*Theoretical and Applied Economics*
*Volume XXI (2014), No. 3(592), pp. 57-66*

# Time series analysis in loan management information systems

**Julian VASILEV**
Varna University of Economics, Bulgaria
vasilev@ue-varna.bg

**Abstract.** *A loan management information system (for transaction processing in credit institutions) records data for given loans, returned sums for principal, interest and taxes. The purpose of this article is to accept or reject our assumption that the more loans are given, the more sums are returned by customers. The main methodology used is time series analysis. Data are analyzed in SPSS. The main proved conclusion is that incoming money flows in credit institutions do not depend on the amount of given loans. This article is published for the first time. This article gives notes for extending existing loan management information systems for loan management in credit institutions and banks in the direction of business analysis.*

**Keywords:** loan management information systems, time series analysis, SPSS, credit institutions, incoming money flows.

**JEL Classification:** E51.
**REL Classification:** 11B, 11C.

## 1. Introduction

Credit institutions use software systems or simply spreadsheets to record data about money flows between them and their customers. Outgoing money flows concern sums for given loans. Incoming money flow concerns principals, taxes and rates. Transaction data consists of the following data: number of contract, date, sum, customer, type of payment. Usual reports in loan management information systems are intended to give summarized information about: (1) money given for certain period of time, (2) money received for certain period of time, (3) transactional data, which may be filtered or sorted. Some loan management information systems have report generators which help credit inspectors to make Pivot tables on the basis of transactional data. Some reports are intended to get a list of customers: (1) who have to pay today, (2) who delayed their payment more than N days.

All these reports focus on transactional data. They are built by selecting data, filtering data with simple or complex criteria and grouping of data. These types of reports are simple but very useful. Extending the functionality of loan management information systems means that transactional data may be used for time series analysis. Since transactional data contain column date, we have data at uniform time intervals. That is why we have a time series. Time series analysis consists of methods for analyzing data. To get some knowledge from data it is recommended to use statistical methods.

Time series analysis is used mainly for forecasting. Trends in economics certainly exist. But by applying statistical methods for time series analysis we me check whether an economic phenomena is dependent by time. We can check whether time is an independent variable that mostly influences a certain economic phenomena.

Analyzing time series differs from other types of data analysis. Decomposition of time series, trend estimation, autocorrelation analysis and spectral analysis has to be made in order to analyze a time series. All these methods have a detailed description in books in Statistics. Before applying regression analysis separation of data is done. Trend, seasonality, slow and fast variation components are calculated. The trend is a long term direction. The seasonal component shows a systematic and calendar effect. The irregular component shows short-term fluctuations.

## 2. Literature review

Loan management systems are mainly used to store transactional data for given loans. They record data for given sums, returned sums for principal, taxes and rates. Some authors (Niţescu, 2012, pp. 53-62) focus on customer factors influencing a bank customer behavior and their impact on early repayment of loans. Niţescu also analyzes the period of loan payment. Others (Kohler et al., 2012, pp. 8-20) connect the term "loan management systems" with libraries. Usually loan management systems operate with confidential databases. Thus they are usually not used in scientific research. Researchers prefer to use other forms of collecting data – for instance questionnaires.

Time series analysis is used mainly for forecasting (Box et al., 2013, Mastrangelo et al. 2013). Comparing modelled and monitored data and calculating errors in time series analysis is a usual technique (Butland et al., 2013, p. 136). Recent research articles focus on the implementation of time series analysis in medicine (Du et al., 2013, p. 296, Doshi-Velez et al., 2014, pp. 54-63). Times series analysis is a common technique used in banks (Faure, 2013, Ghassan et al., 2013).

A lot of articles give good examples on using time series analysis in SPSS (Ma et al., 2014, pp. 35-42, Damiani, 2014, p. 6). But a great deal of them is in the sphere of medicine and pharmaceutical business.

## 3. Collecting data

The use of a loan management information system records a lot of data. By exporting data from reports and sending them to an electrons spreadsheet (MS Excel) we have the following dataset.

**Table 1.** *A part of the Dataset for time series analysis of loans*

| Year | Month | Given_sum | Returned_principal | Received_taxes |
|---|---|---|---|---|
| 2010 | 1 | 920 | 31 | 1 |
| 2010 | 2 | 2256 | 513 | 51 |
| 2010 | 3 | 2078 | 1990 | 277 |

The dataset consists of 48 rows (records) starting from 2010, January, ending with December 2013. We have monthly data. So we may try to predict a trend and try to mark seasonal fluctuations.

## 4. Basic hypothesis

We assume that the more amount of money is given to customers, the more money is returned as principals, interests and taxes. Interests and taxes are combined in one column in our dataset. The less money is given as credits, the less money is returned. In the sphere of loans there is a lag effect. The loan is given during the current month but payments of taxes and the principal come later – in forthcoming months. We may aggregate the dataset for these three years and make the analysis on a quarterly basis. First, we start the analysis on a monthly basis.

## 5. Analyzing the dataset

SPSS may be used to solve different problems in economics. It is a powerful tool for wide variety of analysis including time series analysis. All the data in our dataset has numeric values. SPSS can process only numeric values. That is why our dataset is ready to be transferred in SPSS. In SPSS the following variables are defined: year, month, given_sum, returned_principal and received_taxes. All of them have a scale measure. It is very important to mark the type of measure for each variable. After defining the variables we transfer the dataset from Excel to SPSS by copy-pasting the dataset. We have to check if there is a trend. Since we have a time series dataset, we have to define dates (Data/Define dates).

We continue with making charts for each of the three columns with sums to check graphically if there is a trend.

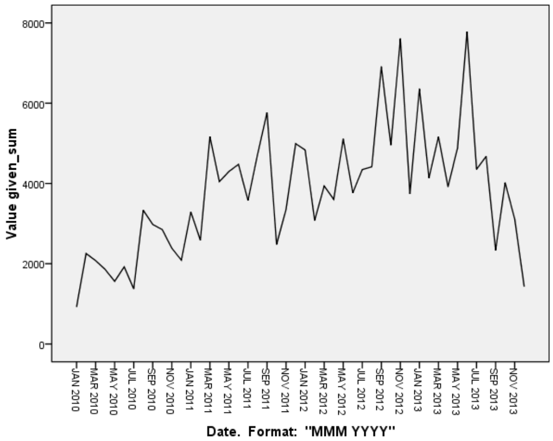

**Figure 1.** *Line chart of "given_sum"*

It is obvious that the value of given_sum increases during the time, but at the end of the dynamic row (the last four months) it decreases. So let's check what happens with the returned principal.

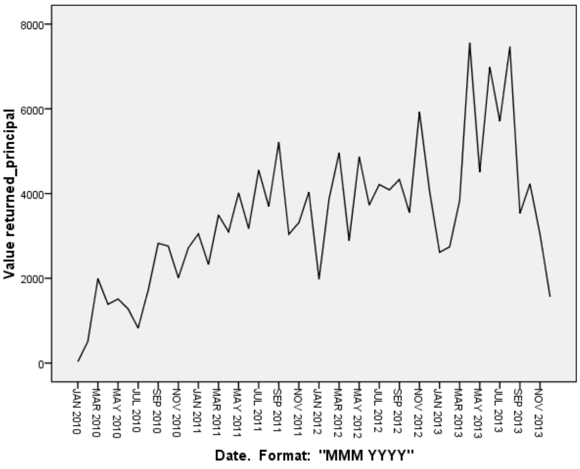

**Figure 2.** *Line chart of "returned_principal"*

The values of "returned_sum" also increase during the time, but during the last three months they decrease. What happens with received sums for interest and taxes?

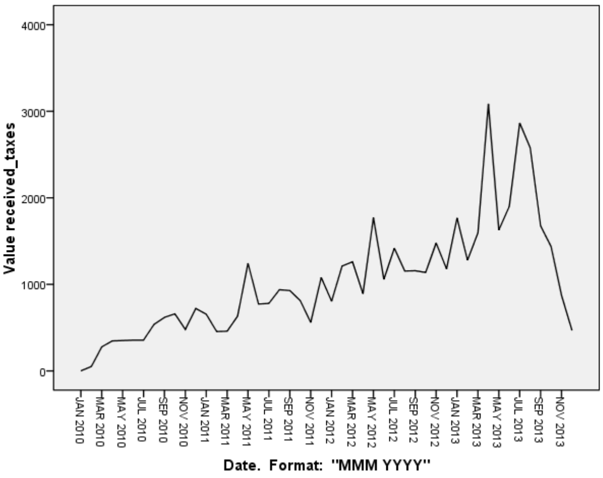

**Figure 3.** *Line chart for "received_taxes"*

The line chart of "received_taxes" shows that they increased 45 months, but the last 3 months they decreased. Surely there is an event in the summer of 2013 which influences the decrease in taking loans and their further payment.

If we omit the four months at the end of year 2013 the trend will show an increase in the three variables. We do not have values for 2014, so we make assumptions or forecasting. Later on the predictions may be checked whether they are correct or not.

It should be checked for correlations between the three time series (Analyze/Correlate/Bivariate). Since the values in the three time series are in cardinal scale, the Pearson correlation coefficient may be used. Two-tailed test of significance is made. The correlation between the "given_sum" and "returned_principal" is 0.692. The correlation between the "given_sum" and "received_taxes" is 0.555. The correlation between the "recived_taxes" and "returned_principal" is 0.845. The three correlation coefficients are significant at 0.01 level. They show that if a sum is given, there is a high probability for the loan to be returned and to get taxes and interest from the customer.

We start with curve estimation for the variable "given_sum" (Analyze/Regression/Curve estimation). We choose all possible models, we include constant in equation and we display the ANOVA table. The highest value of R-square (0.475) has the quadratic function. The unstandardized coefficient B (261.085) is statistically significant, but the constant (703.595) is not statistically significant. The very low value of R-square shows that other factors different from time affect the given sum of a loan.

Calculations may continue again with curve estimation but not including the constant in equation. The highest value of R-square (0.936) is for the power function. The ANOVA table shows that the power model is adequate. For the independent variable SPSS suggests ln(Case Sequence) and the dependent variable is ln(given_sum).

$\mathrm{Ln}(Y) = \mathrm{Ln}(X)^{2.582}$

Figures 1, 2 and 3 graphically show that the more sum is given, the more returned sums are received and the more taxes are received. During the first 44 periods (months) all values increase, but at the end – the last four periods (months) – the end of year 2013 all sums decrease. What will happen in 2014 we do not know but we may predict.

To make regression analysis we have to remove the trend and to make seasonal decomposition. We start with calculating a natural logarithm of the three series of data (Transform/Compute variables). We continue with seasonal decomposition (Analyze/Forecasting/Seasonal Decomposition) of the three data series. We use the additive model type.

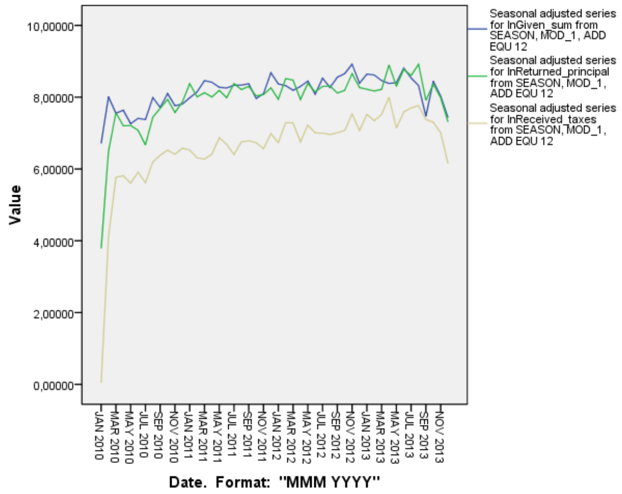

**Figure 4.** *Seasonal adjusted series*

We have cut off all peaks. Stochastic variations are removed. As is it obvious from Figure 4 some months the given sum is more than the received principal, other months – the opposite. The two data series behave like two vines.

By calculating natural logarithms of each value the dispersion is stabilized. A check for autocorrelations for the three seasonal adjusted series has to be made (Analyze/Forecasting/Autocorrelations). The partial autocorrelation coefficients (ACF) are within the upper and the lower confidence limit for the seasonal adjusted series “lnReturned_principal” and “lnReceived_taxes”. For the seasonal adjusted series “lnGiven_sum” there are partial ACF which are outside the upper and the lower confidence limit. We have to create a time series by the using of the first order of the difference function for SAS_1 (Transform/Create time series). A new time series is created SAS_1_1. We check the partial ACF for SAS_1_1. The autocorrelation exists just for the first period (the first month of the time series).

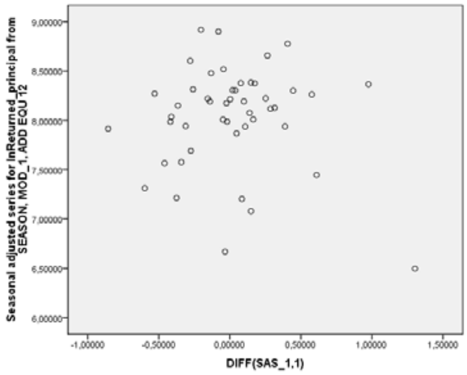

**Figure 5.** *A scatter graph between DIFF(SAS_1) and SAS_2*

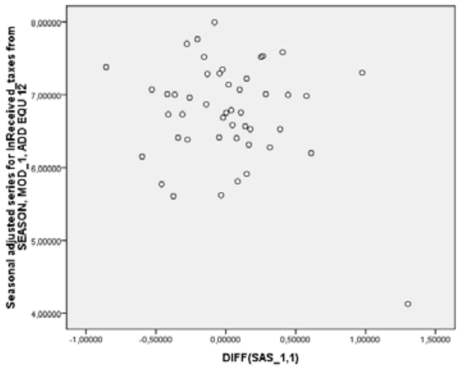

**Figure 6**. *A scatter graph between DIFF(SAS_1) and SAS_3*

The scattered graph 5 shows there is not a correlation between "given_sum" and "returned_principal". The scattered graph 6 shows there is not a correlation between "given_sum" and "received_taxes". Let's check the correlation between SAS_2 and SAS_3.

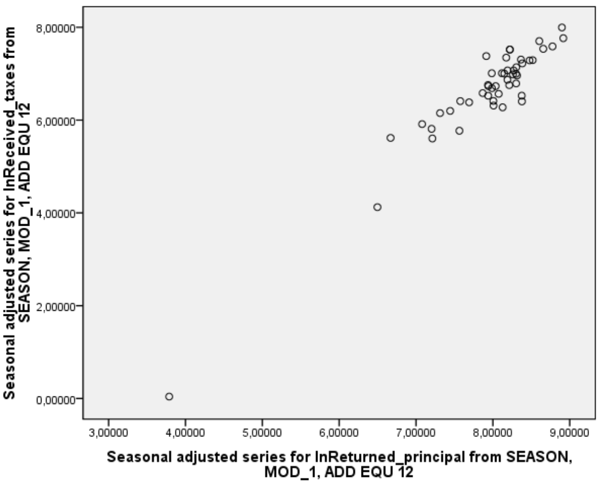

**Figure 7.** *A scatter graph between SAS_2 and SAS_3*

Figure 7 shows a great dependency between the seasonal adjusted values of "lnReturned_principal" and "lnReceived_taxes". The coefficient of Person is 0.955. It is statistically significant at the 0.01 level.

Now a regression model may be made. The dependent variable may be the SAS_2 or SAS_3. We choose SAS_2. The independent variable is DIFF(SAS_1 1). We check the correlation between the two variables. The Pearson coefficient is -0.075. Its sig. 2-tailed is 0.619. So the Pearson correlation coefficient is not significant.

There is not a correlation between DIFF(SAS_1 1) and SAS_2. Since we used logarithms of real values the model may be linear.

Y = B0*X + B1.

To analyze we choose linear regression (Analyze/Regression/Linear). To test residuals the Durbin-Watson coefficient is chosen. The R-square value of the liner model is 0.006. It is very low. It means that the returned sum and paid taxes do not depend on time. The ANOVA test has a significance of the model 0.619. It means that the linear model is not adequate. The coefficient B0 is -0.098 with significance 0.619. Thus B0 is not valid. The constant B1 is 8.036 with significance 0.000. The constant is significant. The calculations made show that the returned sums do not depend on the given sums. We reject the initially formulated hypothesis. The Durbin-Watson value is 0.672. It is outside the range 1.5-2.5. It means that there is autocorrelation in the dependent variable.

The calculations show that the returned sum does not depend on the given sum and time. We may check whether the returned sum depends on time. So the independent variable is the case number. A new variable in SPSS is added. The dependent variable is the SAS_2. The model type is linear. The R-square value is 0.325. The Durbin-Watson value is 0.745. It is outside the range 1.5-2.5. It means that there is autocorrelation in the dependent variable. The ANOVA test shows that the model is adequate. Unstandardized coefficients are statistically significant. The R-square value is 0.339. The model is the following

SAS_2 (predicted) = 0.033 * Case_number + 7.133

Our dataset consists of 48 cases. We may predict the returned sum for January, February and March of year 2014.
SAS_2 predicted for January 2014 = 0.033*49 + 7.133 = 8.75
SAS_2 predicted for February 2014 = 0.033*50 + 7.133 = 8.783
SAS_2 predicted for March 2014 = 0.033*51 + 7.133 = 8.816

Because we used the logarithm function, to predict real sums we have to do the following calculation in Excel:
Predicted returned principal for January 2014 = exp(1)^8.75 = 6310.69
Predicted returned principal for February 2014 = exp(1)^8.783 = 6522.42
Predicted returned principal for March 2014 = exp(1)^8.816 = 6741.25

The average returned sum for the last three months is 2941.67. The calculations show that there is a probability of 34% the returned monthly sum to be twice higher than the average returned sum during the last three months of the time series (Oct-Dec 2013).

## 6. Conclusions

The calculations and the application of statistical methods show that the returned principals do depend neither on time nor on the given sums. The received taxes and interest highly depend on the returned principal. The main hypothesis of the article (that the more sums are given the more sums are received as principals and taxes) is scientifically rejected. The time series analysis shows that incomes in a credit institution do not depend only on time. There are other factors that affect incomes in a credit institution. Future research may focus on other models which include other more appropriate independent variables.

## References

Box, G.E., Jenkins, G.M., Reinsel, G.C. (2013). "Time series analysis: forecasting and control". Wiley. com.

Butland, B.K., Armstrong, B., Atkinson, R.W., Wilkinson, P., Heal, M.R., Doherty, R.M., Vieno, M. (2013). "Measurement error in time-series analysis: a simulation study comparing modelled and monitored data". *BMC medical research methodology*, *13*(1), p. 136

Damiani, G., Federico, B., Anselmi, A., Bianchi, C.B., Silvestrini, G., Iodice, L., Ricciardi, W. (2014). "The impact of Regional co-payment and National reimbursement criteria on statins use in Italy: an interrupted time-series analysis", *BMC health services research*, 14(1), p. 6

Doshi-Velez, F., Ge, Y., Kohane, I. (2014). "Comorbidity Clusters in Autism Spectrum Disorders: An Electronic Health Record Time-Series Analysis", *Pediatrics*, 133(1), pp. e54-e63

Du, C.J., Hawkins, P.T., Stephens, L.R., Bretschneider, T. (2013). "3D time series analysis of cell shape using Laplacian approaches". *BMC bioinformatics*, 14(1), p. 296

Faure, A.P. (2013). "Money and Output: The Missing Links: A Time Series Analysis", *Available at SSRN*

Ghassan, H., Fachin, S., Guendouz, A. (2013). "Financial Stability of Islamic and Conventional Banks in Saudi Arabia: a Time Series Analysis" (No. 2013/1), Centre for Empirical Economics and Econometrics, Department of Statistics

Kohler, E., Theiss, D. (2012). "From Overloaded to Opportunity: The Search for a Low-Cost Interlibrary Loan Management System", *Brick and Click Libraries*, pp. 8-20

Mastrangelo, C.M., Simpson, J.R., Montgomery, D.C. (2013). "Time Series Analysis". In *Encyclopedia of Operations Research and Management Science* (pp. 1546-1552), Springer US

Ma, Y., Zhou, G., Jiao, Y. (2014, January). "Geographical Profile Based on Time-Series Analysis Model". In *Proceedings of the 9th International Symposium on Linear Drives for Industry Applications, Volume 2* (pp. 35-42), Springer Berlin Heidelberg

Nitescu, D. (2012) "Prepayment risk, impact on credit products", *Theoretical and Applied Economics*, 19 (8), pp. 53-62