APPLICATION OF THE METHOD OF INFORMATION DECOMPOSITION FOR REVEALING THE LATENT PERIODICITY IN FINANCIAL TIME SERIES

Rudenko V.M., Turutina V.P., Korotkov E.V.
Bioengineering Center RAS, Moscow, Russia, e-mail: vrud@mail.ru

Abstract

In the given work results of search of the latent periodicity in financial time series are presented. For these purposes we used the method of information decomposition (ID) developed earlier for revealing periodicity in symbolical sequences. Application of this method to numerical rows became possible after conversion of numbers in symbols. It is shown, that the method of ID is more sensitive, than the classical approach applied to search of periodicity - Fourier analysis.

We investigated financial time series from the site $http://www.finam.ru$. Periodicity in 7 days for indexes, exchange rates, and also rates of some stocks has been found out. The received results can be interesting at studying dynamics of the securities market and prediction a behaviour of exchange rates.

1 Introduction

As the object of the research financial time series have been chosen. Revealing of essentially new properties, namely latant periodicity, in the ranks of the numerical nature became possible to application of method ID. Earlier for search of periodicity were used methods based on Fourier transformation [1], Wavelet decomposition and dynamic programming. However the methods based on Fourier transformation and Wavelet transformation develop the statistical significance of long periods (with period length greater than the size of the alphabet used in sequence) into shorter multiple periods [2]. This leads to impossibility of revealing the long fuzzy periodicity of symbolical sequence at statistically significant level. Dynamic programming also can not reveal latent periodicity in symbolical sequence since this method is used of a weight matrix for concurrences of letters (like PAM or BLOSUM) in which the weights of homologous symbol coincidences are a more greater than the weights of non-homologous coincidences.

2 Methods and algorithms

Information decomposition method (ID) has been earlier described in details in publications [2]; therefore in this section we will set it out briefly.

Let there is a symbolical sequence $S = s_1 s_2 \dots s_N$ of length N and the alphabet of the sequence contains n letters. It is necessary to determine if it has periodicity with

length of the period k. For this we compare it with artificial periodic sequence with length of the period k, constructed as follows $1, 2, \ldots k, 1, 2, \ldots k, 1, 2, \ldots k \ldots$. Then we construct matrix M which element $m(i, j)$ equal to quantity of the coincidence i symbol of the alphabet of symbolical sequence and j symbol of artificial sequence at the same position. Using this matrix the mutual information I is calculated under the formula (1):

$$I = \sum_{i=1}^{n} \sum_{j=1}^{k} m(i,j) \ln m(i,j) - \sum_{i=1}^{n} x(i) \ln x(i) - \sum_{j=1}^{k} y(j) \ln y(j) + N \ln N \qquad (1)$$

where $x(i)$, $i = 1, 2, \ldots n$ are the frequencies of symbols in investigated symbolical sequence; $y(j)$, $j = 1, 2, \ldots k$ are the frequencies of symbols in artificial periodic sequence. The value 2I has distribution χ^2 with $(n-1) \times (k-1)$ degrees of freedom, however in case of small statistics it is approximated with this distribution not well. Therefore for an estimation of the statistical importance in case of small samples we apply the method of Monte-Carlo described in [3]. The given method allows estimate the probability p_e of obtaining a value 2I as large as or larger then that obtained from the actual observations. As a quantitative measure of the statistical importance we use Z-value (2):

$$Z = -log(p_e) \qquad (2)$$

The spectrum of values Z for lengths of the periods from 2 up to N/2, where a N - length of the sequence refers to as ID of symbolical sequence.

That the method of ID will be applied to numerical rows, it is necessary to recode somehow numbers in symbols. We have encoded numerical sequence $R0 = r_1 r_2 \ldots r_N$ to symbolical sequence with the alphabet of $log_2 N + 1$ letters. We determined the minimal and maximal element of numerical series $R0$, then divided numerical interval from minimum to maximum into $log_2 N + 1$ intervals such that the number of series elements in each interval was approximately equal to $N/(log_2 N + 1)$. In the case when numerical series contained identical values, we varied the interval borders for all identical values of series to be coded by the same symbol.

We had to compare efficiency of ID method with method of Fourier transformation for search of periodicity in numerical rows. Let there is a numerical sequence $R0 = r_1 r_2 \ldots r_N$of length N. Fourier transformation for it has the form:

$$F = N^{-1/2} \sum_{m=1}^{N} r_m exp(-iq_l m) \qquad (3)$$

where $q_l = 2\pi l/N, l = 0, \ldots, N-1$.

However, for comparative studying it is desirable to consider not Fourier transformation function, but the statistics similar to Z-value which is determined by formula (2). For obtaining of this spectrum we applied as well as in case of ID the Monte-Carlo method [3]. For the numerical row of research 1000 accidental numerical rows have been generated by follow algorithm. Initially numerical row R0 has divided by subsequences $R(i)$, $i = 1 \ldots k$. Elements of subsequences $R(i)$ was equal to r_{i+k*j}, where

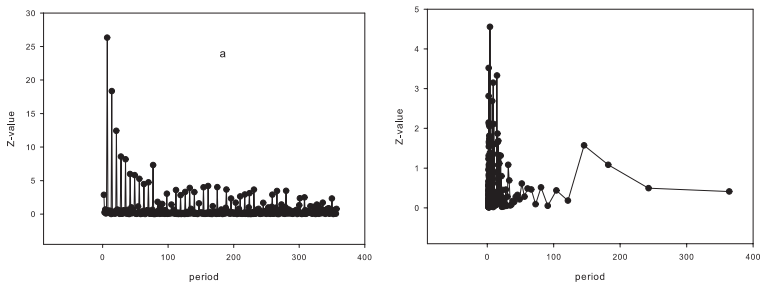

Figure 1: Application of the ID (a) and Fourier transformation (b) methods to the analysis of t_i time series for financial index NASDAQ, for 2004-2005 years.

$j = 0 \ldots N/k - 1$. Then we carried out randomization of elements inside $R(i)$ and union them back. For each of such row the function F (3) has been calculated and p_e and Z-value were determined.

3 Results

Financial time series has been taken from the site $http://www.finam.ru$. We investigated day values financial parameters r_i, where i shows days. To exclude trend influence on periodicity, we transformed initial number series to the series $t_i = (r_i - r_{i-1})/r_i$. In those cases when values r_i or r_{i-1} were missed (the tenders were not held), the value t_i was equated to symbol "-". The symbol "-" we transferred to symbolical sequence without any changing and we consider this symbol as additional letter in the symbolical sequence. When we have applied Monte-Carlo method to estimate the statistical significance by the way of the generation a set of random sequences we saved the string for symbol "-" without any changes, thus we did not consider the contribution of a symbol "-" in periodicity.

We created the software which realize ID method for numerical row and apply it for researhing the financial indexes NASDAQ and DJIA. The results obtained for NASDAQ for 2004-2005 are shown in Fig.1a. It can be seen from the figure that in the given sequence there is a periodicity with length of the period of 7 days. Unlike a method of ID, Fourier decomposition is not revealed the given period Fig. 1b.

The described research have been executed for different exchange rates and stocks. We found out the periodicity of 7 days in many exchange rates. Example for the ID spectrum for €/$ are shown in Fig.2a. Fig.2b. demonstrates that the method of Fourier trnsformation does not reveal this periodicity.

4 Discussion

The method of ID adapted for search of periodicity in numerical sequences, has allowed to find out periodicity in financial time series. It is the third example of using ID method for searching of the latent periodicity which is not found out by other mathe-

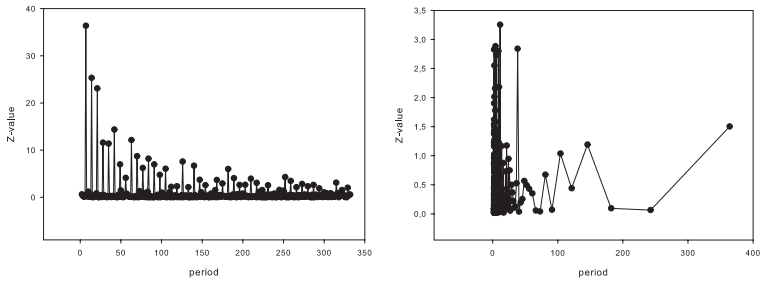

Figure 2: Application of the ID (a) and Fourier transformation (b) methods to the analysis of t_i time series for exchange rate of currencies €/$, for 2004-2005 years.

matical approaches. Earlier periodicity has been found out by us in genetic texts and poetic texts. Thus, we have once again shown, that the method of ID is a universal tool of search of the latent, poorly expressed periodicity as in symbolical and numerical rows.

Research of financial time series has shown presence of periodicity of length in 7 days in exchange rates and financial indexes. Thus periodicity is observed on long term intervals about some years. This fact is interesting enough. This phenomenon can be concerned with different behavior of people at a stock exchange on various weekdays. Presence of periodicity with other length of the period at a significant level was not revealed in currencies and indexes in any of time intervals.

It is characteristic, that periodicity of length 7 is observed for all exchanges and indexes DJIA and NASDAQ. DJIA and NASDAQ - the share indexes considering rates of securities of the basic branches of economy. Therefore it is possible to consider, that exchange rates, also as well as financial indexes, are integrated characteristics. Obviously, their parameters reflect behaviour of the share market as a whole. On the contrary, periodicity in length 7 on rates of separate securities (stocks, bonds, futures) is, more likely an exception to the rules. The common laws as in case of with currencies and financial indexes, for them it has not been allocated. The given fact proves a popular belief that the result of game at a stock exchange is more predicted and brings the stable income to the owner not one, but many types of stocks.

References

[1] Lobzin V.V., Chechetkin V.R. (2000). Order and correlations in genomic DNA sequences. The spectral approach. *Usp.Fiz.Nauk (Russian)*. Vol. **170(1)**, pp. 61-81

[2] Korotkov E.V., Korotkova M.A., Kudryshov N.A. (2003). Information decomposition method for analysis of symbolical sequences. *Phys. Let. A*. Vol. **312**, pp. 198-210.

[3] Roff D.A., Bentzen P. (1989). The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. *Mol.Biol.Evol.*, Vol. **6(5)**, pp. 539-545.