Arabian Journal for Science and Engineering (2019) 44:3733–3743
https://doi.org/10.1007/s13369-018-3647-7

RESEARCH ARTICLE - COMPUTER ENGINEERING AND COMPUTER SCIENCE

The Improvement of Forecasting ATMs Cash Demand of Iran Banking Network Using Convolutional Neural Network

Soodabeh Poorzaker Arabani[1] · Hosein Ebrahimpour Komleh[1]

Received: 16 February 2018 / Accepted: 18 November 2018 / Published online: 3 December 2018

Abstract

One of the problems related to the banking system is Automated Teller Machine (ATM) cash demand forecasting. If an ATM faces a shortage of cash, it will face the decline of bank popularity and in turn will have some costs and the bank will encounter decreasing customers use of these systems. On the other hand, if the bank faces cash trapping at an ATM, regarding inflation in Iran, cash trapping and the lack of using it will have a negative impact on bank profitability. The aim of this study is to predict accurately to eliminate the posed double costs. Since the information related to the amount of cash is daily, each ATM will have a behavior as time series and also because the aim of this study is to predict the demand for cash from the 1056 ATMs, we are facing data from the type of panel. The methods that are used for forecasting ATM cash demand in this research include: forecasting by statistical method, artificial neural network intelligent method, Support vector machine and Convolutional neural network. We will compare the results of these methods and show that intelligent methods in comparison with statistical analysis have higher accuracy.

Keywords Artificial neural network · ATM cash demand · Intelligent forecasting · Statistical forecasting · Convolutional neural network · Support vector machine

1 Introduction

According to the figures from the Central Bank of the Islamic Republic of Iran in the year of 1393, more than 450 million transactions have been recorded for Automated Teller Machines in the country. Considering this amount of transaction for a 1-year period, we face large data. This volume of data is important in two respects. Each bank, as an enterprise, seeks to maximize its profits and to minimize the cost of financial transactions. Therefore, cash demand directly affects the profitability of each bank and the existence or absence of cash in each ATM indirectly affects the reputation of the bank. On the other hand, considering the formation of large volume panel data, the importance of creating a model with the ability to detect time dependencies is necessary for more than a few hundred steps.

The research on cash demand forecast has been important [1]. In addition, the use of artificial neural network has been effective in forecasting cash demand [1]. Jadwal et al. [1] stated that the optimal cash flow forecast is a complex operation. But in our research, we use big panel data. That is, in our data, both there is a time series behavior, and the ATMs are affected by each other.

The aim of this research is to predict ATMs cash demand of Iranian banking network based on traditional statistical method, ANN, SVM, CNN and the comparison of these methods in accuracy prediction to specify the best way for solving the problem of ATMs cash demand forecast and we can reduce the ATMs store complement dual costs of banking system.

2 Literature

There are several researches to predict the demand for cash in the ATMs. In this section, we examine a number of these studies.

The approach proposed by Venkatesh, Ravi, Prinzie and Vanden Poel [2] clustered ATMs into similar patterns by using Taylor-Botyna algorithms in 2014. For each cluster, they used GRNN (Generalized Regression Neural Network)

✉ Soodabeh Poorzaker Arabani
soodabeharabani@gmail.com

1 Department of Engineering, University of Kashan, Kashan, Iran

and MLFF (Multilayer Feed-Forward) and GMDH (Group Method of Data Handling) and WNN (Wavelet Neural Network). Their work results showed that GRNN can obtain the best results according to SMAPE (Symmetric Mean Absolute Percentage Error) test. The SMAPE amount of this method was 18.44%.

Simutis et al. [3] used ANN and SVR techniques and found that the ANN has a better performance than the SVR method. Their data included 15 ATMs and their training collection was the data related to the two years. The MAPE values were between 15 and 28% for ANN and between 17 and 40% for SVR.

Catal et al. [4] considered the increase of the features of dataset to improve the prediction results in their study. In this study, 19 special days such as Mother's Day and New Year's holiday were added to the dataset. The dataset used in their study included 735 days from 111 ATMs in the UK. Their method achieved 21.57% of SMAPE for 56 days forecast after the preprocessing step and exponential smoothing. The findings showed that their method can achieve precise prediction results.

Simutis et al. [5] provided two different methods for predicting daily cash demand for ATM. The first method was flexible ANN. The general characteristics of the ANNs are adaptive adjustment. In the second method of prediction, they used the Support vector machine (SVM) algorithm. In this research, ANN provided a better predictive capacity than SVM.

Ramirez and Acuna [6] reported that MLPNN[1] represents the best performance in ATM cash demand forecast. In their research, they performed ATM cash demand using nonlinear dynamic models AR with external inputs (NARX[2]) and nonlinear ARIMA[3] with external inputs (NARMAX[4]) by ANN and LS-SVM[5] and used to predict one step (OSA) and multistep (MPO). Their purpose was to compare the performance of these two methods to find out which outcomes were the best. They found that the MLPNN had the best output, with an average accuracy of 87% for NARX-OSA and 85% for NARX-MPO. In addition, RBF[6] neural network achieved 82% accuracy for both models and finally, LS-SVM had the worst results with 78% of accuracy for NARX-OSA and 70% for NARX-MPO. They did not find any significant differences between the structures of NARX and NARMAX.

Garsia Pedrero and Gomez Gil [7] represented a new neural network architecture that uses the given information of recurrent neural network and wave analysis. They obtained 27% of SMAPE and reported that this method has better performance than the Feed-Forward Neural Network and Recurrent Neural Network.

Darvish [8] developed a method based on fuzzy logic to solve forecasting cash demand for ATM. To improve the forecasting accuracy of ATM cash demand, an IT2FNN[7] was used.

To improve the accuracy of prediction, Zandevakili and Javanmard [9] used neural network—interval fuzzy type II. The average forecasting accuracy of the proposed method was 72.97%.

Broda et al. [10] presented a method for predicting daily ATM cash demand in Agricole Bank. Their method was based on time series and regression for predicting the amount of money that must be placed daily in every ATM. The approaches used for modeling were ARIMA and exponential smoothing. This method is implemented in R-statistical computing environment. The aim of this study was to minimize the costs of Agricole Bank in Serbia. This work was done to find an optimal cost between two opposite methods.

Dandekar and Ranade [11] found that ATM is one of the most popular banking services. Their proposed model used genetic algorithm to distinguish the strategy to refill money for each ATM. They used all transactions during the years of 2011 to 2012 for Ayande Bank in Iran.

Andrawis et al. [12] presented a predicting model to forecast the 111 time series representing daily cash amounts at ATM machines. They combined ANN, linear models and regressions. They used NN5 dataset and predicted the demanded values for the next 56 days. Their method provided an error rate of 18.95% for SMAPE.

Aseev et al. [13] developed a method to predict ATM cash withdrawal time series. Their hybrid model was based on two methods: Holt-Winters method and Markov chains model. This combination was based on weight factors calculated by each model. The model Holt-Winter predicts time series based on seasonal and trend variables. The Markov chain was able to predict patterns of base time series, such as peaks and holes. The combination of these two methods had better predictive outcomes compared to the situation where each method is used separately.

3 Problem Statement

As presented in the previous section, we have performed a comprehensive literature review on ATM cash demand forecasting. From the literature review, we found out that few researches are conducted to investigate the issue of ATM cash demand forecasting in Iranian banks. The policy that is commonly used by the world's banking system is that the

1 Multilayer perceptron neural network.

2 Nonlinear auto-regressive exogenous.

3 Auto-regressive integrated moving average.

4 Nonlinear auto-regressive moving average with exogenous.

5 Least squares support vector machine.

6 Redial basis function.

7 Interval Type-2 Fuzzy neural network.

bank offers this type of service through contracts with service companies and by paying remuneration to complement the ATMs store and freight costs. These companies are responsible for this task. Researches around the world show [3] that these costs are approximately 35–60% of the total cost of setting up an ATM.

We are confronting a subject that has double behavior, if an ATM faces cash shortage, the reputation of commercial banks will be in jeopardy and customers who cannot receive service from the bank will be dissatisfied. On the other hand, if the accumulation of cash takes place in an ATM, by the given rate of inflation in Iran, confinement of cash will cause loss in the interest of the bank. Therefore, the aim of this research is to accurately predict the amount of cash demand from country banking network ATMs. This will lead to the decrease in the maintenance double costs and ATMs cash completion and the increase in the customers' satisfaction level in using these types of services.

According to the data from the ATMs that is on a daily basis, we can expect the time-dependent behavior from ATMs. On the other hand, because our goal is to predict the amount of all ATMs cash demand of banking networks simultaneously, the data that are used to estimate and predict can be considered as panel data.

Our dataset contains the ATMs daily withdrawal information in the banking network. Inputs are entered into the predictor system as a matrix. Each matrix column includes the cash withdrawal from the ATMs based on one day of the year. Columns are arranged according to the days of the year. Each matrix row contains an ATM cash withdrawal information during the days of the year. Rows are arranged in the matrix based on the geographic neighborhood of ATMs. Because special days and holidays affect the amount of cash withdrawals from ATMs, and on the other hand, since many holidays and special days in the country are religion, and the number of days in lunar and solar year is not the same. For this reason, two years of information has been used for training, until the system detects these changes and has a good prediction.

In this study, the methods which are used to estimate and forecast panel data are:

(1) Statistical traditional ATM cash demand amount panel data forecasting method, and (2) Smart ATM cash demand amount panel data forecasting method by using artificial neural network, SVM and CNN. We show the overall research process in Fig. 1.

First, it is necessary to describe the panel data. Panel data are a set of data that consists of several sections and a period of time. In this research, the meaning of the section is the same as ATM and we show the number of sections with n. The period of time can be a day, a month, a year, the time period in this study is days of the year. We show the period of time with T. Therefore, we show the present observations in the training data with X_{it} where the sections include $i = 1, 2, \ldots, n$ and time includes $t = 1, 2, \ldots, T$. It means that we have totally $n \cdot T$ data. As the panel data reflect both time and inter-sectional changes, it may include additional information. Most of the information that is ignored in the analysis of time series such as heterogeneity that is to be ignored in the analysis of time series, is reviewable in the panel data analysis. Hence, if in ATMs cash demand prediction, we only use the time series data, we will assume in this case that ATMs are homogeneous. In addition, if we only use the ATMs sectional data, we have only noticed ATMs differences in a specific time and we have not given attention to the changes of time. Panel data let us review possibility in both changes (time and sectional).

By using statistical traditional method in the panel data forecast related to country banking network ATMs, at first, we will discuss the correct estimation of regression model parameters. Then, we predict the future amount of ATMs cash demand by using a regression model. In addition, another way that we examine the panel data forecast related to banking network ATMs cash demand is the use of intelligent methods. Our aim of this study is to compare the proposed method in terms of accuracy in prediction and finding the best way to forecast panel data relating to the ATMs cash demand.

4 The Use of Traditional Statistical Methods to Predict

Due to the presence of different types of panel data, there are various statistical traditional methods for estimating and forecasting regression model. Panel data regression models are as follows [14]:

Integration model

Integration model indicates that there are no personal effects and all groups are the same, so the regression equation can be written as follows.

$$Y_{it} = a + BX_{it} + \varepsilon_{it} \quad i = 1, \ldots, n \quad t = 1, \ldots, T \tag{1}$$

This model can be estimated by Ordinary Least Squares (OLS). For this purpose, it is necessary that the sum of squared errors be minimized, $\mathrm{RSS}_{\mathrm{pooled}}$.[8]

$$\begin{aligned}\mathrm{RSS}_{\mathrm{pooled}} &= \sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it} - \hat{y}_{it})^2 \\ &= \sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it} - \hat{a} - \hat{B}X_{it})^2\end{aligned} \tag{2}$$

[8] Residual sum of square.

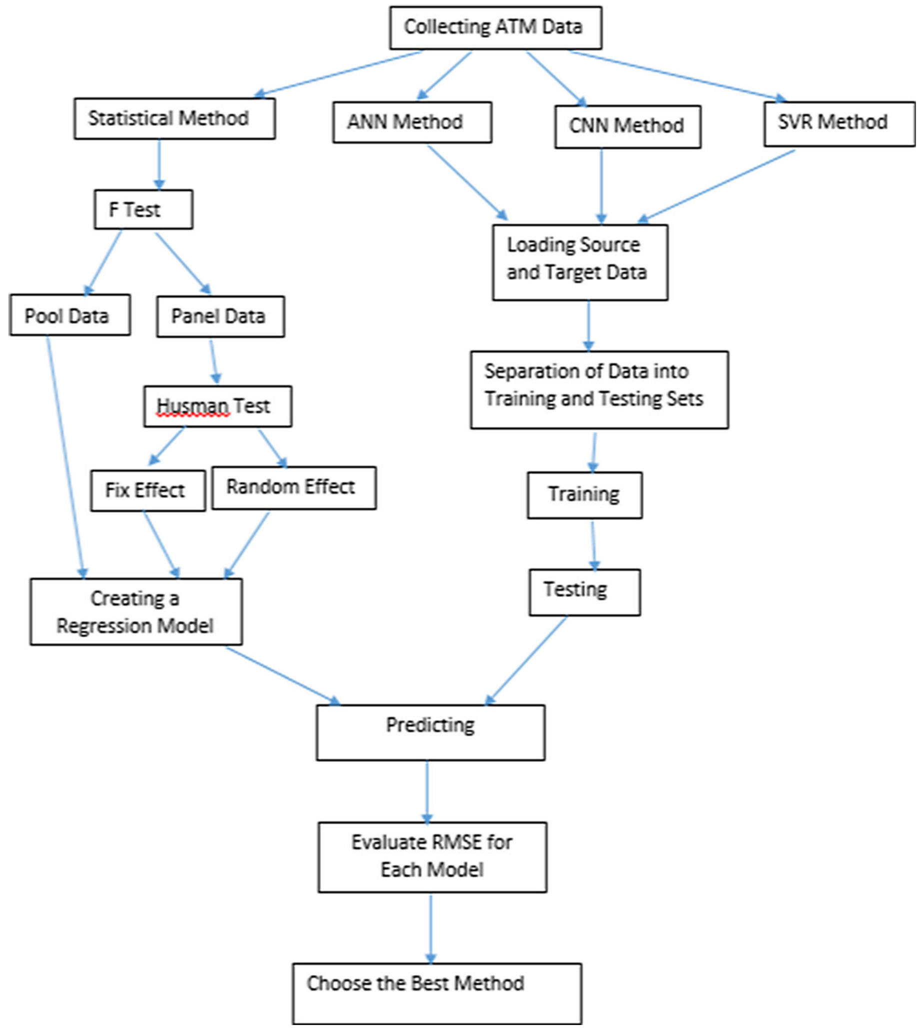

Fig. 1 The overall research process

With the above-mentioned relation derivation related to $\hat{a}$ and $\hat{B}$, we will have:

$$\frac{\partial \mathrm{RSS}_{\mathrm{pooled}}}{\partial \hat{a}} = 0 \Rightarrow$$
$$\sum_{i=1}^{n} \sum_{t=1}^{T} (Y_{it} - \hat{a} - \hat{B} X_{it}) = 0$$
$$\frac{\partial \mathrm{RSS}_{\mathrm{pooled}}}{\partial \hat{B}} = 0 \Rightarrow$$
$$\sum_{i=1}^{n} \sum_{t=1}^{T} (Y_{it} - \hat{a} - \hat{B} X_{it}) X_{it} = 0$$

By solving these equations, integration estimators are obtained.

$$\hat{a}_{\mathrm{pooled}} = \bar{Y} - \hat{B} \bar{X} \quad (3)$$
$$\hat{B}_{\mathrm{pooled}} = \frac{\sum_i \sum_t x_{it} y_{it}}{\sum_i \sum_t x_{it}^2} \quad (4)$$

$\bar{X}$ and $\bar{Y}$ are total averages and lowercase letters show deviation from the average.

$$\bar{\mathrm{X}} = \frac{\sum_{\mathrm{i}} \sum_{\mathrm{t}} \mathrm{X}_{\mathrm{it}}}{\mathrm{nT}} \quad \bar{\mathrm{Y}} = \frac{\sum_{\mathrm{i}} \sum_{\mathrm{t}} \mathrm{Y}_{\mathrm{it}}}{\mathrm{nT}} \quad (5)$$
$$\mathrm{x}_{\mathrm{it}} = \mathrm{X}_{\mathrm{it}} - \bar{\mathrm{X}} \quad \mathrm{Y}_{\mathrm{it}} = \mathrm{Y}_{\mathrm{it}} - \bar{Y} \quad (6)$$

Fixed effects model

In fixed effects model, it is assumed that we can reflect subjective and group differences. Each a_i is an unknown coefficient which should be estimated. a_i represents the effect of all the factors that affect Y_{it} sectional. But the effect of these factors is constant over time. Suppose that Y_i and X_i include observation T for ith group. In this case, we have the following equation.

$$Y_{it} = B X_{it} + a_i + \varepsilon_{it} \quad (7)$$

a_i is different for each group. In order to estimate a_i, a virtual variable for each group is defined; so, by using virtual variables, we can write the model as follows:

$$Y_{it} = BX_{it} + a_1 D_1 + a_2 D_2 + \cdots + a_n D_n + \varepsilon_{it} \tag{8}$$

For example, D_1 is equal to 1 for group 1 and is zero for other groups. D_2 is equal to 1 for the second group and equal to zero for other groups, too. Now, the model coefficients can be achieved by minimizing the sum of squared errors. Because this model is known as the fixed effects model, we show its RSS by $\mathrm{RSS_{FE}}$. Also, because the virtual variables are used in this model, the OLS method is used to estimate its coefficients. Therefore, it is also named "Least Squares Dummy Variables" method (LSDV). Finally, we write the sum of squared errors for the model as:

$$\begin{aligned}\mathrm{RSS_{LSDV}} = \mathrm{RSS_{FE}} &= \sum_i \sum_t e_{it}^2 \\ &= \sum_{i=1}^{n} \sum_{t=1}^{T} (Y_{it} - \hat{a}_i - \hat{B} X_{it})\end{aligned} \tag{9}$$

With derivation related to $\hat{a}_i$ and $\hat{B}$ we will have:

$$\frac{\partial \mathrm{RSS_{FE}}}{\partial \hat{a}_i} = -2 \sum_t (Y_{it} - \hat{a}_i - \hat{B} X_{it}) = 0,$$
$$i = 1, \ldots, n$$
$$\frac{\partial \mathrm{RSS_{FE}}}{\partial \hat{B}} = -2 \sum_{i=1}^{n} \sum_{t=1}^{T} (Y_{it} - \hat{a}_i - \hat{B} X_{it}) X_{it} = 0$$

By solving the above equations, $\hat{a}_i$ and $\hat{B}$ are obtained.

$$\hat{a}_i = \overline{Y_{io}} - \hat{B} \overline{X_{io}} \tag{10}$$

$$\hat{B}_{\mathrm{LSDV}} = \frac{\sum_i \sum_t (X_{it} - X_{io}) \left(Y_{it} - \overline{Y_{io}}\right)}{\sum_i \sum_t (X_{it} X_{io})^2} \tag{11}$$

where

$$\overline{Y_{io}} = \frac{\sum_t Y_{it}}{T} \tag{12}$$

$$\overline{X_{io}} = \frac{\sum_t X_{it}}{T} \tag{13}$$

is group average. It is observed that these results are like integration model except that in which group averages are used instead of total average.

The fixed effects significance test
We use F test for a_i coefficient significance test. In this case, we test whether the effects of different groups are different (a_is are different) or the same (a_is are equal). Thus, the hypothesis is $H0 : a_1 = a_2 = \cdots = a_n = a$. If the hypothesis $H0$ is rejected, we have Eq. (14), that is tested against Eq. (15).

$$Y_t = \sum_{k=1}^{K} B_k X_{KIT} + \sum_{i=1}^{n} a_i D_i + \varepsilon_{it} \mapsto \mathrm{RSS_{UR}}, R_{\mathrm{UR}}^2 \tag{14}$$

$$Y_t = \sum_{k=1}^{K} B_k X_{KIT} + a + \varepsilon_{it} \mapsto \mathrm{RSS}_R = \mathrm{RSS_{pooled}}, R_R^2 = R_{\mathrm{pooled}}^2 \tag{15}$$

The first is *LSDV* regression that defines the differences in the group; therefore, it is called nonconditional regression. The second is integration regression that does not consider group differences and assumes a_i the same, and therefore it is called conditional regression. We calculate *RSS* and R^2 for each of these equations and create F relation.

$$F = \frac{(\mathrm{RSS}_R - \mathrm{RSS_{UR}}) / (n-1)}{\mathrm{RSS_{UR}} / (nt - k - n)} = \frac{nT - k - n}{n - 1} \cdot \frac{R_R^2 - R_{\mathrm{UR}}^2}{1 - R_{\mathrm{UR}}^2} \tag{16}$$

If F is large, it means that $H0$ hypothesis is rejected, therefore, fixed effects are meaningful and a_is are not the same. In other words, subjective and group differences are significant. If the F test probability is smaller than 0.05, $H0$ assumption is rejected and integration model is not appropriate to regression model and we should use the panel model. In the panel model, we should study "If the model has fixed effects in used data or random effects for which we can use Husman test". Husman test is suggested as hypothesis $H0 : E(uiXit) = 0$ that says the random effect is unfeasible if the hypothesis is rejected. We face fixed effects in panel data related to ATMs.

The following results have been obtained by using statistical methods and tests obtained from above on banking network ATMs panel data with the help of Eviews software in the period of 1394. Coefficients of the regression model are estimated for the data in Table 1, which shows that taken cash from ATMs more effects in December, January and February to predict March taken cash. In addition, F test results performed by the Eviews software on the data and the F statistical test that represents zero value show that $H0$ hypothesis of F test is rejected and the suitable regression model for our data is panel model and integration model is not a complete regression method for ATM data effects explanation (Table 2). The prediction results are shown in Fig. 2 by using statistical traditional method. According to the results of prediction by Eviews software, error percentage is estimated 12.48778.

5 The Application of ANN to Predict

The next technique which this study uses is ANN. We compare its performance with the traditional statistical methods in term of accuracy. A comparison of traditional methods of

Table 1 Coefficients of the regression model

Variable	Coefficient	SE	*t* Statistic	Prob
C	19,520,740	4,972,932	3.925398	0.0001
Series01	− 0.021163	0.021750	− 0.972998	0.3309
Series02	0.064472	0.038799	1.661694	0.0970
Series03	0.028709	0.041849	0.686003	0.4929
Series04	0.016833	0.008746	1.924582	0.0547
Series05	− 0.012272	0.005212	− 2.354463	0.0188
Series06	− 0.038198	0.015112	− 2.527746	0.0117
Series07	− 0.001077	0.005974	− 0.180338	0.8569
Series08	− 0.290700	0.064997	− 4.472524	0.0000
Series09	0.150221	0.064331	2.335140	0.0198
Series10	0.183836	0.080849	2.273811	0.0233
Series11	1.104875	0.076654	14.41385	0.0000

Dependent variable: SERIES12
Method: panel least squares
Date: 01/09/17
Time: 19:20
Sample (adjusted): 1 1054
Periods included: 763
Cross sections included: 1
Total panel (balanced) observations: 763

prediction and neurological predictors has begun since the 1990s, attracting the attention of many researchers [15] in the field of computational intelligence. Unfortunately, many of their reports have been contradictory, because their results are based on empirical studies and using different structures of the neural network with different performance functions. In addition, these results were obtained through experiments on different data that was not comparable to each other. However, it seems reasonable to conclude that using a nonlinear neural network model the prediction of linear data collected cannot produce better results than the prediction using linear statistical algorithms. In the opposite way, by using nonlinear data, predicting using a nonlinear neural network can have better results than traditional linear algorithms.

ANNs attempt to model the human understanding through simulation learning process. Intelligent ANNs are a suitable tool to predict the information that may be linear, nonlinear and nonstructural. One of the applications of ANNs can be in ATMs cash demand prediction. With the current form of prediction, ATM's maintenance costs, transportation and completion store will be so costly that it affects an important aspect of the bank's profitability.

An artificial neural system should learn ATM's related information as inputs that are in panel form; and also should learn real requirements of ATM cash as optimal output. The general process of using ANNs in solving problem of ATMs cash demand prediction should include the following steps:

Collect all information in one place
The process of training ANN involves paying attention to all useful or potentially useful information about that issue and includes the use of this information to predict a behavior

Table 2 *F* test results performed by the Eviews software

R-squared	0.971426	Mean dependent var	5.14E+0
Adjusted R-Squared	0.971007	SD dependent var	5.61E+0
SE of regression	95452977	Akaike info criterion	39.6017
Sum Squared resid	6.84E+18	Schwarz criterion	39.6747
Log likelihood	− 15096.07	Hannan–Quinn criterion	39.6298
F statistic	2321.021	Durbin–Watson statistical	2.09191
Prob (*F* statistic)	0.000000	–	–

Cross section fixed (dummy variables)
Effects specification

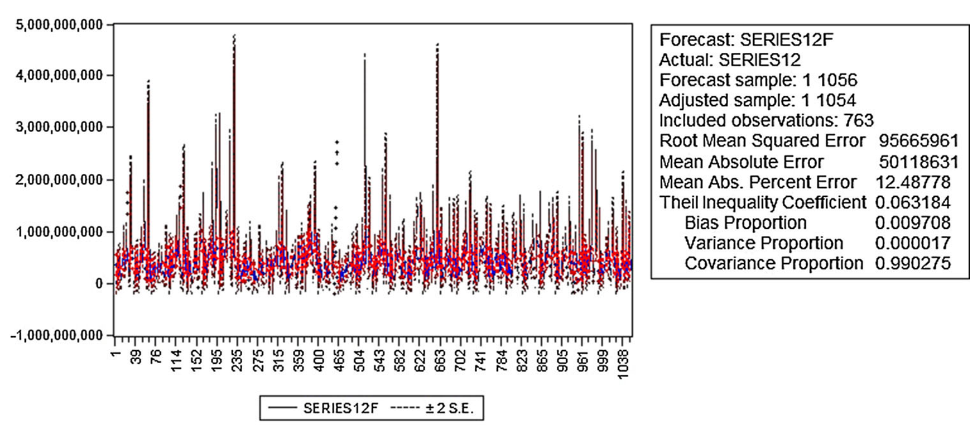

Fig. 2 The results of prediction by Eviews software

or other characteristics. All information should be brought together in order to produce training and testing datasets. To develop a neural network model, the information about ATMs should be collected as a database. A summary of the information related to ATMs includes: ATM names, ATM place addresses, money capacity, money withdrawal in a day.

Separate data for training and testing sets
In the general case for ANNs, an error backpropagation technique will create an ideal organic training set for each of the possible outputs and the ideal test set is the representation of data. In general, a set is extracted for testing experiments for the ANN in the first stage. Then, a training set is selected from remaining samples, that in our model, 80 and 20% of panel data are, respectively, selected for training and testing sets.

Convert data to appropriate inputs for ANN
ANNs usually work with numerical data. Each of the fields in the database must be converted to one or more inputs of the network and each to the appropriate interval. The purpose of this section is to find a way to design that the database fails to enter the ANN. The collected database contains two types of data, which are: (1) numerical data such as money capacity and money taking rate in a day, and (2) character data such as an ATM name and the place address.

The required information for ANN is the amount of taken cash during the days of the year, which is considered as an input for the network; and information such as the name and the place address can be useful to collect additional information, but it cannot be used as an input for the network, this information cannot be used to predict the amount of ATM cash demand.

Training and testing of ANN
Multilayer ANN is considered with error backpropagation for predictive modeling. The selection of a suitable architecture for the neural network is one of the areas [16] which much research has been conducted in this context.

ANN training with backpropagation is composed of two main paths. The first path is called “went path” that in this direction the data is applied to the neural network and its impacts spread through intermediate layers to the output layer. The output that is composed in the output layer is the network actual respond and is considered on went path of constant and unchangeable network parameters. The second path is return path. In this path, parameters are set in the opposite of went path. This adjustment can be set according to error correction rule. Error signal is formed in the output layer. The amount of error is distributed in the whole network after being calculated in the return path from output layer and through the network layers. As this recent distribution is done in the opposite direction of synapse scale communications, the error backpropagation word is selected to explain the network behavioral modification. Network parameters are adjusted in a form that the network actual answer will be closer to the optimal response.

In this study, we use the multilayer perceptron neural network. The multilayer perceptron neural network consists of an input layer, an output layer and one or more hidden layers that are located between the input and output layers. There are a number of simple computational units called neurons in each layer. The network has a hierarchical structure and the number of neurons in each layer depends on the subject and type of work. Multilayer perceptron neural network, if well trained, is a fairly good model for nonlinear data modeling and forecasting. The learning method used in our model is the rule of backpropagation error. The common processing for this learning involves three actions which are: calculating outputs, comparing outputs with optimal answers, adjusting weight and bias values and repeating processing.

Learning by putting weights and bias values begin randomly. The difference between the actual output and the desired output is called the error. We must minimize the error. Error reduction is done by changing the weights and bias values. This type of learning is called supervised learning. There is another type of learning called unsupervised learning. In this type of learning, only the input to the network is given and the network is self-organized and no knowledge is used to classify the outputs. For our network, we use an error backpropagation rule that is a supervised learning.

Artificial neural network topology
Two fundamental questions arise in the ANN topology. The first question is how many hidden layers there should be and the second question is how many neurons there should be in each hidden layer. The proper number of neurons should be commensurate with the work involved.

There are two basic ways to optimize the size of the hidden layer, which is called constructive and destructive. In the constructive method, at first, the network is considered without any hidden layer. In this case, inputs are directly linked to the output. Weights are taught to achieve an error at a constant value. Then, a hidden layer is considered with a neuron and we learn neural network. This work is repeated so that there is no error on its training data. When the growing neurons are stopped, the network performance in this new situation is greater than the previous one with fewer neurons and this new condition does not have a change in the performance improvement. In this case, the last added neuron will be deleted. The destructive method is the opposite of the constructive method. In this method, we start with a large number of neurons in the hidden layer. Then, one by one we reduce the hidden neurons and we review network performance so that improvements are not made to the network performance. In our network, we use the constructive

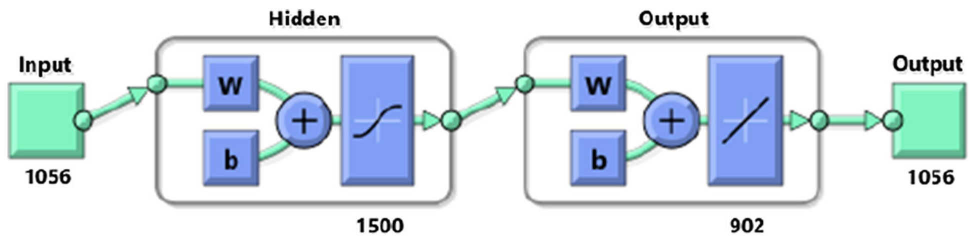

Fig. 3 The ANN topology

method to obtain the number of hidden layer neurons. The ANN network topology is shown in Fig. 3.

Learning rate
The optimal learning rate is at the smooth surface of the error graph. If the error graph in the output layer has a lot of surface change, it shows that the learning rate used is not optimal and it should be reduced to a ratio for all layers of the network. We repeat this step to examine the error graph. Finally, the error graph shows us the best of the situation. Hence, we use this amount for our ANN.

Our dataset contains information on the amount of ATM cash withdrawals per day. Input data have a panel structure. The inputs are entered into the artificial neural network as matrices. The matrix columns of cash withdrawals from ATMs are the days of the year and the matrix rows show the cash withdrawal information of each ATM throughout the year.

In this study, we have used MATLAB software to design the multilayered ANN. Firstly, we define desired input and output information for the neural network in MATLAB software. Input data related to 1056 of country banking network ATMs are in the range of 365 days of a year; and output information is desirable cash demands scale related to these ATMs. By using this information, we train and test the ANN. The designed network consists of 4 layers. The input and output layer contains 1056 neurons and hidden layers of 1500 and 902 neurons, respectively. The results of this design are shown in Fig. 4.

According to the results of ANN experiment with the related panel data to 1056 number of ATMs banking network, which are obtained by MATLAB software, the method achieved the accuracy of about 92.558% for ATMs cash demand prediction. In other words, RMSE error is 7.45% for the test data and 9.69% for the train data. Of course, this error rate is based on the breakdown of information without using cross-validation.

Once again, we have used the ANN model to predict the cash demand of ATMs. This time, we have used the cross-validation method to separate the data. The results obtained from this method have had a significant impact on accuracy. The RMSE error for training data is equal to 1.34% and for testing data is equal to 5.01%.

6 Application of CNN for Prediction

Convolutional neural networks are one of the most important deep learning methods. The convolution neural network consists of three main layers: the convolution layer, the pooling layer and the fully connected layer. Different layers perform different tasks in the network.

There are two stages for training in the convolution network: the feed-forward stage and the backpropagation stage. In the first stage, input is given to the network, and this stage is nothing except a multiplication of the point between the inputs and the parameters of each neuron, and eventually the application of convolution operations in each layer.

$$S(t) = \int x\,(a)\,w\,(t - a)\,\mathrm{d}_a$$

Then, the network output is calculated and for network parameters adjustment or, in other words, in network training, network output is used for failure calculation. To do this, the network output is compared with the correct answer and the failure is obtained. In the next step, based on the calculated failure rate, the backpropagation phase begins. In this step, the gradient of each parameter is computed and all parameters are changed according to their effect on the generated failure in the network. After repeating the appropriate number of these steps, the network training ends [17].

In the general case, the convolution neural network is a hierarchical neural network that its convolution layers and the pooling layers are alternatively, and after that the layer is inserted fully connected. The types of convolutional network layers are:

Convolution layers They include a set of learning able filters. In simple terms, we can say that we face a three dimensional mass. This 3D mass has a length, a width and a depth. If the depth is equal to 1, we are encountering a simple matrix: a two-dimensional array that has length and width. If the depth is equal to two, it means that we have two matrices, and there is a matrix at each depth. Different filters are used in convolutional layers to reduce the number of parameters intensively.

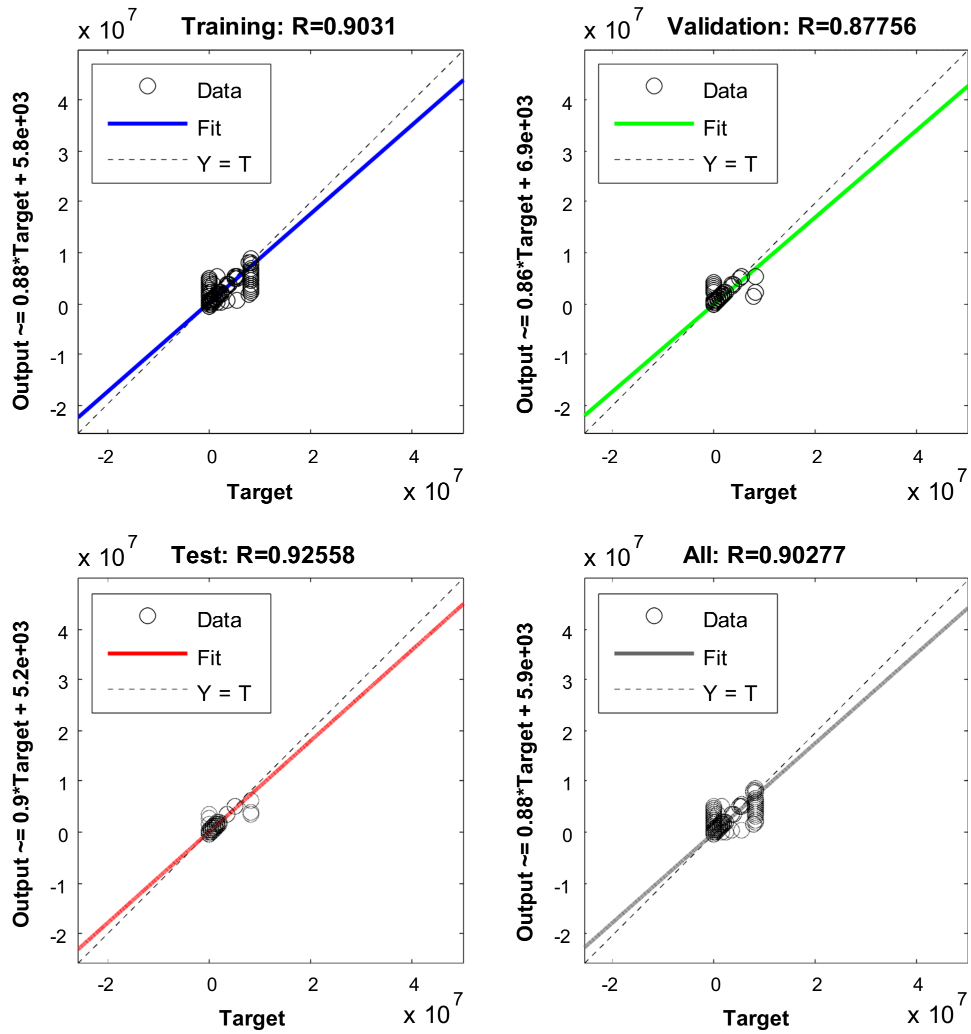

Fig. 4 Regression graphs for ANN model

Pooling layers They are usually inserted after a convolutional layer and it can be used to reduce the size of the network parameters and the feature map. Implementation of the Pooling layers are performed using the Maxpooling function.

The fully connected layer After the last layer of pooling, the fully connected layer is inserted to convert the two-dimensional feature map to the one-dimensional feature vectors. The fully connected layer allows us to give the network result in a vector with a specified size.

In this research, we have used a convolutional network with two convolution and two pooling layers and a fully connected layer. The grid structure is shown in Table 3. First, we separated 80% of the data for network training and 20% for the network test, we have done 60 epoch for network training. The RMSE error rate for training data was 1.47% and it was 9.38% for testing data. Then, we have used the cross-validation method with KFold = 5 to separate the testing and training data. By considering the results of training by this method for training data, the RMSE error was equal to 1.05% and for the test data was equal to 4.47%. It means that the accuracy of the prediction improves significantly to divide the information into training and testing sets by using the cross-validation method.

Table 3 Structure of CNN

In [8]: Model.summary()

Layer (type)	Output shape	Param #
conv1d_1_(Conv1D)	(None, 1, 366)	134,322
max_poolingld_1 (MaxPoolingl	(None, 1, 366)	0
conv1d_2 (ConvlD)	(None, 1, 366)	134,322
max_pooling1d_2 (MaxPooling1	(None, 1, 366)	0
dense_1 (Dense)	(None, 1, 366)	134,322

Total params: 402,966
Trainable params: 402,966
Nontrainable params: 0

Table 4 The results of applied intelligent methods

Method	RMSE_ Train	RMSE_ Test	Time spend on each Epoch
CNN without cross-validation	%1.47	%9.38	3 s, 4 ms
CNN with cross-validation	%1.05	%4.74	3 s, 4 ms
SVM without cross-validation	%15.66	%10.35	–
SVM with cross-validation	%8.87	%13.66	–
ANN without cross-validation	%9.69	%7.45	298 µs
ANN with cross-validation	%1.34	%5.01	313 µs

7 SVR application for prediction

SVR is a class of SVM that tries to minimize overall failure. Regression seeks to find a function that can map input data into real numbers based on learning data. SVR was originally presented by Vapnick et.al in 1997. Support vector machines are a popular learning method for classifying, regression, and other learning operations. SVR is the most common form of SVM application. SVMs project data into a higher-dimensional space and maximize the margin between classes and minimize the margin of error for regression. SVR uses large marginal kernel methods to classify and analyze the regression. The problem of regression is to find a function that can map input data into real numbers, based on learning data, which depends on a difference between the desirable output and the real output [18].

In this study, we have used the SVR model to predict ATM cash demand, which uses the RBF kernel function and gamma parameter equal to 0.1. The dataset was separated into two methods for training and testing. Initially, 20% of the dataset is used for the test and 80% is used for the model training so that the obtained RMSE failure rate for the training set is 15.66% and the RMSE failure rate for the test set is 10.35%. In the next step, we have used the cross-validation with $K\,\text{Fold} = 5$ to separate the information so that the RMSE failure for the training set was 8.87% and the RMSE failure for the test set was 13.66%. Totally, SVR method is not a suitable method in contrast with methods that we mentioned in the previous sections.

8 Conclusions and suggestions

The purpose of this study is to predict the country banking network ATMs cash flow demand by using the statistical traditional method and machine learning intelligent methods, i.e., ANN, CNN and SVR, and the comparison between the results obtained from the employed methods and the introduction of the best way to predict panel data of the country banking network ATMs cash demand rate. In summary, the results are shown in Table 4.

It can be seen that the method of convolutional neural network has shown the best accuracy in this prediction by deep learning and information separation based on cross-validation. After The CNN method, ANN has the best prediction of accuracy using cross-validation. The prediction of accuracy of ANN using cross-validation is very close to CNN prediction accuracy, but the learning speed of ANN is much better than CNN. So, if speed is not important, CNN is the best method among the implemented methods. But if learning speed is important, then ANN is preferred to CNN.

Also, the obtained results in this study have better prediction accuracy than the presented results in the research literature because of two reasons. The first reason is because of the use of large volume panel data, and the other is due to the use of the CNN deep learning method.

It is suggested that we use other deep learning methods for this prediction, such as the deep learning method of the LSTM recursive network [19], the deep belief network [20].

References

1. Jadwal, P.K.; Jain, S.; Gupta, U.; Khanna, P.: K-means clustering with neural networks for ATM cash repository prediction. In: International Conference on Information and Communication Technology for Intelligent Systems. pp. 588–596. (2017)
2. Venkatesh, K.; Ravi, V.; Prinzie, A.; Van den Poel, D.: Cash demand forecasting in ATMs by clustering and neural networks. Eur. J. Oper. Res. **232**, 383–392 (2014)
3. Simutis, R.; Delijonas, D.; Bastina, L.; Friman, J.; Drobinov, P.: Optimization of cash management for ATM network. Inf. Technol. Control. **36**, 117–121 (2007)
4. Catal, C.; Fenerci, A.; Ozdemir, B.; Gulmez, O.: Improvement of demand forecasting models with special days. Proc. Comput. Sci. **59**, 262–267 (2015)
5. Simutis, R.; Dilijonas, D.; Bastina, L.: Cash demand forecasting for ATM using neural networks and support vector regression algorithms. In: 20th International Conference, EURO Mini Conference, "Continuous Optimization and Knowledge-Based Technologies" (EurOPT-2008), Selected Papers, Vilnius. pp. 416–421. (2008)
6. Ramírez, C.; Acuña, G.: Forecasting cash demand in ATM using neural networks and least square support vector machine. In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. pp. 515–522. (2011)
7. Garcia Pedrero, A.; Gomez Gil, P.: Time series forecasting using recurrent neural networks and wavelet reconstructed signals. In:

20th International Conference on Electronics, Communications and Computer (CONIELECOMP), pp. 169–173. (2010)
8. Darwish, S.M.: A methodology to improve cash demand forecasting for ATM network. Int. J. Comput. Electr. Eng. **5**, 405 (2013)
9. Zandevakili, M.; Javanmard, M.: Using fuzzy logic (type II) in the intelligent ATMs' cash management. Int. Res. J. Appl. Basic Sci. **8**, 1516–1519 (2014)
10. Borda, P.; Levajkovic, T.; Kresoja, M.; Marceta, M.; Mena, H.; Nikolic, M.; Stojancevic, T.: Optimization of ATM filling-in with cash. 99th European study group with industry, pp. 1–16 (2014)
11. Dandekar, P.V.; Ranade, K.M.: ATM cash flow management. Int. J. Innov. Manag. Technol. **6**, 343 (2015)
12. Andrawis, R.R.; Atiya, A.F.; El-Shishiny, H.: Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition. Int. J. Forecast. **27**, 672–688 (2011)
13. Aseev, M.; Nemeshaev, S.; Nesterov, A.: Forecasting cash withdrawals in the ATM network using a combined model based on the holt-winters and markov chains. Int. J. Appl. Eng. Res. **11**, 7577–7582 (2016)
14. Brooks, C.: Introductory Econometrics for Finance, pp. 487–509. Cambridge University Press (2014)
15. Palit, A.K.; Povovic, D.: Computational Intelligence in Time Series Forecasting, pp. 129–130. Springer (2005)
16. Theodoridis, S.; Koutroumbas, K.: Pattern Recognition, pp.151–196. Elsevier (1999)
17. Goodfellow, I.; Bengio, B.: Gurville A.: Deep Learning. MIT Press, Cambridge (2016)
18. Anandhi, V.; Manicka Chezian, R.: Support vector regression in forecasting. Int. J. Adv. Res. Comput. Commun. Eng. **2**(10), 4148–4151 (2013)
19. Siami Namini, S.; Siami Namini, A.: Forecasting Economics and Financial Time Series: ARIMA vs. LSTM. (2018). arXiv:1803.06386
20. Zheng, J.; Zhang, G.: Research on exchange rate forecasting based on deep belief network. Neural Comput. Appl. (2017). https://doi.org/10.1007/s00521-017-3039-z