Forecasting Cash Withdrawals in the ATM Network Using a Combined Model based on the Holt-Winters Method and Markov Chains

Mikhail Aseev, Sergei Nemeshaev, and Alexander Nesterov
1National Research Nuclear University MEPhI, Department of Cybernetics
Kashirskoye Highway, 31, Moscow, 115409, Russian Federation.

Abstract
The article describes the method of forecasting time series of cash withdrawals in ATMs. The hybrid model is based on two methods: Holt-Winters additive method and Markov chain-based model. The combination occurs with the help of weight coefficients which are calculated on the basis of work of each model. Holt-Winters method forecasts time series with trend and seasonal variations. Markov chain enables to forecast patterns of basic time series, such as peaks or holes. The composition of these two approaches will allow banks and other financial organizations to predict cash withdrawals more accurately than the methods used separately.

Keywords: Markov Chains, Holt-Winters Method, Time Series, Forecasting, Statistics

INTRODUCTION
Many companies have accumulated the statistical data that can be used to solve a range of tasks related to business analytics and forecasting. The examples of such tasks include identifying needs for goods, defining market trends, forecasting future sales, and other similar tasks. The information obtained in the consideration of these problems is also applied for planning logistics, development of the company strategy, and evaluation of financial risks. The efficiency of this problem solving depends directly on the quality of the data analysis and the forecast accuracy. The banking sector is one of the economic sectors where the application of business forecasting methods is required. Currently, most banks are expanding their ATM networks [1], and the question of minimizing network maintenance costs is becoming increasingly vexed. The mechanisms enabling to determine dates and amount of each ATM network replenishment are required to optimize cash collection. It is evident that this can be only possible if the definition of future cash withdrawals is made with a high accuracy. Currently, this problem is covered in different scientific articles, which gives evidence of its relevance [2-4]. Different forecasting methods help solving the problem of future demand determination. The Holt-Winters model can be distinguished as one of the most common problems among them [5-7]. It was designed as a fast method, using a small memory space, but at the same time quite efficient method of sales volume forecasting, taking into account the trend and seasonality [8]. Another interesting method of forecasting future values is to use a model based on the Markov chains. Assume that the cash withdrawal in ATMs is a random process. In this case, you can use the Markov chains to forecast future values in a series. This approach was also used in different studies [9, 10]. However, the experience has proven that the use of any method does not necessarily always give good results. The recent research often focuses on a combination of several models [11, 12]. The results of these studies show that this approach is efficient and produces better results than any method used separately.

There are different forecasting methods using a combination of several models, for example, adaptive composition of models [8]. In the case of method composition, the forecast is made as a weighted sum of the forecasts calculated according to several individual models. In this case, the weight coefficients are calculated based on the past history of each method, depending on how adequate the data was in each model. This study describes the Holt-Winters model and the forecast method based on the Markov chains, analyzes their possible combinations, and describes the suggested method based on the adaptive method composition. The rest of the work is organized as follows. Section 2 consists of summary theoretical information on the methods forming the basis of the suggested method. Section 3 comprises a description of preliminary data finalization and a description of the proposed hybrid model. Section 4 summarizes the numerical results of the model operation and compares different methods. The study results are given in Section 5.

THEORY
Holt-Winters additive method
The Holt-Winters method enables to forecast time series, taking into account the available trend and seasonal variations. There are two types of this approach, taking into account the additive or multiplicative seasonality. This study uses the additive season model as one of the techniques involved in the combination. Such choice has been made during the preliminary data analysis, including decomposition of the original series and evaluation of the selected components. The description of the model under the terms of cash collection forecast is given below.

Given the time series \( y_1, \ldots, y_t \) of cash withdrawals from the ATM for \( t \) days, \( y_t \) is the last known value of cash withdrawals per day \( t \). It is necessary to calculate the forecast \( \hat{y}_{t+p} \) for \( p \) days ahead. If \( s \) is a duration of the seasonal variation period, the Holt-Winters additive model may be written using the following four equations:

\[
\begin{align*}
\hat{y}_{t+s+p} &= \hat{y}_{t+s} + (y_t - \hat{y}_{t+s})/s \\
\hat{y}_{t+s} &= \hat{y}_{t+s-1} + \alpha (y_t - \hat{y}_{t+s-1}) \\
\gamma_{t+s} &= \gamma_{t+s-1} + \frac{\beta (s - 1)}{s} (\hat{y}_{t+s} - \hat{y}_{t+s-1}) \\
\delta_{t+s} &= \delta_{t+s-1} + \frac{\beta (s - 1)}{s} (\hat{y}_{t+s} - \hat{y}_{t+s-1})
\end{align*}
\]
\[ L_t = \alpha (y_t - S_{t-1}) + (1 - \alpha)(L_{t-1} - T_{t-1}), \]

\[ T_t = \beta (L_t - L_{t-1}) + (1 - \beta)T_{t-1}, \]

\[ S_t = \gamma (y_t - L_t) + (1 - \gamma)S_{t-1}, \]

\[ \hat{y}_{t+p} = (L_t + pT_p)S_{t-s+p}, \]

where
- \( L_t \) — smoothed value of the current series level;
- \( T_t \) — trend assessment;
- \( S_t \) — seasonality assessment;
- \( \alpha \) — smoothing constant for the current series level (0 ≤ \( \alpha \) ≤ 1);
- \( \beta \) — smoothing constant for trend assessment (0 ≤ \( \beta \) ≤ 1);
- \( \gamma \) — smoothing constant for seasonality estimation (0 ≤ \( \gamma \) ≤ 1).

The model assumes that you know the length of the period of seasonal variations. In the course of the spectral analysis of source time series, the frequency of each of them was determined as 7 days. So the period of seasonality is one week.

The weights \( \alpha, \beta, \gamma \) can be selected either subjectively or by minimizing the forecast errors. The higher values of smoothing parameters are taken, the faster the change response will be. This means that the recent data will contribute more to the forecast result. The less the weights are, the weaker the model response to changes will be. In this case, random deviations will be smoothed, and the predicted values will be more stable. [13]

In this study, the parameters \( \alpha, \beta, \gamma \) are determined by minimizing the Mean Squared Error or MSE:

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

Markov chains
Markov chain is a random process, i.e. a sequence of random events, which is characterized by the fact that the probability of transition to the next state only depends on the current state that is on the present, and not on the past. An event in this case means transition from one state to another state.

The process is in one of \( n \) states at any time. A set of states \( S = \{s_1, ..., s_n\} \) is finite or enumerable. The probability of transition from state \( s_i \) to state \( s_j \) is defined as \( p_{ij} \).

The following conditions shall be met:

\[ p_{ij} \geq 0, \]

∀ \( i \) ∑ \( j \) \( p_{ij} = 1. \) (8)

In this case the matrix \( P \) is called stochastic.

At any time, Markov chain \( t \) can be characterized as a row vector \( p(t) = (p_{i_1}^{(t)}, ..., p_{i_n}^{(t)}) \), where \( p_{i_k}^{(t)} \) is the likelihood that in step \( t \) the process is in states \( s_k \). It can be also designated as a probability vector.

Suppose \( p(t) \) is a probability distribution of the current step \( i \), then it is necessary to multiply the current probability vector by the transition matrix in order to find out what condition the chain will be in next time,

\[ p(t+1) = p(t) \times P. \] (9)

If you take into account the associativity of the matrix composition, you need to multiply \( p(t) \) by the transition matrix raised to the power \( t \) in order to find out the probability distribution in \( t \) steps:

\[ p(t+1) = p(t) \times P^t. \] (10)

Then if you introduce a process as the Markov chain, you can simulate it in time and build some forecast. But to do this, you first need to decide how you will define the process state sets, and how you will make the transition matrix, as well as some other issues.

In this paper, the approach based on the Markov chain is used to forecast the future demand for the ATMs and it is described in Section 3.2.

OFFERED METHOD
Data pre-processing
The transactional ATM statistics is the daily information on cash withdrawals or deposits, including the information on the currency in which the transaction was conducted, and the current balance.

The primary statistics used to test the models is the daily data on cash withdrawals in 11 ATMs, owned by one of Ekaterinburg banks, for 8 months.

The problem of missing values was solved in the ATM transaction statistics as part of data pre-processing. The lack of information appears to have resulted from the fact that one day the ATM could be turned off or not work for technical reasons, and the statistics is missing for that day. This situation is unusual.

The time series of cash withdrawals consists of a small amount of omissions (on an average, 10 omissions fall on 244 values of the basic series). Therefore, it is possible to restore the missing data, replacing it with the predicted values. In this case, the forecast of the first missing value is based on the previous statistics. Any further omissions are treated in the same way, and all previous missed values have been already replaced with some values and included in the forecast calculation of the current missing value.

In this study, the Holt-Winters additive model was used to determine the missing values.

Markov chain-based model
First, it is necessary to define a set of states that the process can accept in order to apply the method based on the Markov
chains and forecast the ATM demand. Suppose the statistics of an ATM was selected for a period equal to the training set. Group the number of cash withdrawals to have a certain set of intervals. Denote:

\[ y_{\text{min}} \] - minimum withdrawal from the ATM during the period;

\[ y_{\text{max}} \] - maximum withdrawal from the ATM during the period.

Choose a step \( \Delta \). Then the set of chain states can be defined as follows:

\[
s_1 = [y_{\text{min}}; y_{\text{min}} + \Delta)
\]

\[
s_2 = [y_{\text{min}} + \Delta; y_{\text{min}} + 2\Delta)
\]

\[
\vdots
\]

\[
s_i = [y_{\text{min}} + (i - 1)\Delta; y_{\text{min}} + i\Delta)
\]

\[
\vdots
\]

\[
s_n = [y_{\text{min}} + (n - 1)\Delta; y_{\text{min}} + n\Delta)
\]

where \( n = \left[ \frac{y_{\text{max}} - y_{\text{min}}}{\Delta} \right] \) (brackets \([\ ]\) mean rounding up to the nearest integer upward). Each state is a half-interval, and a set of all these intervals covers completely the basic series of cash withdrawals. To move from the series of cash withdrawals to the series of states, it is necessary to replace every withdrawal \( y_t \) with a certain state. In this case \( y_t \) is replaced with \( s_i \) when \( y_t \in s_i \). After the state sets have been determined, the next problem that has to be solved is how to build a matrix of transition probabilities. Build the transition matrix based on the training set as well. Calculate the number of transitions from state \( s_i \) to state \( s_j \) for all \( i,j \), and define it as \( m_{ij} \). Then calculate the probability of transition from \( s_i \) to \( s_j \):

\[
p_{ij} = \frac{m_{ij}}{n}
\]

If the original statistics does not have transitions to state \( s_k \), define

\[
p_{kj} = \frac{1}{n}, j = 1, \ldots, n.
\]

This decision does not have an effect on the calculation of the transition probability in one or another state. It is required only to fulfill the obligatory conditions of the stochastic matrix:

\[
\forall i \sum_j p_{ij} = 1.
\]

Depending on the source statistics and the selected step \( \Delta \), the transition matrix \( P \), composed of \( p_{ij} \) will be “completed” to a greater or lesser degree. The less the initial data and less the step \( \Delta \) are, the more “unimportant” lines (lines which you cannot go into) are, including the lines, which consist of one unit and zeros.

Now, when you have determined the state system and have built the matrix of transition probabilities, we can predict, what condition the process will be in over time. If \( y_t \) is the latter known value of cash withdrawals per day \( t \), and \( p^{(i)} \) is a current probability vector, the probability distribution in the chain states at the moment \( t + t \) is calculated according to the known formula:

\[
p^{(i+1)} = p^{(i)} \times P^t.
\]

Based on this approach, two models were built as follows: the first one takes into account the transitions from one state to another state per day, and the second one-per week. In the latter case, the basic series was divided into 7 series corresponding to each day of the week, and the whole process was repeated for them. The second approach is supposed to better predict the seasonal component of the series.

Combining the Holt-Winters and the Markov chain models

When several models have been built, it is necessary to determine their combination method. Given a time series \( y_1, \ldots, y_t \) of the daily ATM cash withdrawals, there are \( k \) forecast models, and \( \hat{y}_{jt+d} \) is a forecast of the \( j \)-th model for \( d \) days ahead. First, choose a criterion for selecting the best models in a step. In this study, it is suggested to use weight coefficients for evaluating the efficiency of the model operation. Suppose \( w_{jt} \) is a weight coefficient of the \( j \)-th model per day. One of the ways of choosing coefficients \( w_{jt} \) is an adaptive weight selection [8]:

\[
w_{jt} = \frac{\left( \bar{e}_{jt} \right)^{-1}}{\sum_{s=1}^{k} \left( \bar{e}_{st} \right)^{-1}}
\]

where \( e_{jt} \) is a forecast error of the \( j \)-th model per day \( t \):

\[
e_{jt} = y_t - \hat{y}_{jt}.
\]

\( \bar{e}_{jt} \) is exponentially smoothed forecast error of the \( j \)-th model per day \( t \):

\[
\bar{e}_{jt} = \gamma \left| e_{jt} \right| - (1 - \gamma)\bar{e}_{jt-1},
\]

where \( \gamma \) is a smoothing constant. The parameter \( \gamma \) needs to be selected, but it is recommendation to take \( \gamma = 0.01, \ldots, 0.1 \) [8]. The point of the weight coefficient is that the less the smoothed error \( e_{jt} \) is, the more weight the \( j \)-th model has per day.

Thus, the coefficients are negative, and their sum for every model is equal to one any day \( t \):

\[
\sum_{j=1}^{k} w_{jt} = 1, \forall t,
\]

A validation sample was selected to implement the combined model in addition to the training and test sets. In this paper, its length was chosen to be 42 days. Taking into account the selected length of the validation sample, that study enabled to define \( \gamma = 0.02 \). The validation sample is used to test the models: weaknesses and strengths of each model are determined and the weight coefficients are calculated on their basis \( w_{jt} \). This value is calculated depending on the adequacy of the model operation with this data. The weight coefficient shows what contribution every model shall make to the forecast results.
RESULTS
Any time series of the statistics used have a length of 244 days (8 months). The first 181 values were taken as a training sample, values from 182 to 223-as a validation sample, and values from 224 to 244-as a test set. Thus, the forecasting horizon is equal to 21 days.

The Mean Absolute Percentage Error or \( MAPE \) was used to estimate the forecast accuracy:

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|
\]

(22)

This method is widely used in similar studies to assess the quality of the forecast model and varies, at an average, from 20% to 45% [14]. \( MAPE \) emphasizes the extent of forecast errors as compared to the actual series values. \( MAPE \) can be also used to compare the accuracy of one and the same or different methods for two absolutely different series.

The test results of the models are given in Table 1 (Appendix A). It is worth making a note. The models in the table are shown as the following abbreviations:

- Holt-Winters additive model — AHW;
- Markov chain-based model per day — MCD;
- Markov chain-based model per week — MCW.

Column HW + MCD + MCW corresponds to a combined model.

The error values of the hybrid model, which are less than the similar error produced in the Holt-Winters additive method, are highlighted in bold for every ATM.

The average value \( MAPE \) using the Holt-Winters additive method is 40.45%. The model tries to identify a trend, and find seasonal variations, taking into account the seasonality specified interval, and build a forecast on its basis. However, due to the fact that the basic time series often have a not clearly pronounced seasonal component, the model does not always accurately repeat the seasonal variations in the test set. In general, the value \( MAPE \) is quite high.

Considering the following models, it is worth noting that the table shows the minimum, maximum and average values \( MAPE \) due to nondeterministic methods based on the Markov chains per day and per week, and their combination with the Holt-Winters model.

As the table shows, most of the approaches based on the Markov chains (per day and per week) work worse than the Holt-Winters method. It was noted in the course of the model analysis based on the Markov chains that the forecast built on their basis has a greater dispersion than the Holt-Winters method, so there is a greater number of big forecast errors. Nevertheless, it is worth noting that the mechanism of Markov chain models enables to predict some patterns of the basic time series (for example, peaks or holes) to a good approximation.

The combination of two methods under consideration gives a better result in some cases. In 8 cases, the average \( MAPE \) is less than a similar error in the Holt-Winters method. And the maximum value \( MAPE \) of 4 ATMs is less than the similar error in the Holt-Winters method. The average value of \( MAPE \) the hybrid model is 35.72%. However, this value is quite high.

It is worth noting that on transition to the model combination, indeterminacy of the forecasting algorithm work decreases. Thus, the mean difference between the best and the worst forecast of the hybrid model is 9.31% (against 28.37% and 26.09% in the models based on the Markov chains per day and per week, respectively).

A series of graphs illustrating the forecast built by each model in question for one of the ATMs to demonstrate the model operation is given below. A thin dotted line in the graphs shows the raw data of the test set. The thick blue line corresponds to the forecast.
CONCLUSION

Two different approaches to forecasting an ATM demand were considered in this paper—Holt-Winters additive method and Markov chain-based method. The first method has been widely used for forecasting time series for a long time. The second approach is less traditional. In the second case the cash withdrawal is shown as a random process. Using the machinery of the Markov chain theory, two models were built, the first of which takes into account the transitions from one state to another state per day, and the second-per week. Despite the less accurate results as compared to the results of the Holt-Winters additive method, this model enables to predict some patterns of the basic time series. Based on both approaches, the hybrid model was built. The testing has shown that the combination of the models in question using the weight coefficients in some cases gives a better result than one single approach. Despite this, the high forecast accuracy could not be achieved. The current value of the average value $MAPE$ 35.72% is still high.

In the course of the further work it is planned to build a more complex model based on the Markov chains to increase the forecast accuracy, taking into account not only the daily and weekly state transitions; determine the model behavior based on the Markov chains using special mechanisms; complicate the weight system used in the method combinations, that is instead of using one weight coefficient for each model, use their more complex combination.

REFERENCES


Picture 4: Forecast of the combined model

Week
<table>
<thead>
<tr>
<th>Model ATM</th>
<th>AHW (%)</th>
<th>MCD Min (%)</th>
<th>MCD Max (%)</th>
<th>MCD Mean (%)</th>
<th>MCW Min (%)</th>
<th>MCW Max (%)</th>
<th>MCW Mean (%)</th>
<th>HW + MCD + MCW Min (%)</th>
<th>HW + MCD + MCW Max (%)</th>
<th>HW + MCD + MCW Mean (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>56.8</td>
<td>46.9</td>
<td>60.9</td>
<td>54.8</td>
<td>34.3</td>
<td>63.7</td>
<td>42.0</td>
<td>39.1</td>
<td>47.4</td>
<td>42.1</td>
</tr>
<tr>
<td>2</td>
<td>25.2</td>
<td>29.9</td>
<td>42.5</td>
<td>35.6</td>
<td>44.3</td>
<td>61.9</td>
<td>53.9</td>
<td>20.6</td>
<td>27.0</td>
<td>23.5</td>
</tr>
<tr>
<td>3</td>
<td>32.3</td>
<td>41.2</td>
<td>54.6</td>
<td>49.2</td>
<td>43.5</td>
<td>75.2</td>
<td>57.5</td>
<td>28.6</td>
<td>34.5</td>
<td>31.7</td>
</tr>
<tr>
<td>4</td>
<td>38.7</td>
<td>35.7</td>
<td>55.9</td>
<td>46.3</td>
<td>58.5</td>
<td>74.8</td>
<td>63.5</td>
<td>38.6</td>
<td>44.5</td>
<td>40.4</td>
</tr>
<tr>
<td>5</td>
<td>46.7</td>
<td>37.9</td>
<td>55.3</td>
<td>45.0</td>
<td>37.1</td>
<td>62.1</td>
<td>53.9</td>
<td>32.7</td>
<td>42.0</td>
<td>38.0</td>
</tr>
<tr>
<td>6</td>
<td>47.0</td>
<td>32.4</td>
<td>58.3</td>
<td>43.5</td>
<td>50.4</td>
<td>75.3</td>
<td>60.6</td>
<td>30.8</td>
<td>41.2</td>
<td>34.7</td>
</tr>
<tr>
<td>7</td>
<td>25.2</td>
<td>23.8</td>
<td>39.4</td>
<td>30.4</td>
<td>36.4</td>
<td>51.3</td>
<td>45.2</td>
<td>20.1</td>
<td>28.1</td>
<td>24.4</td>
</tr>
<tr>
<td>8</td>
<td>48.7</td>
<td>63.7</td>
<td>96.1</td>
<td>75.3</td>
<td>63.6</td>
<td>121.3</td>
<td>86.2</td>
<td>46.1</td>
<td>60.4</td>
<td>50.5</td>
</tr>
<tr>
<td>9</td>
<td>20.5</td>
<td>44.1</td>
<td>61.1</td>
<td>50.7</td>
<td>35.7</td>
<td>61.3</td>
<td>41.7</td>
<td>17.9</td>
<td>27.0</td>
<td>22.6</td>
</tr>
<tr>
<td>10</td>
<td>56.5</td>
<td>38.8</td>
<td>106.1</td>
<td>63.0</td>
<td>46.9</td>
<td>63.2</td>
<td>57.2</td>
<td>36.4</td>
<td>48.4</td>
<td>39.8</td>
</tr>
<tr>
<td>11</td>
<td>47.4</td>
<td>46.6</td>
<td>100.9</td>
<td>66.9</td>
<td>57.3</td>
<td>84.9</td>
<td>68.4</td>
<td>40.8</td>
<td>53.6</td>
<td>45.3</td>
</tr>
<tr>
<td>Average</td>
<td>40.45</td>
<td>40.09</td>
<td>68.46</td>
<td>51.64</td>
<td>46.18</td>
<td>72.27</td>
<td>57.28</td>
<td>31.97</td>
<td>41.28</td>
<td>35.72</td>
</tr>
</tbody>
</table>