Financial Time Series Analysis Based On Normalized Mutual Information Functions.

David Blokh

C. D. Technologies Ltd., Israel

david_blokh@012.net.il

ABSTRACT

A method of predictability analysis of future values of financial time series is described. The method is based on normalized mutual information functions. In the analysis, the use of these functions allowed to refuse any restrictions on the distributions of the parameters and on the correlations between parameters. A comparative analysis of the predictability of financial time series of Tel Aviv 25 stock exchange has been carried out.

KEYWORDS

Nonlinear financial time series, Normalized mutual information function

1. INTRODUCTION

Estimation of the possibility of predicting future values of a time series is an important problem in the analysis of financial time series [1]. This problem is applicable in econophysics [2] and, in particular, in the forecast of financial markets [2,3]. In the present paper, the following version of the problem is examined:

A set of time series is specified. Find time series with the "highest" possibility of predicting future values in this set.

The principal method of the study of the time series prediction problem is the use of correlation functions that allow evaluating the possibility of linear prediction of future values of a time series. Thereby, the gaussianity of distribution and the linearity of correlations are assumed. Rather often, these conditions do not hold for financial time series [1,4]. In the present paper, functions of normalized mutual information are used for the analysis of the predictability of financial time series. It allowed to refuse any restrictions on distributions and correlations.

The possibility of using information functions for the analysis of time series was first noted in [5]. Functions of mutual information were used for the analysis in [1,6,7]. However, mutual information is an unnormalized value, and it is impossible to compare mutual information functions of different time series. Therefore, in the present paper, normalized mutual information is a measure of coupling between two time series. On the one hand, this measure allows the estimation of nonlinear correlations between time series, and on the other, the comparison of functions of normalized mutual information of different time series. Normalized mutual information has been applied with significant results in various fields of medicine [8], in

DOI: 10.5121/ijci.2013.2401

particular in oncology [9,10,11]. The approach was proposed earlier [9], as described in the monograph [12].
Using the method described in the present paper, a time series with the highest assessment of the possibility of future values prediction has been chosen from time series with the parameter "Change" of all companies of Tel Aviv 25 stock exchange. An approach to the clustering of financial time series similar to that described in the present paper is presented in [13]. Predictability estimation of financial markets also was carried out using statistical methods [14,15], wavelet analysis [16] and fractals [17].

2. ANALYSIS ALGORITHM

The analysis algorithm consists of four procedures:

- discretization of time series values;
- construction of a normalized mutual information function matrix;
- ranking of columns of the normalized mutual information function matrix;
- application of multiple comparisons method.

2.1 Discretization of Time Series Values

We transform a time series $x_i(t) = (x_i(t_1),\ldots,x_i(t_n),\ldots)$ having continuous values into a time series $y_i(t) = (y_i(t_1),\ldots,y_i(t_n),\ldots)$ having discrete values. Discretization can be performed taking into account the properties of the time series [18]. If the time series does not possess respective properties or we are unaware of them, then we can use formal rules of discretization [19].

2.2. Construction of normalized mutual information function

Let $y_i(t) = (y_i(t_1),\ldots,y_i(t_n),\ldots)$ represent time series having discrete values and $y_i(t+j) = (y_i(t_{1+j}),\ldots,y_i(t_{n+j}),\ldots)$ be a time series $y_i(t)$ with a lag j.

The normalized mutual information of the two time series $y_i(t)$ and $y_i(t+j)$ equals [8,20]

$$c_i(j) = \frac{I(y_i(t); y_i(t+j))}{H(y_i(t+j))} = \frac{H(y_i(t)) + H(y_i(t+j)) - H(y_i(t), y_i(t+j))}{H(y_i(t+j))}$$

where $H(y_i(t))$, $H(y_i(t+j))$, $H(y_i(t), y_i(t+j))$ are entropies of random values $y_i(t)$, $y_i(t+j)$, $y_i(t)\times y_i(t+j)$, respectively.

The normalized mutual information $c_i(j)$ is then calculated as a function of the lag j.

Properties of the Normalized Mutual Information Function $c_i(j)$ [8,21]:

1) $0 \le c_i(j) \le 1$;

2) $c_i(j) = 0$ if and only if $y_i(t)$ and $y_i(t+j)$ are mutually independent;

3) $c_i(j) = 1$ if and only if there exists a functional relationship between $y_i(t)$ and $y_i(t+j)$.

Consider a set of time series

$$X = \{x_1(t) = (x_1(t_1),\ldots,x_1(t_n),\ldots),\ldots,x_m(t) = (x_m(t_1),\ldots,x_m(t_n),\ldots)\}$$

having continuous values. Performing discretization, we obtain a set of time series

$$Y = \{y_1(t) = (y_1(t_1),\ldots,y_1(t_n),\ldots),\ldots,y_m(t) = (y_m(t_1),\ldots,y_m(t_n),\ldots)\}$$

We compute the normalized mutual information functions $c_i(j)$ for the time series y_i $1 \le i \le m$, $1 \le j \le k$ and obtain $m \times k$ matrix of function values $[c_i(j)]$.

2.3. Ranking of Columns of Normalized Mutual Information Function Matrix

Each row of $[c_i(j)]$ matrix is a normalized mutual information function of time series, and each column contains the values of normalized mutual information functions corresponding to the same lag. For each column of $[c_i(j)]$ matrix, we rank its entries and assign the rank 1 to the smallest entry of the column. We obtain $m \times k$ matrix of ranks $[r_i(j)]$, with each column of the matrix containing ranks from 1 to m.

We estimate the predictability of the i-th time series as compared to other time series by the sum of all the entries of i-th row of the matrix $[r_i(j)]$.

Such estimation allows us to use multiple comparisons of rank statistics for the comparison of time series predictability.

2.3. Application of Multiple Comparisons Method

We compare rank sums using the Newman-Keuls test [22]. The application of this test allows us to obtain an estimation adequate to the problem content. This test is successfully used for the analysis of biomedical data [22,23,24] and clustering financial time series [13].

3. PREDICTABILITY ESTIMATION OF TEL AVIV 25 STOCK EXCHANGE COMPANIES

We perform the predictability estimation of Tel-Aviv 25 stock exchange companies. Obviously, similar estimates can be also performed for other markets. The importance of such estimation is noted in [2]. The complexity of this problem consists in the fact that the financial time series contain nonlinearities [4].

Let $p_i(t_j)$ constitute the Adjusted Closing Price of *i*-th company on the t_j day, and $x_i(t_j)$ – the parameter “Change” equal to

$$x_i(t_j) = \frac{p_i(t_j) - p_i(t_{j-1})}{p_i(t_{j-1})}$$

.

We consider the changes in the parameter “Change” for all Tel-Aviv 25 stock exchange companies [25] in the period from June 8, 2010 till December 14, 2011 (374 days). Thus, we are dealing with a set of time series $\{x_1(t), \ldots, x_{25}(t)\}$, and the value of each series is the parameter “Change”. Thus, we have 25 time series, and each series contains 374 elements.
We carry out, step by step, the four procedures mentioned in the previous section.

1. The discretization rule is as follows [19]:

$$y_i(t_j) = -2, \text{ if } x_i(t_j) < -0.0147;$$
$$y_i(t_j) = -1, \text{ if } -0.0147 \le x_i(t_j) < -0.0044;$$
$$y_i(t_j) = 0, \text{ if } -0.0044 \le x_i(t_j) < 0.0044;$$
$$y_i(t_j) = 1, \text{ if } 0.0044 \le x_i(t_j) < 0.0147;$$
$$y_i(t_j) = 2, \text{ if } 0.0147 \le x_i(t_j).$$

Having performed the discretization, we obtain a set of time series $\{y_1(t), \ldots, y_{25}(t)\}$, whose elements assume the values of –2, –1, 0, 1, 2. The discretization is such that the quantities of elements in each category are “approximately equal”.

2. Then we compute the matrix of normalized mutual information functions – Table 1.

3. We rank entries of each column of the normalized mutual information function matrix – Table 2.

Let us consider Table 2 as the Friedman statistical model [26], and examine the row effect of this table.

Hypotheses
H_0: There is no row effect (“null hypothesis”).

H_1: The null hypothesis is invalid. *Critical range.* The sample is “large”, therefore, the critical range is the upper 1%-range of χ^2_{24} distribution.

Let us calculate the χ^2-criterion [26]. This gives us $\chi^2 = 56.05$.

The critical range is $\chi^2_{24} > 42.98$. Since 56.05 > 42.98, the null hypothesis with respect to Table 2 is rejected. Thus, according to the Friedman test, the row effect has been found. Hence, there is a difference between the rows under consideration.

4. For multiple comparisons, we use the Newman-Keuls test [22].

$\left|R_j - R_{j+1}\right| > 12.08$, where R_j and R_{j+1} are elements of the column “Sum of ranks” in the j-th and *(j+1)*-th rows of Table 3, respectively. By multiple comparisons, we construct the predictability estimation shown in Table 3.

The obtained predictability estimation possesses the following properties: for two neighboring sets of Table 3, the smallest element of one set and the greatest element of another located nearby are significantly different (α_T =0.01); elements belonging to the same set do not differ from each other (α_T =0.01). The differences between Set 1 (STRAUSS GROUP) and all the elements (companies of the stock exchange) are statistically significant (α_T =0.01). Thus, the time series of the parameter “Change” of STRAUSS GROUP company is the most predictable in comparison with time series of other Tel Aviv 25 Stock Exchange companies.

Table 1. Matrix of normalized mutual information functions.

No	Symbol \ Lag	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	AVNR.L	.012	.017	.015	.011	.015	.014	.018	.013	.023	.011	.013	.013	.007	.013	.021	.024	.014	.021	.013	.007	.010	.016
2	AZRG	.013	.028	.016	.015	.020	.022	.022	.012	.018	.012	.014	.011	.016	.015	.011	.011	.025	.015	.016	.017	.019	.021
3	ORL	.015	.013	.013	.008	.021	.022	.019	.011	.023	.019	.016	.022	.023	.013	.025	.013	.006	.011	.008	.010	.012	.009
4	BEZQ	.007	.016	.018	.017	.019	.006	.009	.009	.013	.035	.005	.014	.020	.016	.019	.019	.030	.015	.013	.017	.011	.022
5	CEL	.015	.020	.018	.009	.022	.016	.020	.015	.018	.010	.013	.029	.013	.011	.017	.013	.010	.011	.021	.014	.011	.014
6	DEDR.L	.024	.025	.019	.018	.009	.016	.012	.016	.012	.010	.018	.020	.013	.014	.018	.017	.016	.015	.023	.018	.011	.021
7	DLEKG	.016	.021	.012	.013	.017	.023	.015	.027	.015	.014	.014	.022	.014	.027	.017	.030	.014	.012	.012	.018	.015	.023
8	DSCT	.017	.022	.015	.014	.011	.014	.022	.010	.015	.009	.027	.015	.016	.015	.024	.012	.011	.015	.016	.015	.012	.007
9	ESLT	.022	.028	.007	.019	.026	.014	.028	.020	.019	.006	.010	.008	.016	.013	.016	.016	.012	.012	.014	.009	.015	.013
10	GLOB	.014	.012	.012	.016	.021	.016	.016	.018	.012	.008	.016	.020	.020	.012	.017	.014	.017	.011	.011	.022	.014	.015
11	HOT	.013	.018	.013	.012	.005	.016	.011	.024	.013	.012	.014	.009	.011	.011	.013	.012	.016	.019	.014	.016	.013	.014
12	ICL	.017	.010	.012	.020	.019	.014	.023	.016	.011	.014	.018	.013	.012	.007	.027	.010	.010	.017	.013	.013	.009	.017
13	ILCO	.032	.020	.012	.015	.017	.024	.017	.010	.014	.015	.012	.025	.011	.014	.017	.015	.017	.014	.011	.032	.016	.016
14	ISRA.L	.017	.019	.017	.010	.018	.025	.023	.009	.018	.010	.013	.007	.014	.028	.026	.012	.017	.009	.017	.017	.033	.027
15	LUMI	.024	.026	.016	.016	.022	.013	.011	.015	.030	.015	.013	.015	.027	.015	.010	.016	.018	.010	.013	.007	.012	.018
16	MLNX	.019	.013	.017	.019	.014	.011	.016	.017	.014	.014	.020	.014	.011	.015	.011	.021	.015	.013	.017	.013	.013	.025
17	MZTF	.011	.023	.006	.014	.010	.016	.011	.011	.013	.027	.011	.005	.013	.019	.010	.017	.016	.015	.014	.018	.026	.015
18	NICE	.029	.016	.019	.018	.015	.015	.015	.011	.008	.013	.022	.013	.015	.011	.010	.008	.014	.011	.014	.008	.021	.010
19	OSEM	.031	.014	.015	.011	.011	.027	.024	.018	.020	.015	.017	.012	.031	.013	.015	.020	.021	.007	.012	.011	.017	.012
20	PTNR	.014	.018	.020	.022	.013	.012	.020	.016	.010	.014	.008	.023	.014	.016	.010	.022	.022	.014	.016	.021	.022	.019
21	PZOL	.018	.013	.017	.018	.007	.008	.010	.016	.013	.014	.020	.017	.025	.021	.030	.016	.020	.013	.011	.011	.012	.012
22	PRGO	.020	.015	.005	.013	.010	.011	.016	.011	.018	.017	.009	.014	.010	.021	.014	.010	.016	.014	.013	.012	.016	.012
23	POLI	.018	.027	.012	.020	.017	.011	.008	.009	.016	.011	.016	.019	.015	.013	.013	.013	.010	.012	.008	.018	.024	.015
24	STRS	.027	.029	.030	.034	.026	.035	.021	.025	.024	.033	.018	.018	.036	.021	.013	.013	.031	.011	.034	.028	.020	.025
25	TEVA	.029	.010	.022	.017	.018	.008	.020	.021	.013	.012	.019	.012	.030	.015	.011	.009	.010	.007	.022	.012	.014	.012

Table 2. Ranked entries of each column of the normalized mutual information function matrix.

No	Symbol	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Sum of ranks
1	AVNR.L	3	11	11	4	10	11	14	11	22	8	9	10	1	10	20	24	10	25	11	1	2	15	243
2	AZRG	5	23	15	11	19	19	20	10	18	10	13	5	17	13	5	5	23	20	19	16	19	19	324
3	ORL	8	6	10	1	21	20	15	6	23	22	14	22	20	8	22	10	1	5	1	5	8	2	250
4	BEZQ	1	9	19	16	17	1	2	2	10	25	1	13	18	19	19	20	24	18	10	17	4	21	286
5	CEL	9	15	20	2	23	18	17	13	19	4	7	25	9	4	16	11,5	2	9	22	12	5	9	271,5
6	DEDR.L	19	20	22	18	3	17	7	17	5	5	19	19	7	12	18	18	12	22	24	18	3	20	325
7	DLEKG	10	17	6	7	13	21	9	25	14	13	11	21	12	24	17	25	9	11	6	20	15	22	328
8	DSCT	13	18	13	9	6	10	21	5	13	3	25	14	16	14	21	7	6	21	17	13	7	1	273
9	ESLT	18	24	3	21	24	12	25	21	20	1	4	3	15	9	13	15	7	12	13	4	14	8	286
10	GLOB	7	3	7	14	20	15	12	20	4	2	15	20	19	5	15	13	17	6	5	23	12	11	265
11	HOT	4	13	9	6	1	14	4	23	9	9	12	4	5	2	9	8	14	24	14	14	10	10	218
12	ICL	12	1	4	23	18	9	22	15	3	16	18	9	6	1	24	4	3	23	12	10	1	16	250
13	ILCO	25	16	8	12	14	22	13	4	12	18	6	24	4	11	14	14	18	15	4	25	17	14	310
14	ISRA.L	11	14	17	3	15	23	23	3	17	6	10	2	10	25	23	6	16	3	20	15	25	25	312
15	LUMI	20	21	14	13	22	8	6	12	25	20	8	15	22	17	3	17	19	4	8	2	9	17	302
16	MLNX	16	5	18	20	9	6	10	18	11	17	22	11	3	15	7	22	11	13	21	11	11	24	301
17	MZTF	2	19	2	10	4	16	5	8	8	23	5	1	8	20	1	19	13	19	15	19	24	13	254
18	NICE	22	10	21	19	11	13	8	7	1	12	24	8	13	3	2	1	8	8	16	3	21	3	234
19	OSEM	24	7	12	5	7	24	24	19	21	19	17	6	24	6	12	21	21	1	7	6	18	7	308
20	PTNR	6	12	23	24	8	7	16	14	2	15	2	23	11	18	4	23	22	16	18	22	22	18	326
21	PZOL	15	4	16	17	2	2	3	16	6	14	23	16	21	21	25	16	20	14	3	7	6	4	271
22	PRGO	17	8	1	8	5	4	11	9	16	21	3	12	2	23	11	3	15	17	9	9	16	6	226
23	POLI	14	22	5	22	12	5	1	1	15	7	16	18	14	7	8	11,5	5	10	2	21	23	12	251,5
24	STRS	21	25	25	25	25	25	19	24	24	24	20	17	25	22	10	9	25	7	25	24	20	23	464
25	TEVA	23	2	24	15	16	3	18	22	7	11	21	7	23	16	6	2	4	2	23	8	13	5	271

Table 3. Predictability estimation of Tel Aviv 25 Stock Exchange Companies

No. of set	The degree of predictability	Company name	Symbol	Sum of ranks
1	Large predictability	STRAUSS GROUP	STRS	464
2.1	Intermediate predictability	DELEK GROUP	DLEKG	328
2.2		PARTNER	PTNR	326
2.3		DELEK DRILLIN L	DEDR.L	325
2.4		AZRIELI GROUP	AZRG	324
2.5		ISRAMCO L	ISRA.L	312
2.6		ISRAEL CORP	ILCO	310
2.7		OSEM	OSEM	308
2.8		LEUMI	LUMI	302
2.9		MELLANOX	MLNX	301
3.1	Small predictability 1	BEZEQ	BEZQ	286
3.2		ELBIT SYSTEMS	ESLT	286
4.1	Small predictability 2	DISCOUNT	DSCT	273
4.2		CELLCOM	CEL	271,5
4.3		PAZ OIL	PZOL	271
4.4		TEVA	TEVA	271
4.5		GAZIT GLOBE	GLOB	265
4.6		MIZRAHI TEFAHOT	MZTF	254
4.7		POALIM	POLI	251,5
4.8		ICL	ICL	250
4.9		BAZAN	ORL	250
4.10		AVNER L	AVNR.L	243
4.11		NICE	NICE	234
4.12		PERRIGO	PRGO	226
4.13		HOT	HOT	218

4. CONCLUSION

The analysis of financial time series based on the functions of normalized mutual information allows predictability assessment without the assumptions of gaussianity and linearity. The use of these functions also allows us to compare the predictabilities of different time series. At present, there are no such advanced prediction methods for the functions of normalized mutual information as for the analysis based on correlation functions. The development of prediction methods for time series analysis based on the functions of normalized mutual information should become an object of further research.

REFERENCES

[1] M. Small (2005) "Applied Nonlinear Time Series Analysis. Applications in Physics, Physiology and Finance," World Scientific Publishing Co. Pte. Ltd., Singapore.

[2] R.N. Mantegna, H.E. Stanley, (2000) "An Introduction to Econophysics. Correlations and Complexity in Finance," Cambridge University Press, Cambridge.

[3] R. N. Mantegna, Z. Palágyi, H.E. Stanley, (1999) "Applications of statistical mechanics to finance," Physica A: Statistical Mechanics and its Applications, Vol. 274, No. 1-2, pp. 216-221.

[4] Cont, (2002) "Empirical properties of asset returns: stylized facts and statistical issues," Quantitative Finance, Vol. 1, No. 2, pp. 223-236.

[5] W. Li,(1999) "Mutual Information Functions Versus Correlation Functions," Journal of Statistical Physics, Vol. 60, No. 5-6, pp. 823-837.

[6] G. A. Darbellay, D.Wuertz, (2000) "The entropy as a tool for analysing statistical dependences in financial time series," Physica A: Statistical Mechanics and its Applications, Vol. 287, No. 3-4, pp. 429-439.

[7] A. Dionisio, R. Menezes, D. Mendes, (2004) "Mutual information: a measure of dependency for nonlinear time series," Physica A: Statistical Mechanics and its Applications, Vol. 344, No. 1-2, pp. 326-329.

[8] J. Zvarova, M. Studeny, (1997) "Information theoretical approach to constitution and reduction of medical data," International Journal of Medical Informatics, Vol. 45, No. 1-2, pp. 65-74.

[9] D. Blokh, I.Stambler, E. Afrimzon, et al. (2007) "The information-theory analysis of Michaelis-Menten constants for detection of breast cancer," Cancer Detection and Prevention, Vol. 31, No. 6, pp. 489-498.

[10] D. Blokh, N. Zurgil, I. Stambler, et al. (2008) "An information-theoretical model for breast cancer detection," Methods of Information in Medicine, Vol. 47, no. 4, pp. 322-327.

[11] D. Blokh, I. Stambler, E. Afrimzon, et al. (2009) "Comparative analysis of cell parameter groups for breast cancer detection," Computer Methods and Programs in Biomedicine, Vol. 94, No. 3, pp. 239-249.

[12] P. J. Gutierrez Diez, I.H. Russo, J. Russo, (2012) "The Evolution of the Use of Mathematics in Cancer Research," Springer, New York.

[13] D. Blokh, (2012) "Clustering financial time series via information-theory analysis and rank statistics," Journal of Pattern Recognition Research, Vol. 7, No. 1, pp. 106-115. http://www.jprr.org/index.php/jprr/article/viewFile/396/166

[14] D. K. Patro, Y. Wu, (2004) "Predictability of short-horizon returns in international equity markets," Journal of Empirical Finance, Vol. 11, No. 4, pp. 553-584.

[15] M. Cooper, R. C. Gutierrez, Jr., B. Marcum, (2005) "On the Predictability of Stock Returns in Real Time," Journal of Business, Vol. 78, No. 2, pp. 469-500. http://www.jstor.org/stable/full/10.1086/427635

[16] N.A. Kyaw, C.A. Los, and S. Zong, (2006) "Persistence characteristics of Latin American financial markets", Journal of Multinational Financial Management, Vol. 16, No. 3, pp. 269–290.

[17] G. R. Richards, (2004) "A Fractal Forecasting Model for Financial Time Series", Journal of Forecasting, Vol. 23, No. 8, pp. 587-602.

[18] G. Nicolis, I. Prigogine, (1990) "Exploring Complexity," W.H. Freeman, New York.

[19] G.V. Glass, J.C. Stanley, (1970) "Statistical Methods in Education and Psychology," Prentice- Hall, New Jersey.

[20] A. Renyi, (1959) "On measures of dependence," Acta Mathematica Academiae Scientiarum Hungaricae, Vol. 10, No. 3-4, pp. 441-451.

[21] T.M. Cover, J.A. Thomas, (2006) "Elements of Information Theory," 2nd ed., Wiley-Interscience, New York.

[22] S.A. Glantz, (1994) "Primer of Biostatistics," 4th ed., McGraw-Hill, New York.

[23] S.M. Pupa, S. Giuffre, F.Castiglioni,"Regulation of Breast Cancer Response to Chemotherapy by Fibulin-1," (2007) Cancer Research, Vol. 67, No.9, pp. 4271-4277.

[24] S.S. Wu, W. Wang, D.H. Annis, "On Identification of the Number of Best Treatments Using the Newman-Keuls Test," (2008) Biometrical Journal, Vol. 50, No. 5, pp. 861-869.

[25] Tel-Aviv,25stockexchange http://www.tase.co.il/TASEEng/MarketData/Indices/MarketCap/IndexMainDataMarket.htm?Action=1&IndexID=142) [26] W.J. Conover, (1999) "Practical Nonparametric Statistics," Wiley-Interscience, New York.