Modeling the Trend of Credit Card Usage Behavior for Different Age Groups Based on Singular Spectrum Analysis

Wei Nai 1, Lu Liu 2,*, Shaoyin Wang 1 and Decun Dong 3

1 Department of Electronic and Information Engineering, Tongji Zhejiang College, Jiaxing 314051, China; alexni@tongji.edu.cn (W.N.); 82730@tongji.edu.cn (S.W.)
2 Whitman School of Management, Syracuse University, Syracuse, NY 13244, USA
3 Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China; ddc@tongji.edu.cn

* Correspondence: lliu123@syr.edu; Tel.: +1-(315)-420-0628

Received: 12 December 2017; Accepted: 26 January 2018; Published: 29 January 2018

Abstract: Credit card holders from different age groups have different usage behaviors, so deeply investigating the credit card usage condition and properly modeling the usage trend of all customers in different age groups from time series data is meaningful for financial institutions as well as banks. Until now, related research in trend analysis of credit card usage has mostly been focused on specific group of people, such as the behavioral tendencies of the elderly or college students, or certain behaviors, such as the increasing number of cards owned and the rise in personal card debt or bankruptcy, in which the only analysis methods employed are simply enumerating or classifying raw data; thus, there is a lack of support in specific mathematical models based on usage behavioral time series data. Considering that few systematic modeling methods have been introduced, in this paper, a novel usage trend analysis method for credit card holders in different age groups based on singular spectrum analysis (SSA) has been proposed, using the time series data from the Survey of Consumer Payment Choice (SCPC). The decomposition and reconstruction process in the method is proposed. The results show that the credit card usage frequency falls down from the age of 26 to the lowest point at around the age of 58 and then begins to increase again. At last, future work is discussed.

Keywords: statistics of credit card usage; singular spectrum analysis (SSA); time series; behavior analysis; trend modeling

1. Introduction

In modern society, credit cards have become a fact of everyday life for most consumers. A survey about credit card usage has shown evidence of their pervasiveness [1]. As of 2011, seventy-seven percent of U.S. adults owned at least one credit card, with a total of 1.4 billion cards in circulation. The average cardholder owns 7.7 cards and uses a credit card 119 times a year, charging an average of 88 U.S. dollars per transaction or 10,500 U.S. dollars annually [2]. The ease of access of credit cards has given consumers increased opportunities for making purchases. However, while many consumers are able to use credit cards wisely, others seem to be unable to control their consuming behavior [3]. Over the past two decades, the use of credit cards has become an area of economic and social concern [4].

In order to master the credit card usage situation and all of the other economic behaviors that happened around the country of the U.S., and to provide a publicly available time series dataset to support research on consumer payments, the Survey of Consumer Payment Choice (SCPC) was carried out by Federal Reserve Bank of Boston [5]. There are three broad categories of SCPC variables. Of them, the first set of variables comprises My Household Questionnaire (MHQ) variables. MHQ...
is used to gather demographic data about each respondent, including age, gender, and household income. The second set of variables comprises survey variables, which are the actual results from the SCPC survey questions. The third set of variables comprises created variables, which are created to help people understand this survey better.

In this paper, we make the first effort to analyze the trend of credit card usage as age grows with the aid of Singular Spectrum Analysis (SSA) based on SCPC time series data. This paper is organized as follows: in Section 2, we make a comprehensive literature review from three aspects: analysis based on SCPC data, developments and applications of SSA, and the trend studies of credit card usage. In Section 3, we present the SSA methodology we use and how we model the credit card usage with SSA. In Section 4, we describe the SCPC survey and the credit card usage data we obtained from the survey. Section 5 is our case study, involving how we applied SSA to our credit card usage data. Last but not least, in Section 6, we make our summary, conclusions, and comments upon future work.

2. Literature Review

2.1. On Analysis Based on SCPC Data

As has been discussed above, the SCPC has been carried out so as to let government officials, researchers, and the public to obtain an easier approach to consumer payments behavior, and to obtain a better understanding of the usage situation of credit cards. Until now, various research efforts has been made by employing the data in this survey. By statistical analysis, Foster et al. [5] have pointed out that the average consumer held 5.1 of the nine common payment instruments and used 4.2 of them during a typical month in the U.S. of the year 2008. Stavins et al. [6] have tried to discover whether U.S. merchants are using their recently granted freedom to offer price discounts and other incentives to steer customers to pay with methods that are less costly to merchants, and have found that only a very small fraction of transactions received a cash or debit card discount, and even fewer were subjected to a credit card surcharge. Koulayev et al. [7] developed and estimated a structural model of adoption and use of payment instruments by U.S. consumers. After that, they utilized a cross-section from the Survey of Consumer Payment Choice. Finally, they evaluated substitution and income effects.

2.2. On Developments and Applications of Singular Spectrum Analysis

Singular Spectrum Analysis was put forward by Broomhead and King [8,9]. At first, it was widely used because of its effectiveness in reducing residuals. Later, Vautard et al. [10,11] applied this new SSA method in dealing with time series data and showed good results, especially for short, noisy, chaotic signals. At the same time, Ghil and Vautard [12] used SSA to analyze the time series of global surface air temperatures for the past 135 years. They found out that SSA can generate a secular warming trend and a small number of oscillatory modes; after the residual is separated, the original pattern of the series can be better explained. In 2002, Ghil et al. [13] reviewed different SSA methods used in the area of climate variables prediction. The various steps, advantages, and disadvantages were all discussed. In 2007, Golyandina and Osipov [14] analyzed the problem of applying SSA to time series with missing data. Their proposed algorithms result in the extraction of additive components of time series with simultaneous filling in of the missing data. SSA has also been applied to the area of finance and economics. Hassani and Thomakos [15] reviewed recent developments in the theoretical and methodological aspects of SSA in the area of economic and financial time series, and also present some new results. Again, Hassani et al. [16] applied univariate and multivariate singular spectrum analysis for predicting the value of changes in the daily pound/dollar exchange rate. A comparison to other models was presented, and the results showed that SSA is superior to other benchmarking models. Also, Hassani et al. [17] developed the multivariate SSA (MSSA) technique and demonstrated that MSSA can be a powerful method for time series analysis and forecasting. UK Industrial Production series were used to illustrate the main findings, and the result showed better accuracy compared with the autoregressive integrated moving average (ARIMA) and vector autoregressive (VAR) models.
During recent years, the SSA algorithm has been widely used in various industries, and Safari et al. [18] even upgraded it as multi-scaled SSA (MSSSA) in making short-term forecasts for certain objects, such as wind power, which may have many chaotic components.

2.3. On Trend Studies of Credit Card Usage

Until now, a bunch of efforts has been made in the usage trend analysis of credit cards. Some of the related research has focused on the usage trend of a specific group of people. Dellutri et al. [19] studied the rising trend of senior citizen credit card debt, and pointed out that increased health care costs, gambling, lower interest rates on investment, the loss of jobs before planned retirement, and low retirement income are the main causes by a thorough statistics. Adams and Moore [20] analyzed the risky behavior of college students, for they are of more convenience for researchers to acquire card usage sample data on; while some of the related research has focused on the usage trend of specific behavior, such as for the misuse of a credit card, by simply investigating statistics data. Manning [21] and Ladka [22] pointed out that the popularization of credit cards had increased the trend of personal bankruptcy, and by a statistical analysis, the American Bankruptcy Institute [23] showed the result that about 1 personal bankruptcy happens in every 175 adults, and by frequent itemset mining, Seeja and Zareapoor [24] developed a credit card fraud detection model to identify misbehaving tendencies. As for the number of cards owned and the total amount of items to be repaid, Wang and Xiao [25] showed the result of a mean of 2 cards with a range of 1 to 18 cards owned by each holder, and that the total amount of items to be repaid has risen from around $250 to almost $1500 in the past 30 years. When it comes to the comprehensive analysis of credit card usage trends, Yang et al. [26] pointed out that trend analysis is important for credit card issuers as emerging consumer psychology and consumer-behavior-related subjects; however, the only work that comprehensively analyzed the trend of credit card usage, done by Mansfield et al. [27], focused more on the research of a usage trend by reviewing 537 research reports in the past 40 years than on the trend of the credit card usage itself. It can be seen that while analyzing credit card usage by simply enumerating or classifying raw data is a common practice in related research, there is an obvious lack of support in specific mathematical models on this issue.

3. Singular Spectrum Analysis

The Singular Spectrum Analysis (SSA) method was developed since 1970s. SSA is a model-free approach because SSA decomposes an original time series into a trend, seasonal and semi-seasonal components, and residuals based on singular value decomposition (SVD) [28]. The new decomposed series can help us to understand the trend of the original time series and to extract seasonal or monthly components and residuals. The basic concept of SSA consists of four steps: embedding, singular value decomposition (SVD), grouping, and diagonal averaging. The basic SSA algorithm decomposes an initial time series into a new time series which consists of the trend, the seasonal and semi-seasonal components, and residuals. The basic SSA algorithm has two stages: decomposition and reconstruction.

3.1. Decomposition

The first step in the basic SSA algorithm is embedding. In this step, the original time series is converted into a trajectory matrix. For an original time series \( X = x_1, \ldots, x_N \) of length \( N \) with no missing value, a window of length \( L \) is chosen \((2 < L < N/2)\) to embed the original time series.
Then, the original time series $X$ is mapped into $L$ lagged vectors, $X_i = x_i, \ldots, x_{i+L-1}$ for $i = 1, \ldots, K$, where $K = N - L + 1$. Thus, $T_X$ is written as:

$$T_{i,j} = \begin{pmatrix} x_1 & x_2 & \cdots & x_L \\ x_2 & x_3 & \cdots & x_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_K & x_{K+1} & \cdots & x_N \end{pmatrix}.$$  

After the embedding step, SVD is applied to the trajectory matrix $T_X$ and the decomposed trajectory matrices $T_i$ are obtained. $U_i$ for $1 < i < L$ is a $K_i \times L$ orthonormal matrix, $D_i$ for $1 < i < L$ is a diagonal matrix order of $L$, and $V_i$ for $1 < i < L$ is an $L \times L$ square orthonormal matrix. In this step, $T_X$ has $L$ many singular values, which are:

$$T_X = \sum_{i=1}^{d} \lambda_i U_i V_i^T. \quad (1)$$

Then, we can calculate the ratio of each eigenvalue. The ratio of each eigenvalue is the contribution of the matrix $T_i$ to $T_X$.

3.2. Reconstruction

The grouping step of the reconstruction stage is to decompose the $L \times K$ matrix $T_i$ into subgroups according to the trend, the seasonal and semi-seasonal components, and residuals. The grouping step of the reconstruction stage is a partition of the set of indices $1, \ldots, d$ into the collection of $m$ disjoined subsets of $I = I_1, \ldots, I_m$. Thus, $T_i$ corresponds to the group $I = [I_1, \ldots, I_m]$. $T_h$ is a sum of $T_j$, where $j \in I$. So, $T_X$ can be expanded as

$$T_X = T_1 + \ldots + T_L = T_{h_1} + \ldots + T_{h_m}. \quad (2)$$

Assume that there are two groups of eigentriples of the trajectory matrix $T_X$: $T_L$ and $T_R$. The whole set will be $I = [1, \ldots, d], R \cup L = I$. However, $R$ is not a subset of $L$. $T_I$ is

$$T_I = \sum_{i \in I} \sqrt{\lambda_i} U_i V_i^T \quad (3)$$

and we can calculate $T_L = T_I - T_R$ under the assumption of weak separability. Thus, $T_L$ can be written as

$$T_L = \sum_{i \in L} \sqrt{\lambda_i} U_i V_i^T. \quad (4)$$

The final step is diagonal averaging. In the basic SSA algorithm, the diagonal averaging step is to transform the grouped matrices $T_i$ into a new time series of length $N$. We obtain the time series $\tilde{T}$ from an averaging of the corresponding diagonals of the matrix $T_i$.

$$\tilde{T}_{i,j} = \begin{cases} \frac{1}{s-1} \sum_{j=1}^{s-1} x_{j,s-j} & 2 \leq s \leq L \\ \frac{1}{L} \sum_{j=1}^{L} x_{j,s-j} & L+1 \leq s \leq K+1 \\ \frac{1}{N-s+2} \sum_{j=s-K}^{N-s+2} x_{j,s-j} & K+2 \leq s \leq N+1 \end{cases}. \quad (5)$$
4. Data Description

4.1. The Survey of Consumer Payment Choice

The Survey of Consumer Payment Choice (SCPC) was produced by the Consumer Payments Research Center (CPRC) in the research department at the Federal Reserve Bank of Boston. One of the major goals of the Survey of Consumer Payment Choice (SCPC) was to provide a publicly available, consumer-level longitudinal dataset to support research on consumer payments and to provide aggregate data on trends in U.S. consumer payments.

There are 2065 respondents in the 2012 SCPC, and it contains 49 tables with detailed estimates of the number of consumer payments, rate of adoption, and share of consumers using nine common payment instruments, which are cash, checks, money orders, travelers checks, debit cards, credit cards, prepaid cards, online banking bill payments (OBBP), and bank account number payments (BANP) plus payments made directly from consumers’ income source. The report also contains estimates of consumer activity related to banking, cash management, and other payment practices; consumer assessments of payment characteristics; and a rich set of consumer and household demographic characteristics.

The analysis focuses on the generations aged 18 through 94 and how they made payments, including the median dollar value of payments, the frequency of the payment instrument used, the device used to pay (including mobile, computer, or mail order) and the top five merchant categories where the majority of transactions were made.

4.2. Credit Card Usage

Of the many tables in SCPC 2012, we picked up the table containing information about credit card usage. The survey asked questions about the time of usage of a credit card in the last month. The data was first filtered, excluding those that were invalid data. After that, we had 1248 data points in total. The scatter plot is provided below, as is shown in Figure 1. As we can see from the figure, it is pretty messy and it seems that there is no particular pattern in the series of data. The plot of the original credit card usage series is shown in Figure 2.

![Figure 1. Scatter plot of credit card usage per month with age.](image-url)
5. Model Results

SSA is a subspace-based method which works in four steps. First, we select a maximum lag $L$ ($1 < L < N$, where $N$ is the number of data points), and a trajectory matrix is created with $L$ columns (lags 0 to $L - 1$) and $N - L + 1$ rows. Second, the SVD is calculated of the trajectory matrix. Third, we use diagnostics to determine what eigenvectors are grouped to form bases for projection. Fourth, a reconstructed series is formed for each group of eigenvectors.

5.1. Decomposition

The first step of decomposition is embedding. To perform embedding, the original time series is mapped into a sequence of lagged vectors of size $L$ by forming $K = N - L + 1$ lagged vectors. Let $X$ be the initial time series $\{x_1, x_2, \ldots, x_n\}$ with no missing values of length $N$.

In our case of a credit card usage series, we have tried multiple values of $L$, and finally $L = N/4$ was determined to be the optimum value.

Next, we decomposed the trajectory matrix into eigenvectors. After we decomposed our series of credit card usage, what we got first is a series of eigenvectors as is shown in Figure 3. The graph of eigenvectors reflects the contribution of the leading eigentriple. Figure 3 shows the form of the ten eigenvectors. The eigenvectors have almost constant coordinates, and therefore they correspond to a pure smoothing process. As we can see from Figure 3, the first eigentriple contributes the most (48.69%) of all the eigentriples, while the other ones contain several high-frequency components and therefore are not related to the trend.
5.2. Reconstruction

The stage of reconstruction can be viewed as the formation of an elementary series and then taking a sum of some of them depending on the grouping chosen. What we do first is eigentriples grouping. This process of grouping corresponds to splitting the original matrices $X_i$ into several groups and then summing the matrices within each group. If $I = i_1, \ldots, i_p$ is one such group, then the matrix $X_I$ corresponding to the group $I$ is defined as: $X_I = X_{i_1} + \ldots + X_{i_p}$. For $m$ such groups, $X$ will be given as: $X = X_{i_1} + \ldots + X_{i_m}$. The contribution of component $X_I$ is measured by the share of the corresponding eigenvalues.

The choice of leading eigentriples corresponds to the approximation of the time series in view of the optimality property of the SVD. Then, we perform the step of diagonal averaging. At this step, we transform each matrix of the grouped decomposition into a new series. In the stage of reconstruction, we can see the components of the original series as is shown by the following figures from Figures 4–6.

In our case, the problem of finding a reconstructed structure of the original series by SSA is the same as the identification of the eigentriples of the SVD from the trajectory matrix of this series, which correspond to the trend, the seasonal component, the semi-seasonal component, and noise. In practice, the means we used for noise extraction was the grouping of the eigentriples, which excludes the elements of trend, seasonal component, and semi-seasonal component.

After we eliminate the seasonal, semi-seasonal, and residual components of the original credit card usage series, we can finally get the trend of the series, as is shown by Figure 7.

From the figure, we can see that following the steps of SSA, the originally messy credit card usage series showed an obvious trend. First, it increases as age grows and reaches its first peak at around the age 42, then for some unknown reasons it falls down to its lowest point at around the age 58, and then it begins to increase for the second time monotonely to the end. Finally, there is another growing trend, the seasonal component, the semi-seasonal component, and noise.
trend, but we believe this irregular pattern is due to the lack of data points when the age reaches 80. The trend of the credit card usage series is shown by Figure 7.

Figure 4. Seasonal component of the original credit card usage series.

Figure 5. Semi-seasonal component of the original credit card usage series.
Figure 6. Residue component of the original credit card usage series.

Figure 7. Trend of the original credit card usage series.
6. Summary, Conclusions, and Future Work

In this paper, for the first time we applied Singular Spectrum Analysis (SSA) to credit card usage data to analyze the trend of credit card usage as age grows. Following the four steps of SSA, after eliminating the seasonal, semi-seasonal, and residual components of the original time series, we obtain our trend of credit card usage. The result showed that, at first, the credit card usage increases as age grows, and reaches its first peak at around the age 26, then for some unknown reasons it falls down to its lowest point at around the age 58, and then it begins to increase for the second time monotonely to the end. At last, there is another growing trend, but we believe this irregular pattern is due to the lack of data points when the age reaches 80.

To the best of our knowledge, this is the first time that SSA has been applied to analyze the trend of credit card usage over age. Previous studies concerning credit card usage have mainly focused on credit card debt trends, the risky credit card usage behavior of college students, or the relation between credit card usage and personal bankruptcy. The result we generated is novel and important considering the field of credit card usage.

In the future, we will do research about how to explain this trend: why does the credit card usage reach its peak at around the age of 26? Will the position of this peak be changed? In addition, why does it fall down afterward and why does it grow up again? We will try to analyze this interesting and counter-intuitive phenomenon by the demographic data collected by the American Life Panel (ALP) survey.

Acknowledgments: This work is supported by National Key Research and Development Program of the 13th Five-year Plan in China under Grant No. 2016YFB1200401, the Fund of Shanghai Cooperative Center for Maglev and Rail Transit, and the Fund of Shanghai Fei Tu You Science and Technology Company Limited under Grant No. 0217013.

Author Contributions: Wei Nai and Lu Liu proposed the idea of employing the SSA algorithm into the credit card usage behavioral analysis; Lu Liu conceived and designed the experiments; Wei Nai set up and performed the experiments; Lu Liu collected the data; Shaoyin Wang and Decun Dong analyzed the data; Wei Nai and Lu Liu wrote the paper; and Xinxian Lei (not in the author list) and Decun Dong gave many helpful insights into the paper’s organization.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).