is known as "robust optimization". Classical robust optimization methods are mainly based on experience to obtain uncertainty sets, but data-driven robust optimization method uses past data to construct uncertainty sets, which improves the uncertainty sets of traditional robust optimization method.

In principle, it is based on the concept that the output of the model should not be very sensitive to the values of the input parameters. The importance of this feature in modeling has caused the application of robust discussion in various fields to face increasing growth in recent years. In general, the models introduced in robust planning are divided into two general forms:

- Models that are defined based on discrete scenarios.
- Models that are based on the concept of uncertainty sets and are based on the fluctuation of parameters in an interval, that uncertainty is modeled in the form of bounded uncertainty sets, and the goal is to find an answer that is suitable for most parameters Uncertainty is not sensitive. In general, the uncertainty sets include box, ellipse and multifaceted.

3. Methodology

3.1. DEA Robust Optimization Model

3.1.1. proposed model

In the model presented for the DEA method, the decision variables and parameters are all deterministic. However, in a real scenario, this will lead to errors if there are uncertain factors. Since the operational efficiency of the banking system is affected by adaptive policies and economic shocks, therefore, the DEA model is not suitable for real situations under certain conditions. Therefore, it is necessary to consider the impact of these two uncertainties in calculations and evaluations, and for this purpose, optimization is based on an approach that seeks the optimal solution in the worst case.

Here, three sets of box, ellipse and polyhedral uncertainty are examined to describe different types of uncertainty. According to the results obtained from this research, the polyhedral uncertainty set is the worst result and the oval uncertainty set is the most robust and desirable. It has the most efficiency.

In general, we express the set of uncertainty as follows:

$$U = \{Y_{r0}^{D} = Y_{r0} + \sum_{l=1}^{L} Y_{r0l}^{F} \varepsilon_l \ \ , \ Y_{rj}^{D} = Y_{rj} + \sum_{l=1}^{L} Y_{rjl}^{F} \varepsilon_l \ \ , \ \varepsilon_l \in Z\} \quad \text{Formula 3.1}$$

Finally, the following linear model can be obtained:

$$\max \theta$$

$$\theta - \sum_{r=1}^{s} u_r \left(Y_{r0} + \sum_{l=1}^{L} Y_{r0l}^{F} \varepsilon_l \right) \leq 0$$

$$\sum_{i=1}^{m} (v_i X_{io}) \leq 1$$

$$\sum_{r=1}^{s} u_r \left(Y_{rj} + \sum_{l=1}^{L} Y_{rjl}^{F} \varepsilon_l \right) - \sum_{i=1}^{m} (v_i X_{ij}) \leq 0 \quad \forall j$$

$$u_r, v_i \ \geq 0$$

$$i = 1,2, \dots , m \quad j = 1,2, \dots , n \quad r = 1,2, \dots , s$$

Formula 3.2

3.1.2. Proposed model sets

Set of inputs: I

The set of decision-making units: J

The set of outputs: R

The set of uncertainty factors: L

3.1.3. Decision variables of the proposed model

The amount of r_{th} output from the j_{th} decision making unit (s): $\boldsymbol{Y_{rj}}$

Amount of i_{th} input to j_{th} decision making unit (m): $\boldsymbol{X_{ij}}$

The j_{th} decision making unit (n): ***DMU$_j$***

Fluctuations of r_{th} output from j_{th} decision making unit with different uncertainty factors: $\boldsymbol{Y_{rjl}^{F}}$

L_{th} uncertainty factor (L): $\boldsymbol{\varepsilon_l}$

3.1.4. Numerical parameters of the proposed model

Variable coefficient of r_{th} output rate: $\boldsymbol{u_r}$

variable coefficient of i_{th} input amount: $\boldsymbol{v_i}$

3.1.5. Proposed model with elliptical uncertainty set

The elliptic uncertainty set is expressed using the second soft (standard Euclidean) uncertainty data vector, whose mathematical relationship is as follows:

$$Z = \left\{ \varepsilon \middle| \|\varepsilon\|_2 \le \Omega \right\} = \{ \varepsilon | \sqrt{\sum_{l\in L} \varepsilon_l^2} \le \Omega \}$$
Formula 3.3

In this mathematical relationship, Ω is an adjustable parameter that controls the size of the uncertainty set. $\sum_{l=1}^{L}(\sum_{r=1}^{S} u_r Y_{r0l}^F)^2$ shows the uncertainty of the output data that the effect of L causing disruption in Displays the output data. Finally, the proposed model is as follows:

$$\max \theta$$

$$\theta + \Omega \sqrt{\sum_{l=1}^{L}(\sum_{r=1}^{S} u_r Y_{r0l}^F)^2} - \sum_{r=1}^{s} u_r(Y_{r0}) \le 0$$

$$\sum_{i=1}^{m}(v_i X_{io}) \le 1$$

$$\Omega \sqrt{\sum_{l=1}^{L}(\sum_{r=1}^{S} u_r Y_{r0l}^F)^2} - \sum_{i=1}^{m}(v_i X_{ij}) + \sum_{r=1}^{s} u_r(Y_{rj}) \le 0 \quad \forall j$$

$$u_r, v_i \ge 0$$

$$i = 1,2, \dots, m \quad j = 1,2, \dots, n \quad r = 1,2, \dots, s$$
Formula 3.4

3.2. Data-Driven Robust Optimization (DDRO) Model

3.2.1. Proposed Model

In the model presented for the robust DEA method, it was assumed that the output variables are uncertain because of the presence of uncertain factors in the real environment. In this method, there was not enough information and the number of fluctuations of the output variable was calculated experimentally. It also has a very conservative approach to ensure absolute robustness. This is despite the fact that in real conditions, we can roughly expect the government's tendency towards macro-policy. At the same time, we can act experimentally instead of not having useful information available from some historical data. Unlike the previous section, here the set of observations of the output variables named P is considered, which practically includes the raw data of the output, which is defined as follows:

$$P = \{ Y_{r1}, \dots, Y_{rM} \} \quad Y_{rq} \in \boldsymbol{R}, \boldsymbol{q = 1, \dots, M}$$

$$\hat{Y}_r = \frac{1}{M} \sum_{q=1}^{M} Y_{rq}$$
Formula 3.5

In this section, the elliptic uncertainty set is considered to describe the uncertainty, and finally the following model can be obtained:

$$\max \left\{ \min \sum_{r=1}^{S} (u_r \hat{Y}_r), \hat{Y}_r \in \boldsymbol{Z} \right\}$$

$$\sum_{i=1}^{m}(v_i X_{io}) \le 1$$

$$\sum_{r=1}^{s}(u_r Y_{rj}) - \sum_{i=1}^{m}(v_i X_{ij}) \le 0 \quad \forall j$$

$$u_r, v_i \ge 0$$

$$i = 1,2, \dots, m \quad j = 1,2, \dots, n \quad r = 1,2, \dots, s$$
Formula 3.6

3.2.2. Proposed model sets

Set of inputs: I

The set of decision-making units: J

The set of outputs: R

3.2.3. Decision variables of the proposed model

The amount of r_{th} output from the j_{th} decision making unit (s): $\boldsymbol{Y_{rj}}$

Amount of i_{th} input to j_{th} decision making unit (m): $\boldsymbol{X_{ij}}$

The j_{th} decision making unit (n): ***DMU****j*

3.2.4. Numerical parameters of the proposed model

Variable coefficient of r_{th} output rate: $\boldsymbol{u_r}$

variable coefficient of i_{th} input amount: $\boldsymbol{v_i}$

3.2.5. Proposed model with elliptical uncertainty set

The ellipse uncertainty set consists of observations whose geometric shape of the isopycnal obtained from their multivariate normal distribution is an ellipse. Therefore, the maximum fit of output variables P is a normal distribution N (μ, ∑) with μ= $\hat{Y}_r$ *and* $\sum = \frac{1}{M} \sum_{q=1}^{M}(Y_{rq} - \mu)(Y_{rq} - \mu)^T$. Finally, for the model, we assume an ellipse with the shape $Z = \{Y: (Y_r - \hat{Y}_r)^T \sum^{-1} (Y_r - \hat{Y}_r) \le r\}$r} and ω is an adjustable parameter showing the size of the uncertainty set. Additionally, according to Bertsimas and Sim's approach, the parameter γ is considered for each uncertain variable, which is not necessarily correct. Moreover, its role is to adjust the robustness of the proposed method against the level of conservatism of the answer. The proposed model is determined as follows:

$$\min - \sum_{r=1}^{s}\left(u_r\hat{Y}_{ro}\right) + \gamma$$

$$\omega . u_r^T \sum u_r \leq \gamma^2 \quad , \gamma \in [S]$$

$$\omega . u_r^T \sum u_r \leq \gamma^2 + \sum_{i=1}^{m} v_i X_{ij} - \sum_{r=1}^{s}\left(u_r Y_{rj}\right) \ , \forall j$$

$$\sum_{i=1}^{m}\left(v_i X_{io}\right) \leq 1$$

$$u_r, v_i \geq 0$$

$$i = 1,2, \dots, m \quad j = 1,2, \dots, n \quad r = 1,2, \dots, s$$

Formula 3.7

3.3. Determining inputs and outputs

In this paper, the data from the financial statements of 2018 to 2020 banks in Codal are used. In addition, the desired approach to select the input and output variables from among the data of financial statements, according to the policies and industry of the investigated organizations and in consultation with experts and experts, considering banks as Service intermediary is a cost approach.

After surveying and consulting with the experts, the inputs include ($X_{ij}(\mathrm{m} = 4)$) the cost of interest on deposits, the cost of fees, administrative and general costs, and the expense of doubtful claims, and the outputs include ($\boldsymbol{Y_{rj}}$ $(s = 3)$) is the income of granted facilities, deposit income and fee income.

Table 3.1. *Banks' data table to calculate efficiency (numbers are in millions of Rials)*

No	Bank	Year	Deposit interest expense (i=1)	Bank service charge (i=2)	Administrative and general expenses (i=3)	The expense of doubtful claims (i=4)	Income from granted facilities (r=1)	deposit income (r=2)	Fee income (r=3)
1	Pasargad	2020	266,150,927	4,994,601	27,314,352	20,241,492	141,549,306	11,358,619	29,672,340
	Pasargad	2019	156,624,399	3,457,536	19,649,235	38,091,334	116,837,587	9,558,098	17,450,088
	Pasargad	2018	127,388,555	4,002,094	14,446,456	10,726,337	94,747,899	6,876,490	9,835,617
	Pasargad	2017	93,855,838	2,199,937	11,906,096	18,780,085	97,498,825	3,234,953	7,231,312
	Pasargad	2016	85,162,995	2,122,640	9,256,886	16,522,253	125,406,301	2,390,230	4,084,550
2	Parsian	2020	195,667,722	7,307,558	24,444,487	20,070,194	103,711,345	12,291,487	17,261,164
	Parsian	2019	161,793,356	4,961,989	18,619,316	13,515,400	96,813,475	5,668,334	7,092,306
	Parsian	2018	147,433,577	3,114,842	11,946,357	4,536,868	103,229,653	4,656,623	3,575,527
	Parsian	2017	132,151,924	2,749,393	9,862,256	2,942,020	28,787,224	3,541,577	3,047,223
	Parsian	2016	116,367,907	2,101,381	8,561,397	7,443,049	23,976,848	3,988,487	2,556,613
3	Dey	2020	73,180,168	694,091	7,392,951	13,210,026	19,456,832	2,410,074	2,002,118
	Dey	2019	53,572,566	355,361	5,187,111	946,344	14,623,031	2,220,291	1,035,509
	Dey	2018	55,233,048	163,144	3,862,604	0	18,347,258	2,319,888	562,719
	Dey	2017	45,909,458	141,367	3,354,478	6,521,565	597,220,337	1,056,143	508,739
	Dey	2016	46,616,735	122,455	3,158,094	3,759,219	362,972,208	1,415,476	533,099
4	Mellat	2020	363,328,277	20,778,897	241,110,610	12,469,868	201,262,790	63,572,036	47,646,875
	Mellat	2019	216,735,346	16,479,242	192,846,278	59,379,833	172,432,086	76,655,990	29,194,491
	Mellat	2018	154,963,273	13,813,278	88,112,948	69,931,978	163,601,905	49,497,113	19,005,700
	Mellat	2017	151,698,318	12,072,446	58,949,683	80,656,199	335,386,023	37,646,175	15,963,749

No	Bank	Year	Deposit interest expense (i=1)	Bank service charge (i=2)	Administrative and general expenses (i=3)	The expense of doubtful claims (i=4)	Income from granted facilities (r=1)	deposit income (r=2)	Fee income (r=3)
	Mellat	2016	159,921,856	9,611,092	49,584,533	16,198,957	226,289,614	22,205,190	14,802,151
5	Tejarat	2020	314,354,242	5,700,293	144,863,062	50,234,884	172,218,699	96,197,267	43,663,820
	Tejarat	2019	196,318,317	4,247,222	107,317,463	16,011,051	159,941,134	61,350,673	24,252,872
	Tejarat	2018	167,980,235	4,124,383	51,883,155	16,412,472	122,889,479	22,900,760	17,719,299
	Tejarat	2017	146,412,554	3,106,542	39,603,361	17,614,332	396,165,439	4,787,003	11,206,293
	Tejarat	2016	153,671,267	3,313,607	34,107,891	15,197,371	265,663,492	1,583,692	7,744,877
6	Saderat	2020	324,936,242	8,551,500	130,634,255	36,017,446	204,047,038	14,801,488	36,193,505
	Saderat	2019	239,987,322	6,872,868	101,906,517	26,730,340	156,606,994	18,789,685	25,584,573
	Saderat	2018	167,083,472	5,997,316	73,171,269	17,011,421	140,149,281	8,999,229	18,622,528
	Saderat	2017	147,975,148	5,781,583	59,002,197	24,130,428	38,329,114	10,461,137	12,925,884
	Saderat	2016	142,969,871	5,791,672	48,654,014	14,736,161	24,692,618	15,529,843	11,410,844
7	Khavare Myane	2020	35,226,702	374,320	3,597,952	2,046,084	12,427,834	14,961,931	2,953,555
	Khavare Myane	2019	22,416,790	272,686	2,673,858	2,115,498	11,208,402	11,169,622	1,748,992
	Khavare Myane	2018	10,008,961	132,013	1,859,466	390,157	9,135,805	5,000,340	1,350,564
	Khavare Myane	2017	8,654,844	140,950	1,078,255	477,041	62,891,605	2,295,437	700,013
	Khavare Myane	2016	8,132,004	89,344	754,191	591,360	74,390,617	1,986,448	731,258
8	Gardeshgari	2020	118,733,668	579,985	8,723,722	7,098,684	58,145,780	3,405,711	1,701,565
	Gardeshgari	2019	86,833,770	400,148	7,244,530	15,199,195	37,998,033	765,682	911,813
	Gardeshgari	2018	63,381,803	327,572	5,300,703	10,270,582	27,941,227	989,653	692,773
	Gardeshgari	2017	47,469,388	255,947	3,468,331	136,177	149,235,235	1,314,681	524,674
	Gardeshgari	2016	40,171,154	178,207	2,927,907	8,172,299	106,775,304	1,189,904	422,020
9	Shahr	2020	160,854,621	4,318,876	23,931,556	4,722,986	82,364,295	2,778,856	23,080,541
	Shahr	2019	135,416,646	3,536,232	16,363,967	31,217,770	76,215,337	2,041,238	8,192,916
	Shahr	2018	125,810,128	2,455,514	10,471,568	26,576,990	64,009,347	2,561,608	3,653,291
	Shahr	2017	105,338,384	1,075,401	8,518,193	18,468,895	200,864,107	1,618,631	2,411,391
	Shahr	2016	96,648,836	688,618	7,016,188	8,412,774	94,262,996	980,629	2,829,012
10	Eqtesade Novin	2020	171,133,626	1,520,207	18,809,029	9,007,514	35,564,083	10,499,613	16,006,176

No	Bank	Year	Deposit interest expense (i=1)	Bank service charge (i=2)	Administrative and general expenses (i=3)	The expense of doubtful claims (i=4)	Income from granted facilities (r=1)	deposit income (r=2)	Fee income (r=3)
	Eqtesade Novin	2019	82,734,953	1,011,151	12,103,309	13,343,673	40,346,118	4,867,172	9,475,597
	Eqtesade Novin	2018	51,829,438	1,007,991	8,364,804	2,759,809	47,520,702	6,125,352	4,757,076
	Eqtesade Novin	2017	47,230,683	919,481	6,868,461	2,598,549	0	1,521,571	3,976,309
	Eqtesade Novin	2016	51,567,982	854,640	5,826,330	3,442,894	0	1,037,997	3,958,498
11	Qarz Al-Hasaneye Resalat	2020	0	2,725,389	13,296,845	2,976,593	0	23,223,380	10,343,748
	Qarz Al-Hasaneye Resalat	2019	0	2,553,524	8,285,109	1,127,763	0	5,363,913	7,355,239
	Qarz Al-Hasaneye Resalat	2018	0	2,074,796	6,266,152	500,934	0	5,182,926	4,852,342
	Qarz Al-Hasaneye Resalat	2017	0	1,602,505	4,785,869	451,774	102,406,344	2,350,241	4,319,872
	Qarz Al-Hasaneye Resalat	2016	0	1,195,509	3,148,293	332,828	49,463,049	1,913,827	3,117,535
12	Saman	2020	89,117,389	5,029,323	18,908,608	8,000,000	29,069,859	5,363,881	7,975,952
	Saman	2019	54,434,294	4,023,794	13,824,016	4,000,000	24,463,856	9,289,884	4,539,669
	Saman	2018	42,942,724	3,096,222	8,787,295	3,000,000	18,223,887	5,566,004	2,996,895
	Saman	2017	36,851,249	2,458,648	6,505,564	1,100,000	4,152,891	2,178,681	2,392,080
	Saman	2016	33,822,475	1,951,868	5,146,638	1,200,000	4,331,073	1,355,993	1,640,633
13	Sarmaye	2020	30,208,456	309,811	5,362,556	2,033,847	2,910,075	345,084	540,868
	Sarmaye	2019	27,742,496	296,546	3,889,769	3,979,064	4,182,611	872,360	389,269
	Sarmaye	2018	30,326,405	203,519	3,131,671	8,340,642	17,164,286	888,066	339,886
	Sarmaye	2017	35,907,096	185,348	2,606,463	17,337,002	30,055,350	900,287	363,186
	Sarmaye	2016	40,277,886	143,127	2,371,799	13,590,003	17,624,841	1,035,938	291,908
14	Post Bank	2020	14,725,916	9,935,491	11,622,769	996,587	11,522,933	9,785,961	9,426,783
	Post Bank	2019	9,167,696	5,889,876	7,870,546	1,303,701	7,567,876	4,913,751	5,411,327

No	Bank	Year	Deposit interest expense (i=1)	Bank service charge (i=2)	Administrative and general expenses (i=3)	The expense of doubtful claims (i=4)	Income from granted facilities (r=1)	deposit income (r=2)	Fee income (r=3)
	Post Bank	2018	9,497,856	3,411,285	6,617,538	632,292	7,103,834	5,404,553	3,241,539
	Post Bank	2017	9,760,261	2,943,646	5,048,803	1,306,734	14,973,998	3,729,615	3,112,434
	Post Bank	2016	7,804,499	1,391,344	3,742,758	1,973,544	4,536,986	2,686,367	3,209,826
15	Iran Zamind	2020	98,132,272	654,385	9,178,729	749,003	4,320,823	512,171	1,750,597
	Iran Zamind	2019	74,610,305	527,351	7,719,263	1,133,111	4,787,342	116,895	905,628
	Iran Zamind	2018	59,717,744	382,511	4,240,499	404,619	5,189,666	360,856	464,073
	Iran Zamind	2017	36,788,441	327,207	3,348,847	363,845	57,289,158	17,108,422	369,432
	Iran Zamind	2016	33,180,298	380,628	2,627,688	494,264	29,860,274	319,544	330,714
16	Kar Afarin	2020	48,377,506	257,565	10,711,788	2,467,437	18,395,013	8,454,026	4,054,071
	Kar Afarin	2019	27,797,974	284,224	5,957,284	2,494,023	16,884,179	6,455,922	2,366,281
	Kar Afarin	2018	18,874,679	177,879	3,374,079	2,102,647	18,003,874	4,957,648	1,326,562
	Kar Afarin	2017	18,303,164	138,863	2,665,461	1,510,175	49,767,329	3,442,379	1,102,814
	Kar Afarin	2016	17,535,451	99,677	2,292,304	1,921,401	34,276,434	1,362,832	847,679
17	Sina	2020	38,110,738	906,728	15,445,092	2,188,051	24,702,548	9,112,675	5,014,773
	Sina	2019	26,126,089	646,360	11,889,445	1,952,577	23,619,478	4,095,358	3,300,794
	Sina	2018	20,623,435	535,614	6,429,580	2,567,833	23,155,769	2,918,189	2,114,649
	Sina	2017	20,710,112	516,072	4,703,564	2,104,274	248,210,430	1,298,977	1,788,105
	Sina	2016	21,786,578	480,265	4,008,047	925,613	231,054,134	1,859,384	1,447,388
18	Ayande	2020	477,813,768	2,156,596	28,500,182	10,510,173	215,310,140	3,443,315	4,382,806
	Ayande	2019	364,389,466	1,486,046	20,402,932	5,544,945	235,400,502	7,336,231	2,597,541
	Ayande	2018	333,794,219	1,152,777	20,538,707	204,823,400	128,801,456	2,040,139	1,909,374
	Ayande	2017	246,910,074	2,218,093	15,547,059	36,728,529	220,894,750	1,296,271	2,894,310
	Ayande	2016	186,195,610	589,981	10,647,870	33,434,629	141,377,537	997,160	7,530,557
19	Refahe Kargaran	2020	196,331,615	5,009,352	75,067,705	11,577,595	127,646,611	6,945,626	17,973,525
	Refahe Kargaran	2019	141,161,624	4,063,449	57,781,005	8,033,339	100,559,976	11,733,734	10,871,380
	Refahe Kargaran	2018	118,604,764	3,779,226	30,433,196	6,689,000	89,922,377	1,850,798	7,355,219
	Refahe Kargaran	2017	90,891,626	3,356,217	33,267,214	2,519,371	141,549,306	1,919,386	7,261,149
	Refahe Kargaran	2016	74,511,740	2,921,351	21,923,307	3,716,541	116,837,587	761,188	5,429,622

3.4. Weighting to input and output variables

In addition, the weighting of the variables was done by surveying the experts of the banking industry.

Table 3.2. *Table of bank data coefficients to calculate efficiency*

(The sum of input and output coefficients separately is 100 and each numerical coefficient is considered between 0 and 100.)

		INPUT COEFFICIENTS					OUTPUT COEFFICIENTS			
No	name	Deposit interest expense v_1	Bank service charge v_2	Administrative and general expenses v_3	The expense of doubtful claims v_4	Total	Income from granted facilities u_1	Deposit income u_2	Fee income u_3	Total
1	Expert 1	55%	5%	10%	30%	100%	25%	40%	35%	100%
2	Expert 2	55%	25%	10%	10%	100%	65%	15%	20%	100%
3	Expert 3	81%	2%	10%	7%	100%	93%	1%	6%	100%
4	Expert 4	83%	2%	9%	6%	100%	90%	3%	7%	100%
5	Expert 5	60%	20%	10%	10%	100%	70%	10%	20%	100%
6	Expert 6	60%	20%	10%	10%	100%	50%	20%	30%	100%
7	Expert 7	10%	30%	40%	20%	100%	20%	35%	45%	100%
8	Expert 8	10%	30%	40%	20%	100%	20%	35%	45%	100%
9	Expert 9	81%	2%	10%	7%	100%	93%	1%	6%	100%
	Weighting the variables	0.55	0.15	0.17	0.13	1.00	0.58	0.18	0.24	1.00

3.5. Examining the results of solving the proposed model

3.5.1. Efficiency calculation based on DEA model

In the robust data-driven DEA model, non-deterministic output variables are assumed. In our model, the effect of the changes of the two uncertainty factors of adaptive policies and economic shocks on the output variables is investigated. Therefore, l=2 is considered $Yr0lF$=0.02 $Yr0$and $YrjlF$=0.02 Yrj are also Proposed. It is worth noting that the rate of change of the output variables is assumed to be from zero to 10%. As a result, the uncertainty parameter is changed between 0 and 5 with a step of one to observe the gradual changes in efficiency. 2020 data is used in this model.

Table 3.3. *Efficiency of banks by DEA model based on elliptic uncertainty set*

No	**Bank**	**1**	**2**	**3**	**4**	**5**	**R**
1	Eqtesade Novin	0.708	0.687	0.666	0.646	0.625	3
2	Ayande	0.253	0.246	0.239	0.231	0.224	8
3	Ayande	0.056	0.054	0.053	0.051	0.050	17
4	Parsian	0.177	0.172	0.167	0.162	0.156	10
5	Pasargad	0.885	0.859	0.834	0.808	0.782	1
6	Post Bank	0.784	0.761	0.738	0.715	0.693	2

No	Bank	1	2	3	4	5	R
7	Khavare Myane	0.098	0.095	0.092	0.089	0.086	16
8	Khavare Myane	0.105	0.102	0.099	0.096	0.093	15
9	Dey	0.341	0.331	0.321	0.311	0.301	7
10	Refahe Kargaran	0.390	0.379	0.368	0.356	0.345	5
11	Saman	0.156	0.152	0.147	0.142	0.138	11
12	Sarmaye	0.198	0.192	0.187	0.181	0.175	9
13	Sina	0.009	0.009	0.009	0.009	0.008	19
14	Shahr	0.111	0.108	0.104	0.101	0.098	14
15	Saderat	0.032	0.031	0.030	0.029	0.028	18
16	Qarz Al-Hasane Resalat	0.119	0.115	0.112	0.108	0.105	12
17	Kar Afarin	0.114	0.111	0.108	0.104	0.101	13
18	Gardeshgari	0.383	0.372	0.360	0.349	0.338	6
19	Mellat	0.423	0.411	0.398	0.386	0.374	4

3.5.2. Efficiency calculation based on robust data-driven DEA model

In this proposed model, unlike the previous model, the set of observations of the output variables named P is considered, which practically includes the raw data of the output, which is defined as follows:

$P = \{ Y_{r1}, \dots, Y_{r5} \} \quad Y_{rq} \in \boldsymbol{R}, \boldsymbol{q} = \mathbf{1}, \dots, \mathbf{5}$

$\hat{Y}_r = \frac{1}{5} \sum_{q=1}^{5} Y_{rq}$

Formula 3.8

Elliptical uncertainty set is applied, and accordingly, the maximum fit of output variables P is a normal distribution N (μ, ∑) with $\mu = \hat{Y}_r$ and $\sum = \frac{1}{M} \sum_{q=1}^{M} (Y_{rq} - \mu)(Y_{rq} - \mu)^T]\!]$. Finally, for the model, we assume an ellipse with the shape $Z = \{Y: \left(Y_r - \hat{Y}_r\right)^T \Sigma^{-1} (Y_r - \hat{Y}_r) \le r\}$ and ω is an adjustable parameter between zero 5 and has been changed with step one to observe the gradual changes in efficiency. Furthermore, for each non-deterministic variable the parameter γ is taken into account which is not necessarily correct and its role is to adjust the degree of robustness of the proposed method against the level of conservatism of the answer. In this model, data from 2017 to 2020 has been used.

Table 3.4. *The efficiency of banks by the robust data-driven DEA model with ellipse uncertainty set*

No	Bank	1	2	3	4	5	R
1	Eqtesade Novin	0.517	0.516	0.515	0.513	0.509	4
2	Iran Zamin	0.289	0.287	0.284	0.283	0.282	7
3	Ayande	0.056	0.055	0.053	0.051	0.049	17
4	Parsian	0.995	0.995	0.994	0.994	0.993	1
5	Pasargad	0.727	0.726	0.723	0.721	0.719	2
6	Post Bank	0.698	0.696	0.695	0.693	0.691	3

No	Bank	1	2	3	4	5	R
7	Tejarat	0.077	0.077	0.075	0.071	0.069	16
8	Khavare Myane	0.115	0.113	0.111	0.109	0.107	11
9	Dey	0.269	0.267	0.264	0.263	0.261	8
10	Welfare of workers	0.252	0.250	0.248	0.247	0.244	9
11	Saman	0.084	0.083	0.084	0.084	0.077	15
12	Sarmaye	0.141	0.141	0.141	0.141	0.133	10
13	Sina	0.019	0.019	0.019	0.019	0.012	19
14	Shahr	0.094	0.094	0.094	0.094	0.087	13
15	Saderat	0.034	0.034	0.034	0.034	0.026	18
16	Qarz Al-Hasane Resalat	0.091	0.091	0.091	0.091	0.083	14
17	Kar Afarin	0.100	0.100	0.100	0.100	0.093	12
18	Gardeshgari	0.458	0.458	0.458	0.458	0.450	5
19	Mellat	0.373	0.373	0.373	0.373	0.365	6

3.6. Ranking of the investigated organizations

As shown in Table 4.1, the appropriate data to calculate the efficiency of banks were extracted from the financial statement data from 2018 to 2020 from 19 banks, and then the variables were weighted by surveying the experts and specialists of the banking industry. In addition, the desired approach for selecting input and output variables is the cost approach.

Based on the mentioned contents, the efficiency of the mentioned 19 banks has been calculated by the robust DEA model and the data-driven robust DEA model and using "GAMS" software, and the results of the calculations have been given in Tables 4.3 and 4.4. But the critical point is the difference in the ratings obtained by the banks in the evaluation made with the DEA model and the data-driven DEA model, which is summarized in Table 4.5. Since the accuracy of the estimation and calculation of efficiency is based on the DEA model, it is mostly data-driven and, as a result, the ratings obtained based on the data-driven model will be more reliable.

Table 3.5. *Comparing the efficiency of banks by the two investigated models*

Bank	**The rating obtained is based on the robust DEA model**	**The rating obtained based on the data-driven DEA model**
Pasargad	1	2
Post Bank	2	3
Eqtesade Novin	3	4
Mellat	4	6
Welfare of workers	5	9
Gardeshgari	6	5
Dey	7	8

Bank	The rating obtained is based on the robust DEA model	The rating obtained based on the data-driven DEA model
Iran Zamin	8	7
Sarmaye	9	10
Parsian	10	1
Saman	11	15
Qarz Al-HasaneResalat	12	14
Kar Afarin	13	12
Shahr	14	13
Khavare Myane	15	11
Tejarat	16	16
Ayande	17	17
Saderat	18	18
Sina	19	19

4. Conclusion

According to the investigations carried out in previous studies on three sets of box, ellipse and multifaceted uncertainty to describe different types of uncertainty, the set of multifaceted uncertainty is the worst result and the set of oval uncertainty is the most robust and desirable. It has the most efficiency and, as a result the elliptical uncertainty set was applied in this study.

Based on the obtained results, increasing the adjustable parameter Ω and ω from one to five with step one, which shows the size of the uncertainty set, it leads to an increase in the conservatism of the limits and a decrease in the calculated efficiency. Moreover, in the second approach, the parameter γ is considered for each non-deterministic variable, which is not necessarily correct. In addition, its role is to adjust the degree of robust of the proposed method against the level of conservatism of the answer. Therefore, if it is equal to S, it indicates the highest level of protection regarding the uncertainty. As a result, it leads to the change of data during the implementation of the initial model as much as possible to create the least changes in the initial model and its justification is more likely.

Comparing these two proposed approaches, Data-Driven Robust Optimization (DDRO) Model (DEA model) and Robust Optimization Model (DEA mode) , the former shows the performance of a larger period of time and is more accurate than the latter. In data-driven methods, more information is provided, so we can describe the uncertainty more accurately, leading to a more satisfactory result. Meanwhile, in the first approach, the uncertainty parameter was effective in the size of the uncertainty set, but in the second approach, in addition to this parameter, the data set also affects the size of the uncertainty set.

References

[1] Zedgenizova I, Ignatyeva I, Zarubaeva E, Teplova D. IT opportunities: increasing the level of financial security in digital economy. J Adv Pharm Educ Res. 2021;11(3):157- 61.

[2] Smirnov DA, Trofimov MS, Botasheva LE, Melnikova VA. Legal features of the use of big data in the financial activities of the state. J Adv Pharm Educ Res. 2021;11(4):24-8.

[3] Viet H, Do V, Phan NH, Lan PD, Ngoc DV. Determinants Influencing Non-Performing Loan Ratio of Joint Stock Commercial Banks in Vietnam. J. Organ. Behav. Res. 2023: 8(1); 214-30.

[4] Vetrova EA, Kabanova EE, Maslikov VA, Bogacheva TV, Kubasova EI. Impact of the financial and economic activities of university on its development. J Adv Pharm Educ Res. 2021;11(4):59- 67.