Available online at www.sciencedirect.com

ScienceDirect

Physica A 376 (2007) 525–531

PHYSICA A

www.elsevier.com/locate/physa

Dynamical structures of high-frequency financial data

Kyungsik Kim[a,*], Seong-Min Yoon[b], SooYong Kim[c], Ki-Ho Chang[d], Yup Kim[e], Sang Hoon Kang[f]

[a]*Department of Physics, Pukyong National University, Pusan 608-737, Republic of Korea*
[b]*Division of Economics, Pukyong National University, Pusan 608-737, Republic of Korea*
[c]*Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Republic of Korea*
[d]*Remote Sensing Research Laboratory, Meteorological Research Institute, KMA, Seoul 156-720, Republic of Korea*
[e]*Department of Physics, Kyung Hee University, Seoul 130-701, Republic of Korea*
[f]*School of Commerce, University of South Australia, SA 5001, Australia*

Received 10 April 2006; received in revised form 10 October 2006
Available online 13 November 2006

Abstract

We study the dynamical behavior of high-frequency data from the Korean Stock Price Index (KOSPI) using the movement of returns in Korean financial markets. The dynamical behavior of a binarized series of our models is not completely random. In addition, the conditional probability is numerically estimated from a return series of KOSPI tick data. Non-trivial probability structures can be constituted from binary time series of autoregressive (AR), logit, and probit models, for which the Akaike Information Criterion shows a minimum value at the 15th order. From our results, we find that the value of the correct match ratio for the AR model is slightly larger than that derived by other models.

Keywords: Correlation function; Conditional probability distribution; AR model; Logit and probit models; KOSPI

1. Introduction

Recently, the study of differently scaled economic systems has received a considerable attention as an interdisciplinary field of physicists and economists [1–7]. One of the challenging issues is to test efficient market hypotheses from the perspective of empirical observations and theoretical considerations [8]. Realization of the capacity to exploit or predict the dynamical behavior of continuous tick data for various financial assets [9,10] is extremely desirable. Some financial markets are not perfectly efficient as are others at some points in time. In such cases, investors or agents in the financial market can accrue profits by accurate prediction using market information and strategies, and their trading behavior usually reinforces the effective network among them. For example, when the rice stock rises or falls in price, a trader's decision to buy or sell is influenced by various strategies, external information, and other traders. One such strategy is to apply the

*Corresponding author. Tel.: +82 51 620 6354; fax: +82 51 611 6357.
E-mail address: kskim@pknu.ac.kr (K. Kim).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2006.10.054

up and down movement of returns to a correlation function and the conditional probability. This strategy, which is pivotal for predicting an investment, is a useful tool for understanding the stock transactions of a company whose stock price is rising or falling. Ohira et al. [9] discussed conditional probability and the correct match ratio of high-frequency data for the yen–dollar exchange rate; they showed that such dynamics are not completely random and that a probabilistic structure exists. Sazuka et al. [10] used the order $k = 10$ of the Akaike Information Criterion (IC) to determine the predictable value of the autoregressive (AR) model. In contrast, they numerically calculated the fifth order of the logit model [11]. Motivated by such research, we apply and analyze the AR, logit, and probit models in application to the Korean financial market, which, in contrast to active and well-established financial markets, is now in a slightly unstable and risky state.

Interest in nonlinear models has recently grown, particularly in the social, natural, medical, and engineering sciences. Statistical and mathematical physics provide powerful and rigorous tools for analyzing social data. Moreover, several papers have focused on social phenomena models based on aspects of stochastic analyses, such as the diffusion, master, Langevin, and Fokker–Planck equations. Many researchers in econometrics or biometrics have proposed the use of AR, logit, and probit models in the formulation of the discrete choices, including binary analyses. Interestingly, Nakayama and Nakamura [16] associated the fashion phenomena of the bandwagon and snob effects with the logit model. To our knowledge, in addition to the Akaike IC, there are at least two other similar standards, the Hannan–Quinn IC and the Schwarz IC. However, we restrict ourselves to utilizing the Akaike IC as the residual test in order to minimize the remained value for the binary analysis. Moreover, after calculating the binary structures and their Akaike IC value, we compute the correct match ratio, or the power of predictability. Although the dynamical behavior of logit and probit models has been calculated and analyzed in scientific fields such as mathematics, economics, and econophysics, until now these models have not been studied in detail with respect to financial markets [12–15].

In this letter, we present the future predictability function of the AR, logit, and probit models using a tick data analysis of the Korean Stock Price Index (KOSPI) for the Korean financial market. The obtained results are of great importance for developing a powerful and capable tool that can be used to investigate properties of efficient and predictable markets. In Section 2, by examining the binary phenomena of a financial time series in terms of a nontrivial probability distribution, we show that the high-frequency data of our model follows a special conditional probability structure for the up and down movement of returns. In Section 3, the dynamical behavior for our models is calculated numerically. Conclusions are presented in the final section.

2. Models

In our calculations, the return of the tick data at time t is $R(t) = \ln p(t+\Delta t)/p(t)$ for the price $p(t)$, and the return increment is $D(t) \equiv R(t+1) - R(t)$ for every time t. From the series of tick data in one asset, we can binarize the $\{X(t)\}$ series as follows: $X(t) = +1$ if $D(t) > 0$ and $X(t) = -1$ if $D(t) < 0$. We can then extend the $\{X(t)\}$ series to a random walk formalism as $Z(t+1) = Z(t) + X(t)$. Moreover, we can determine the cumulative probability distribution and the conditional probabilities from the random walk of the one-directional zigzag motion. The correlation function can also be calculated as

$$C(u) = \langle D(t+u)D(t) \rangle. \tag{1}$$

We now introduce the AR, logit, and probit models [11–14] for an $\{X(t)\}$ series of continuous tick data. The AR model is defined by

$$\mathrm{AR}(k) = \alpha_0 + \sum_{i=1}^{k} \alpha_i X(t-i) + \varepsilon(t), \tag{2}$$

where $\varepsilon(t)$ is a white noise with Gaussian distribution of zero mean and variance σ. The standard logit model for the binary analysis [12] is described as

$$\log \mathrm{it}(p) = \log \frac{p}{1-p} = \beta_0 + \sum_{i=1}^{k} \beta_i X(t-i) + \varepsilon(t), \tag{3}$$

where p is a dummy variable between 0 and 1. The linear probit model from Eq. (3) is represented in terms of

$$\text{probit}(p) = \Phi^{-1}(p) = z_p, \tag{4}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function, and the standard normal cumulative distribution function is given by $\Phi(z_p) = \Pr(z \leqslant z_p) = (1/\sqrt{2\pi}) \int_{-\infty}^{z_p} \mathrm{d}z \exp(-z^2/2)$.

Furthermore, we make use of Eqs. (2)–(4) to determine the binary structure and its correct match ratio. These mathematical techniques yield more general predictability results [17–21]. To determine the minimized order k of our model, we define the Akaike IC [17,18] as

$$\text{AIC} = \frac{2}{T}[-\ln Ml + \ln Mp], \tag{5}$$

for a sample size T, where Ml and Mp denote the maximum likelihood and the number of parameters, respectively.

3. Numerical results

To analyze the correlation function and the conditional probability, we introduce our underlying asset into the KOSPI in the Korean financial market. First, we consider two delivery periods: the first set of data, Data A, was from January 1997 to December 1998; the second set, Data B, was from January 2004 to December 2004. The lag time of the two sets of tick data is about one minute. Data A contains 133, 823 items of data and Data B contains 86, 561 items.

From the two tick data, we computed two series: the $X(t)$ series and the $Z(t)$ series, where $Z(t)$ represents a one-dimensional zigzag motion. This computation refers to a binary strategy of the buy and sell trend of traders in financial markets. Fig. 1 plots the correlation function, $C(u)$, which we obtained from the return increment $D(t)$. The plot suggests that the minute returns for Data A of the KOSPI are not entirely independent of, or different from, the random walk model but almost independent for long periods. Given the probabilistic structure of our model, we can deduce from the correlation function that the dynamical behavior is completely nonrandom.

By quantitative analysis, we can relate the $X(t)$ series to the conditional probability. To analyze the high-frequency data of the KOSPI, we concentrated on the up and down return movements in terms of conditional probability. The parameter $P(+|+,+)$ refers to the conditional probability that the price returns will likely move in the same direction; that is, that the price is likely to rise after two consecutive steps in the same direction. Table 1 summarizes the results of various conditional probabilities for Data A and Data B of the KOSPI. Fig. 2 shows that the conditional probability of $P(+|+,+,+)$ has a remarkably larger value than the probability of $P(+|m=3)$, except $P(+|+,+,+)$. From our results, we can give the relation of the three parameters as $P(+|+,+) = p$, $P(+|-,+) = q$, and $P(+|+,+,+) = p + \alpha$ for $0 < \alpha < p < 1$, $0 < q < 1$.

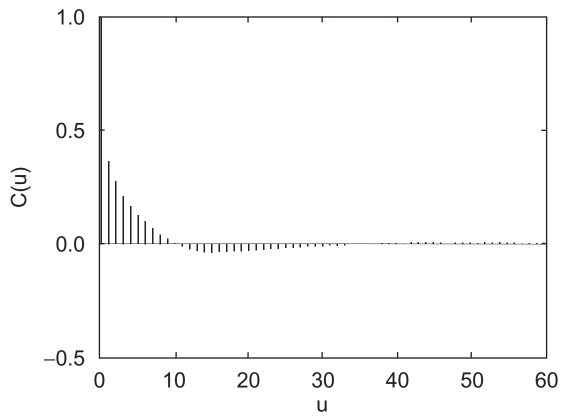

Fig. 1. Plot of correlation function $C(u)$ from the set of minutely tick data, Data A of the KOSPI; the data were collected from January 1997 to December 1998.

Table 1
Values of conditional probability from the simulation results of Data A and Data B

KOSPI	Data A	Data B
NP	133, 823	86, 561
$P(+)$	48.65	49.80
$P(+\|+)$	57.85	50.37
$P(+\|-)$	39.93	49.23
$P(+\|+,+)$	68.48	52.25
$P(+\|+,-)$	53.10	51.62
$P(+\|-,+)$	43.26	48.46
$P(+\|-,-)$	31.18	46.92
$P(+\|+,+,+)$	74.70	53.27
$P(+\|+,-,+)$	50.90	48.98
$P(+\|+,-,-)$	42.98	48.41
$P(+\|-,-,+)$	34.61	47.90
$P(+\|-,-,-)$	25.83	45.61
$P(+\|+,+,+,+)$	77.51	54.04
$P(+\|+,+,+,-)$	66.42	53.05
$P(+\|+,+,-,+)$	56.17	50.10
$P(+\|+,-,-,+)$	40.60	47.48
$P(+\|+,-,-,-)$	34.24	46.12
$P(+\|-,+,+,+)$	66.39	52.39
$P(+\|-,-,+,-)$	40.21	49.88
$P(+\|-,-,-,+)$	30.09	48.31
$P(+\|-,-,-,-)$	22.90	45.19

NP stands for the number of tick data points.

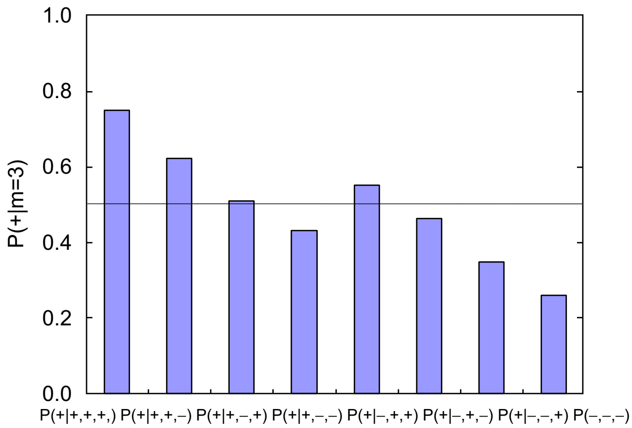

Fig. 2. Conditional probabilities $P(+|m=3)$ for the set of minutely tick data, Data A of the KOSPI.

Fig. 3 shows the conditional probability in the case of the same direction of price movement. The conditional probability $P(+|m)$ ($P(-|m)$) has a larger value than $P((+|m+1)$ ($P((+|m-1))$, which exists for one selling state (buying state) after $m-1$ selling states ($m-1$ buying) states, where m is the value of buying (selling) states in the same direction of price increasing (decreasing) movement. When we compare this result to that of the yen–dollar exchange rate of the Japanese financial market, our conditional probabilities for $m<5$ have a slightly larger value than those of the yen–dollar exchange rate [9]. The values of $P(+|m)$ and $P(-|m)$ for $m<6$ increase continuously while the two values for $m\geqslant 6$ are almost constant; in this case, the period of the m states is about m minutes in real time. We predict this result to be consistent with the buy–sell strategy of dealers who can change in a few minutes. Note that although Data A and Data B share significant

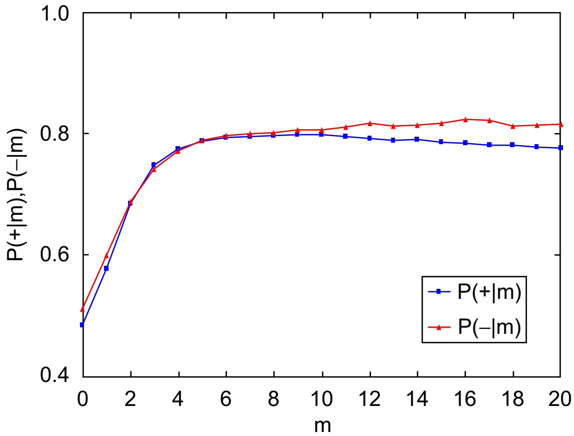

Fig. 3. Plot of conditional probabilities $P(+|m)$ and $P(-|m)$ for the set of minutely tick data, Data A of the KOSPI.

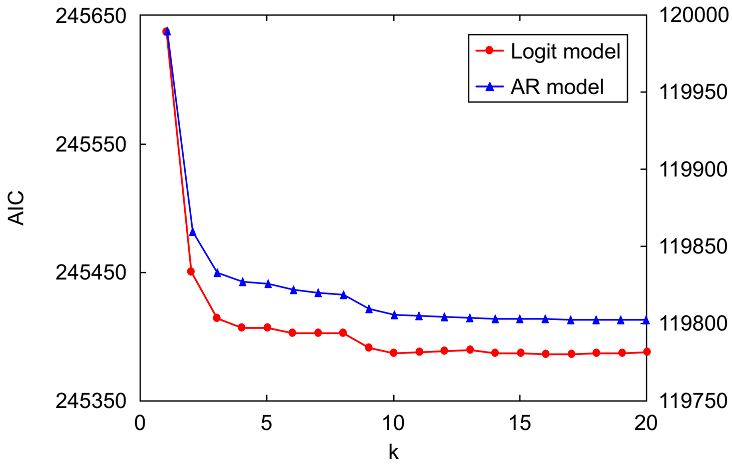

Fig. 4. Plot of the Akaike IC values for the AR model (the value of the left y-axis) for Data A and of the logit model (the value of the right y-axis) for Data B; in each case, the Akaike IC value decreases gradually as the order of model grows.

similarity, we cannot understand the behavior of these data sets from a random walk model that has fixed values for conditional probabilities.

For simplicity, we used the AR, logit, and probit models to analyze the $X(t)$ series for high-frequency tick data of the Korean financial market. As shown in Fig. 4, we found that the Akaike IC values for the AR and logit models decrease gradually as the order of the models increases. Because the Akaike IC for the three models has approximately the same value in a range larger than an order of $k = 15$, we consider this value to be the minimum value; in addition, this value is similar to the 10th order of the AR model of the yen–dollar exchange rate [11]. Hence, the function shape of the logit model is similar to that of the probit model, and each probability structure tends to move continuously in the same direction. By minimizing the Akaike IC value of our model, we were also able to calculate the correct match ratio. The correct match ratio is a measure that can evaluate the predictability of a forecasting model. It measures the degree of consistency between the real direction of price movement and the predicted one. Table 2 shows the values of the correct match ratios for Data A and Data B. The AR model of Data A has a higher value than other models for the correct match ratio; in contrast, the logit model of Data B has a smaller value.

Table 2
Values of the correct match ratio from the simulation results of Data A and Data B

KOSPI	Data A(%)	Data B(%)
NP	133, 823	86, 561
AR model	65.3	52.2
Probit model	51.4	50.2
Logit model	48.6	49.8

4. Conclusions

We used the AR, logit, and probit models to determine the probability structure of high-frequency tick data of the KOSPI in the Korean financial market. The value of our conditional probability of the KOSPI is slightly greater than that of the yen–dollar exchange rate. Our results show that the Korean financial market is slightly unstable and less systematic than other financial markets, though the results may be related to actual transactions of all assets. In addition, by using the AR, probit, and logit models, we deduced that the forecasted (or simulated) sign is equal to the sign of the actual returns. This deduction enables us to obtain the correct match ratio. Moreover, because the match ratio is always greater than 0.5, we can conclude that our model has improved forecasting capability. The AR model, which is expected to have a higher predictable value only in the Korean financial market, robustly supports the future predictability of price movement trends in financial markets. We also note that, with nonlinear models of data analysis, international finance theories can offer an enhanced interpretation of results. For the past decade, many econophysical investigations have led to greater appreciation of, and insight into, scale invariance and the universality of statistical approaches to physics and economics. Our results should encourage interdisciplinary research of physics and economics.

Acknowledgment

This work was supported by Grant no. R01-2006-000-10470-0 from the Basic Program of the Korea Science and Engineering Foundation.

References

[1] R.N. Mantegna, H.E. Stanley, An Introduction to Econophysics: Correlation and Complexity in Finance, Cambridge University Press, Cambridge, 2000.
[2] Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng, H.E. Stanley, Phys. Rev. E 60 (1999) 1390.
[3] X. Gabaix, P. Gopikrishnan, V. Plerou, H.E. Stanley, Nature 423 (2003) 267.
[4] F. Lillo, R.N. Mantegna, Phys. Rev. E 68 (2003) 016119.
[5] P.Ch. Ivanov, A. Yuen, B. Podobnik, Y. Lee, Phys. Rev. E 69 (2004) 056107.
[6] E. Scalas, R. Gorenflo, H. Luckock, F. Mainardi, M. Mantelli, M. Raberto, Phys. Rev. E 69 (2004) 011107;
J. Masoliver, M. Montero, G.H. Weiss, Phys. Rev. E 67 (2003) 021112.
[7] L.S. Sabatelli, S. Keating, J. Dudley, P. Richmond, Eur. Phys. J. B 27 (2002) 273;
K. Yamasaki, L. Muchnik, S. Havlin, A. Bunde, H.E. Stanley, Proc. Natl. Acad. Sci. 102 (2005) 9424.
[8] Y.-C. Zhang, Physica A 269 (1999) 30.
[9] T. Ohira, N. Sazuka, K. Marumo, T. Shimizu, M. Takayasu, H. Takayasu, Physica A 308 (2002) 368.
[10] N. Sazuka, T. Ohira, K. Marumo, T. Shimizu, M. Takayasu, H. Takayasu, Physica A 324 (2003) 366.
[11] N. Sazuka, Physica A 355 (2005) 183.
[12] W. Weidlich, Phys. Rep. 204 (1991) 1;
W. Weidlich, G. Haag, Concepts and Models of a Quantatitive Sociology: The Dynamics of Interacting Populations, Springer, Berlin, 1983.
[13] D. McFadden, Frontiers in econometrics: conditional logit analysis of qualitative choice behavior, in: P. Zarembka (Ed.), Academic Press, New York, 2000, pp. 105–142.
[14] M. Ben-Akiva, S.R. Lerman, Discrete Choice Analysis: Theory and Application to Travel Demand, MIT Press, Cambridge, MA, 1985.

[15] J. Aldrich, F. Nelson, Linear Probability, Logit, and Probit Models, Sage Publications, Beverly Hills, CA, 1984.
[16] S. Nakayama, Y. Nakamura, Physica A 337 (2004) 625.
[17] H. Akaike, Canonical correlation analysis of time series and the use of an information criterion, in: R.K. Mehra, D.G. Lainotis (Eds.), Academic Press, New York, 1976, pp. 52–107.
[18] H. Akaike, in: B.N. Petrov, F. Csaki (Eds.), Second International Symposium on Information Theory, Akademia Kiado, Budapest, 1973, pp. 267–281;
R.S. Tsay, Analysis of Financial Time Series, Wiley, New York, USA, 2002, pp. 314–322.
[19] E.J. Hannan, B.G. Quinn, J. R. Stat. Soc. B 41 (1979) 190.
[20] E.J. Hannan, Ann. Stat. 8 (1980) 1071.
[21] G. Schwarz, Ann. Stat. 6 (1978) 461.