Application of Quadratic Exponential Smoothing and Markov Chain in Computer Predicting Total Amount

Hongyun Gao1, Qianyong Tang2, Dan Li1, Shuang Chen1,∗

1College of Information Engineering Dalian University, Dalian China
2School of Economics and Management China University of Geoscience Beijing China

*Corresponding author: shuangchen2020@dlu.edu.cn

Abstract. Deposits are the main source of funds for commercial Banks in China. The growth rate of total bank deposits is relatively stable, which has a crucial impact on the development of national economy and banking industry. Therefore, it is very important to accurately predict the total amount of bank deposits, which can provide a reliable basis for the scientific management of Banks. This paper takes the total deposits of ICBC in each quarter in the past decade as the historical data, which is used to choose smoothing times, initial smoothing values, and optimal smoothing coefficient. Markov chain is combined with second exponential smoothing model, which is used to predicting total deposits of ICBC in 2020.

Keywords: Quadratic Exponential Smoothing Method, Markov Chain, Total Deposits.

1. Introduction
The purpose of commercial bank in China is profit. Commercial bank needs to take deposits and issue loans. Therefore, deposits are very important part for all the commercial banks. After 40 years of reform and opening-up policy, the economy developed rapidly which leads the total deposits increasing. However, the money flew into real estate markets, and online finance is rising since 2010. The growth rate of total deposits is gradually slowing down. Prediction of total deposits is important for projects in future. To stimulate the economy, policy will be adjusted according to the forecast.

Industrial and Commercial Bank of China (ICBC) is one of the largest nationalized banks in China. ICBC is a representative research project.

The prediction of total deposits has been studied by many researchers, using various methods. Most of the methods are divided into several models roughly.

1.1. Regression Model
Regression model takes relationship among total deposits and its influence factors into account [1]. The process is searching the factors which affect the total deposits. All the factors are quantized and normalized. All the factors are rank due to the degree of influence. And main factors are selected as independent variables, and obviously total deposits is the dependent variable. Using regression model
to find function relationship between independent and dependent variables. The function could be used to predict the future total deposits.

The regression model mainly used to predict medium- and long-term macro-forecast.

1.2. Grey Model Theory
Since the total deposit is a complex system, and some factors are known, some are not. The system is considered as grey system [2]. The main factors are not taken into account. Using the data from total deposits sequence, the inherent law and useful information are obtained, which is used to establish the grey model.

The grey model theory mainly used to predict short-term macro forecast. The character of the data is less quantity and variation, and incomplete.

1.3. Time Series Forecasting Method
The method is based on time series. All the total deposits are ranked due to time series. As the time goes on, the total deposits are analyzed and combined with appropriate algorithm, total deposits in future could be estimated in reasonable range [3].

Compared to other method, time series forecasting method is easy to understand. Moreover, it is effective in short-term macro forecast prediction.

Although the methods mentioned above are effective in specific situation, model based on single method still has defect. In the paper, a model is supplied using Markov chain and quadratic exponential smoothing method.

2. Method

2.1. Markov Chain
A random sequence \(\{X(t), t \in T\}, T = \{0,1,2,\ldots\}\), and state space \(S = \{s_0, s_1, s_2, \ldots\}\). Taking any random moment \(t\), and random state \(s_0, s_1, \ldots, s_{t-1}, s_t, s_j\), contingent probability meets

\[
P\{X_{t+1} = s_j|X_t = s_i, X_{t-1} = s_{t-1}, \ldots, X_1 = s_1, X_0 = s_0\} = P\{X_{t+1} = s_j|X_t = s_i\} \tag{1}
\]

The random sequence \(\{X(t), t \in T\}\) is defined as Markov Chain [4, 5].

Transition probability is a very important parameter for probability property of Markov Chain. Assume the step number of random states in sequence \(x_1, x_2, \ldots x_n\) is \(k\), therefore the random state could be shown as \(S = \{s_0, s_1, s_2, \ldots, s_k\}\). If \(f_{ij}\) represents the counts from state \(s_i\) to state \(s_j\), \((s_i, s_j \in S, i, j \in I = \{1,2,\ldots, k\})\). The matrix \(\{f_{ij}\}_{i,j \in I}\) is the corresponding state transfer counts matrix. Then normalize the matrix \(\{f_{ij}\}_{i,j \in I}\) using equation:

\[
p_{ij} = \frac{f_{ij}}{\sum_{j=1}^{k} f_{ij}} \tag{2}
\]

The matrix of elements \(p_{ij}\) \((i, j \in I)\) is defined as

\[
P_1 = \{p_{ij}\}_{i,j \in I} \tag{3}
\]

\(P_n\) is defined as the state transfer frequency matrix from state \(S_i\) to state \(S_j\) using transfer times of \(n\). Future state could be predicted by studying \(P_n\).

For homogeneous Markov Chain, the transfer frequency \(p_{ij}\) is related to \(S_j\) and \(t\). The relationship between \(P_n\) and \(P_1\) is represented as:

\[
P_n = P_1^n \tag{4}
\]
Assume the original state frequency distribution is $m_0$, and the transfer frequency distribution of step $n$ is:

$$m_n = m_0 \cdot P_n = m_0 \cdot P_1^n$$  \hspace{1cm} (5)

Comparing the frequency distribution in $m_0$, then state corresponding to the maximum of frequency is the most possible result in the prediction.

All the result is tested using Markov test, since the data should meet Markov property. The test statistic is

$$\chi^2 = 2 \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} \ln \frac{p_{ij}}{p_{i}} \sim \chi^2_{(n-1)^2}$$  \hspace{1cm} (6)

$n$ is defined as number of possible random sequences.

2.2. Quadratic Exponential Smoothing Method

Quadratic Exponential Smoothing Method is one kind of time series analysis method. The method has regularity, stability and systematism, and could be postponed reasonably. The state and tendency in time series will postpone to future to a certain extent [6, 7].

The smoothing exponential should be selected based on actual situation, since different smoothing exponentials lead different results on the same original data. The history data are ranked by time series, if the history data has inconspicuous change, then single exponential smoothing is used. If the history data has linear trend, quadratic exponential smoothing and so on. In the paper, we use Quadratic Exponential Smoothing method.

The definition of Quadratic Exponential Smoothing Method is:

$$\hat{y}_{t+T} = a_t + b_t T$$  \hspace{1cm} (7)

$T$ means advanced periods in prediction; $\hat{y}_{t+T}$ represents predicted result in time section $t + T$; and $a_t, b_t$ are model parameters.

The equations for model parameters $a_t, b_t$ are:

$$a_t = 2S_t^{(1)} - S_t^{(2)}$$  \hspace{1cm} (8)

$$b_t = \frac{a}{1-a} (S_t^{(1)} - S_t^{(2)})$$  \hspace{1cm} (9)

$$S_t^{(1)} = \alpha y_t + (1 - \alpha) S_{t-1}^{(1)}$$  \hspace{1cm} (10)

$$S_t^{(2)} = \alpha S_t^{(1)} + (1 - \alpha) S_{t-1}^{(2)}$$  \hspace{1cm} (11)

$S_t^{(1)}$ and $S_t^{(2)}$ are quadratic exponential smoothing results; $y_t$ is original data; and $\alpha$ is smoothing coefficient. Smoothing initial values are $S_1^{(1)}$ and $S_1^{(2)}$. Then using equation (10) and (11), all the smoothing values are calculated. In general, initial values are determined by the variable $t$.

- If $t \geq 15$, initial value is defined by first observed value:

$$S_1^{(1)} = S_1^{(2)} = y_1$$  \hspace{1cm} (12)

- If $t < 15$, initial value is defined by the average values of the first three observed values.

$$S_1^{(1)} = S_1^{(2)} = \frac{y_1 + y_2 + y_3}{3}$$  \hspace{1cm} (13)
The result of prediction in quadratic exponential smoothing method is mainly affected by the accuracy of smoothing coefficient. In the paper, smoothing coefficient is selected using Levenberg-Marquardt method, which is widely applied in economic projection, analysis of management optimization problems, physical problem analysis, optimization, mechanical and electronic design and so on. The basic thought of the method is minimizing sum of absolute square-error of observed values and corresponding model matching values.

If the observed value, predicted value and smoothing coefficient are $y_t$, $\hat{y}_t$ and $\alpha$. The definition of sum of absolute square-error is:

$$F(\alpha) = \sum_{t=1}^{n} f_t^2(\alpha) = \sum_{t=1}^{n} (y_t - \hat{y}_t)^2$$  \hspace{1cm} (14)

From the equation (14), $F(\alpha)$ is a function of independent variable $\alpha$. To minimize the $F(\alpha)$, equation (14) is translated to optimization problem:

$$\min_{\alpha} F(\alpha) = \min_{\alpha} \sum_{t=1}^{n} f_t^2(\alpha)$$  \hspace{1cm} (15)

3. Overview Of Model and Data

The difference between predicted result and actual value will be significant, if time is not taken into account. Quadratic Exponential Smoothing Method is a weighting algorithm method based on time span. Less time span leads more affection to the prediction. The prediction of Quadratic Exponential Smoothing Method is not well if the data sequence has property of strong randomicity and great fluctuation. However, Markov Chain has property of non-aftereffect, which is suitable for data with property of strong randomicity and great fluctuation. In the paper, based on Quadratic Exponential Smoothing Method, all the prediction is modified using Markov Chain [8]. The flow chart is shown in Fig 1.

Data used in the paper is from balance sheet of Industrial and Commercial Bank of China LTD published by Shanghai Stock Exchange (SSE). Total deposits of ICBC taken from 2010 to 2019 are ranked as time series, which is used as original value of the model. All the data are shown in Table 1.

![Figure 1. Flow Chart](image-url)
Table 1. Deposits from 2010 to 2019

| year | Quarter 1  | Quarter 2  | Quarter 3  | Quarter 4  |
|------|------------|------------|------------|------------|
| 2010 | 103326.16  | 108327.89  | 112825.90  | 111455.57  |
| 2011 | 117644.24  | 120471.38  | 121442.32  | 122612.19  |
| 2012 | 126152.81  | 131805.97  | 136335.02  | 136429.10  |
| 2013 | 143842.73  | 145084.02  | 146927.18  | 146208.25  |
| 2014 | 151609.80  | 157283.32  | 153399.64  | 155566.01  |
| 2015 | 158481.71  | 162877.68  | 165218.28  | 162819.39  |
| 2016 | 170380.81  | 173695.24  | 177544.18  | 178253.02  |
| 2017 | 185650.09  | 190211.71  | 193296.43  | 192263.49  |
| 2018 | 201978.26  | 208180.42  | 213077.07  | 214089.34  |
| 2019 | 225746.51  | 231254.37  | 233681.59  | 229776.55  |

4. Empirical Analysis
First, all the history data are analyzed, and Quadratic Exponential Smoothing Method is chosen. Since the number of samples is 49, the first actual value is taken as initial value, which is $S_1^{(1)} = S_1^{(2)} = 103326.16$. With iterative, the optimal smoothing coefficient $\alpha = 0.3860$, relative error $\Delta_t(\%)$ range is $[-4.3\%, 4.9\%]$. Model coefficients $a_{40} = 233562.97, b_{40} = 3803.99$. Therefore, Quadratic Exponential Smoothing Method is

$$\hat{y}_{40+t} = 233562.97 + 3803.99t$$ (16)

Second, with predicted values and relative error the number of states is 3~5, and every state has principle. 40 samples are divided into 4 states (Table 2). In Table 3, states corresponding to relative errors of total deposits of ICBC is shown.

Third, Markov state transfer counts matrix $(f_{ij})_{i,j\in I}$ is calculated using Table 3, which is

$$(f_{ij})_{i,j\in I} = \begin{pmatrix} 0 & 1 & 2 & 0 \\ 2 & 2 & 4 & 3 \\ 2 & 6 & 9 & 1 \\ 0 & 2 & 3 & 0 \end{pmatrix}$$ (17)

Table 2. State division

| State | Description     | Relative Errors |
|-------|----------------|-----------------|
| $S_1$ | Seriously Overrated | $(-5\%, -2.5\%)$ |
| $S_2$ | Overrated        | $(-2.5\%, 0)$ |
| $S_3$ | Underrated       | $(0, 2.5\%)$ |
| $S_4$ | Seriously Underrated | $(2.5\%, 5\%)$ |
is taken as modified coefficient, and the prediction quarter results of 2020 are 

\[ m_0 = (0.1019, 0.2990, 0.4967, 0.1024), \]

\[ m_4 = (0.1096, 0.2949, 0.4865, 0.1052). \]

From the matrix, state transfer frequency matrix \( P_t \) (equation (18)) is calculated. With iteration, subsequent transfer frequency matrices are calculated.

\begin{table}[h]
\centering
\caption{State and its relative errors in Quadratic Exponential Smoothing Method}
\begin{tabular}{|c|c|c|c|c|c|c|c|c|}
\hline
\textbf{year} & \multicolumn{2}{c|}{\textbf{Quarter 1}} & \multicolumn{2}{c|}{\textbf{Quarter 2}} & \multicolumn{2}{c|}{\textbf{Quarter 3}} & \multicolumn{2}{c|}{\textbf{Quarter 4}} \\
\hline
 & \textbf{Error} & \textbf{\( s_1 \)} & \textbf{Error} & \textbf{\( s_1 \)} & \textbf{Error} & \textbf{\( s_1 \)} & \textbf{Error} & \textbf{\( s_1 \)} \\
\hline
10 & & & 5.00 & \( s_4 \) & -0.74 & \( s_2 \) & & \\
11 & 3.75 & \( s_4 \) & 1.97 & \( s_3 \) & -0.50 & \( s_2 \) & -1.18 & \( s_2 \) \\
12 & 0.66 & \( s_3 \) & 2.79 & \( s_4 \) & 2.26 & \( s_3 \) & -1.50 & \( s_2 \) \\
13 & 2.54 & \( s_4 \) & -0.63 & \( s_2 \) & -1.29 & \( s_2 \) & -3.11 & \( s_1 \) \\
14 & 0.82 & \( s_3 \) & 2.24 & \( s_3 \) & -3.73 & \( s_1 \) & -1.47 & \( s_2 \) \\
15 & 0.06 & \( s_3 \) & 1.51 & \( s_3 \) & 0.56 & \( s_3 \) & -2.78 & \( s_1 \) \\
16 & 2.38 & \( s_3 \) & 1.40 & \( s_3 \) & 1.13 & \( s_3 \) & -0.90 & \( s_2 \) \\
17 & 2.14 & \( s_3 \) & 1.39 & \( s_3 \) & 0.14 & \( s_3 \) & -2.49 & \( s_2 \) \\
18 & 2.36 & \( s_3 \) & 1.99 & \( s_3 \) & 0.94 & \( s_3 \) & -1.40 & \( s_2 \) \\
19 & 2.75 & \( s_4 \) & 1.13 & \( s_3 \) & -0.95 & \( s_2 \) & -4.37 & \( s_1 \) \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\centering
\caption{Results of Quadratic Exponential Smoothing and Markov Chain}
\begin{tabular}{|c|c|c|c|}
\hline
\textbf{Quarter} & \textbf{Predicted result} & \textbf{Predicted range} & \textbf{Modified result} \\
\hline
1 & 237366.96 & \( (237366.96, 243301.14) \) & 240334.05 \\
2 & 241170.95 & \( (241170.95, 247200.23) \) & 244185.59 \\
3 & 244974.94 & \( (244974.94, 251009.32) \) & 248037.13 \\
4 & 248778.93 & \( (248778.93, 254998.40) \) & 251888.67 \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\centering
\caption{Markov modified results and relative errors}
\begin{tabular}{|c|c|c|c|c|c|c|c|c|c|c|c|}
\hline
\textbf{year} & \multicolumn{2}{c|}{\textbf{Quarter 1}} & \multicolumn{2}{c|}{\textbf{Quarter 2}} & \multicolumn{2}{c|}{\textbf{Quarter 3}} & \multicolumn{2}{c|}{\textbf{Quarter 4}} \\
\hline
 & \textbf{Modified Result} & \textbf{Relative Error} & \textbf{Modified Result} & \textbf{Relative Error} & \textbf{Modified Result} & \textbf{Relative Error} & \textbf{Modified Result} & \textbf{Relative Error} \\
\hline
2010 & 11206.72 & 1.44 & 110881.50 & 0.52 \\
2011 & 117476.00 & 0.14 & 119575.31 & 0.74 & 120523.92 & 0.76 & 122502.73 & 0.09 \\
2012 & 126889.55 & -0.58 & 132935.91 & -0.86 & 134924.65 & 1.03 & 136741.11 & -0.23 \\
2013 & 145448.98 & -1.12 & 144177.51 & 0.62 & 146969.46 & -0.03 & 145107.08 & 0.75 \\
2014 & 152242.04 & -0.42 & 155685.26 & 1.02 & 153137.81 & 0.17 & 155875.60 & -0.20 \\
2015 & 160364.67 & -1.19 & 162423.05 & 0.28 & 166342.93 & -0.68 & 161070.05 & 1.07 \\
2016 & 168408.20 & 1.16 & 171259.98 & 1.40 & 177740.99 & -0.11 & 177609.99 & 0.36 \\
2017 & 183957.34 & 0.91 & 189918.82 & 0.35 & 195441.59 & -0.11 & 194583.76 & -1.21 \\
2018 & 199670.80 & 1.14 & 206578.95 & 0.77 & 213721.84 & -0.30 & 214376.98 & -0.13 \\
2019 & 227771.79 & -0.90 & 231507.12 & -0.11 & 232954.30 & 0.31 & 230826.03 & -0.46 \\
\hline
\end{tabular}
\end{table}

Forth, based on significance level \( \alpha = 0.05 \), and 9 degrees of freedom, \( \chi^2 = 19.168 \) using equation (6). The result fits the Markov property.

Last, initial state probability distribution of relative error sequence in quarter 4 of 2019 is shown as \( m_0 \). State probability distribution in quarter 3 and 4 of 2020 is predicted as \( m_3 = (0.1019, 0.2990, 0.4967, 0.1024), m_4 = (0.1096, 0.2949, 0.4865, 0.1052) \). From the result, quarter 3 and 4 mostly in state \( S_3 \). Moreover, the middle of range (0.25\%\) of predicted relative error of state \( S_3 \) is taken as modified coefficient, and the prediction quarter results of 2020 are 240334.05, 244185.59, 248037.13, 251888.67 billion CNY.
The prediction of Quadratic Exponential Smoothing Method is the max relative error is 5.00%, and average relative error is 1.81%. Although the accuracy is relatively high, the results modified using Markov Chain are much more accuracy. The max relative error is 1.44%, and the average relative error is 0.64%. And related results are shown in TABLE IV and TABLE V.

5. Conclusion
In the paper, Markov Chain and Quadratic Exponential Smoothing Method are combined as an aggregative model. First, total deposit is predicted using Quadratic Exponential Smoothing Method, the smoothing coefficient is modified and data are optimized using Levenberg-Marquardt algorithm thought. Moreover, relative errors probability matrix of total deposit is calculated using Markov Chain. And the results predicted in Quadratic Exponential Smoothing Method is modified.

The model takes full advantage of the raw data, especially for the data which has feature of linear upward trend, and random fluctuation. Prediction of the model is much more accurate than Quadratic Exponential Smoothing Method. From the state of relative errors, the state of relative errors of quarter 1 and 2 mainly in $S_3, S_4$. The predicted results are lower than actual total deposits, and state of relative errors of quarter 3 and 4 mainly in $S_1, S_2$, which is overrated. Although the prediction is still overrated after modification, the predication is much more accuracy than Quadratic Exponential Smoothing Method.

Since the epidemic situation of coronavirus in 2020, economic development is influenced. The actual total deposits are lower than prediction. However, whether the prediction in the latter half of the year is overrated or underrated is still unknown, since the epidemic situation in global and China is uncertain. In the following, multiple methods combined with Smoothing Method should be studied.

Acknowledgments
The paper is sponsored by NSFC No. 11501074 and 11701061.

References
[1] Mirea M, Florea I M, Aivaz K A. The Impact of the Interest Rate and the Income of the Households on the Dynamics of Bank Deposits [J]. Ovidius University Annals, Economic Ences Series, 2019, xix.
[2] Mi J, Fan L, Duan X, et al. Short-Term Power Load Forecasting Method Based on Improved Exponential Smoothing Grey Model[J]. Mathematical Problems in Engineering, 2018, (2018-3-25), 2018, 2018(PT.3): 1 - 11.
[3] Li J, Shang P. Financial time series analysis using Total-CApEn and Avg-CApEn with cumulative histogram matrix [J]. Communications in Nonlinear ence & Numerical Simulation, 2018, 63.
[4] Olmedo M T C, Mas J F. Markov Chain [M]// Geomatic Approaches for Modeling Land Change Scenarios. 2018.
[5] Kamatani K. Efficient strategy for the Markov chain Monte Carlo in high-dimension with heavy-tailed target probability distribution [J]. Eprint Arxiv, 2018, 24 (4).
[6] Tratar L F, Mojskerc B, Toman A. Demand forecasting with four-parameter exponential smoothing [J]. International Journal of Production Economics, 2016, 181 (pt.A): 162 - 173.
[7] Lin H, Mu B L. An Empirical Study on Exponential Smoothing and Seasonal Model [J]. Applied Mechanics and Materials, 2014, 599 - 601: 2088 - 2091.
[8] Li X, Cao J, Du D. Comparison of Levenberg-Marquardt Method and Path Following Interior Point Method for the Solution of Optimal Power Flow Problem [J]. International Journal of Emerging Electric Power Systems, 2012, 13 (3): 23 pp.--23 pp.