数理解析研究所講究録
第 1620 巻 2009 年 27-38

Nonlinear Time Series Analysis of Stock Return Variation

Takaya Miyano* and Kenichi Tatsumi†

*Department of Micro System Technology, Ritsumeikan University
1-1-1 Noji-Higashi, Kusatsu, Shiga 525-8577, Japan

†Graduate School of Economics, Gakushuin University
1-5-1 Mejiro, Toshima-ku, Tokyo 171-8588, Japan

Abstract

We tested for the days-of-the-week effect on stock return variation using NIKKEI 225 stock index data ranging from the 4th of January in 1989 to the 29th of August in 2003. Nonlinear time series analysis with the surrogate method suggests that there may be a particular weekly pattern in stock return variation such that stock return is significantly lower on Monday.

1 Introduction

There has been a long controversial conjecture that stock return might display dependence on the days of the week [1]. We may call this as the days-of-the-week effect, which is one of anomalies that cannot be explained within the framework of the standard theories of economics. In the United States, there has been an uncertain observation such that returns on investments in stocks might be lower on Monday, which is called the Monday effect. Possible existence of the days-of-the-week effect has been investigated using regression analysis with inconclusive results in previous literature.

In this study, we test for the days-of-the-week effect on daily stock return in the Japanese stock market using nonlinear time series analysis. Our approach is practical but has not been taken in previous literature. We use a time series of NIKKEI 225 stock index, i.e., mean price over 225 stocks at the Tokyo Stock Exchange. The index data are deficient on Saturdays, Sundays and the national holidays when the stock market is closed. To recover the missing data, linear interpolation is applied to the time series. As the methods for time series analysis, we apply the diagnostic algorithm developed by Wayland et al. [2] and the surrogate method introduced by Theiler et al. [3, 4]. These methods have been devised to test for chaos from complex time series. We may not expect that stock return variation represents chaotic behavior. Nevertheless, the days-of-week effect is supposed to be captured in terms of the degree of complexity in stock return variation. These mathematical methods are suitable for this purpose.

This paper is organized as follows. In section 2, we describe the method for interpolating data, i.e., linear interpolation. Linear interpolation smoothly recovers the data deficiencies, in particular, those on between Friday and Monday. This property is very important for our purpose, as discussed in this paper. The mathematics for the time series analysis are presented in section 3. Results of numerical analysis are given in section 4. We discuss the results and make conclusions in section 5.

2 Interpolating data

We used a time series of daily closing stock index ranging from the 4th (Wednesday) of January in 1989 to the 29th (Friday) of August in 2003, issued by the Tokyo Stock Exchange. The stock data are deficient on Saturdays, Sundays and the national holidays. Linear interpolation was used to recover these missing data points: We synthesized n successive missing data points, denoted as $x(t+i)$ $(i = 1, \cdots, n)$, between stock data $x(t)$ and $x(t+n+1)$ by

$$x(t+i) = \frac{n-i+1}{n+1}x(t) + \frac{i}{n+1}x(t+n+1) \tag{1}$$

We thus obtained an interpolated time series consisting of $N = 5348$ data points, as shown in Fig. 1. Daily stock return $r(t)$ $(t = 2, \cdots, N)$ was estimated from the interpolated time series by

$$r(t) = \frac{x(t) - x(t-1)}{x(t)}. \tag{2}$$

Figure 2 shows the time series $\{r(t)\}_{t=2}^{N}$.

3 Methods for time series analysis

Given a time series $\{x(t)\}_{t=1}^{N}$ as a set of observational data for a single observable quantity in a system of Q degrees of freedom, we can represent the dynamical behavior of the system using embedology [5, 6]: We reconstruct trajectories as lagged sequences of data points $\mathbf{x}(t) = (x(t), x(t+\Delta t), \cdots, x(t+(D-1)\Delta t)$ where D is the embedding dimension and Δt is

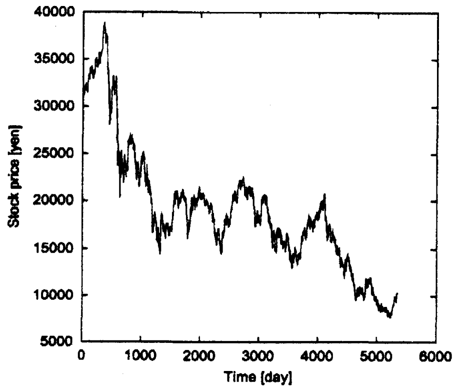

Figure 1: NIKKEI stock index with interpolations on Saturdays, Sundays, and national holidays.

an appropriately chosen time lag. The sufficient condition for a map of the trajectories from Q-dimensional space to D-dimensional reconstructed space to be a diffeomorphism is $D \geq 2Q+1$ [5]. In practical applications, Q is usually unknown and D is determined by trial-and-error. The optimal time lag Δt can be inferred from autocorrelation function or mutual information. In this work, we determined Δt on the basis of mutual information [7, 8].

3.1 Mutual information

Mutual information as a function of the time distance between data points of a time series, introduced by Fraser and Swinney [7, 8], provides useful information for determining the optimal time lag for embedding as well as for measuring the rate of spontaneous decay of information from the time series. Let U and V be sets of realizations of the random variables u and v, respectively. We approximate the probability density functions $p_U(u)$ and $p_V(v)$ and the joint probability density function $p(u,v)$ from 1-dimensional and 2-dimensional histograms generated from sets of sample data $\{u_i\}_{i=1}^N$ and $\{v_i\}_{i=1}^N$. Here, N is the number of examples. The information entropies $H(U)$ and $H(V)$ and the joint information entropy $H(U,V)$ are

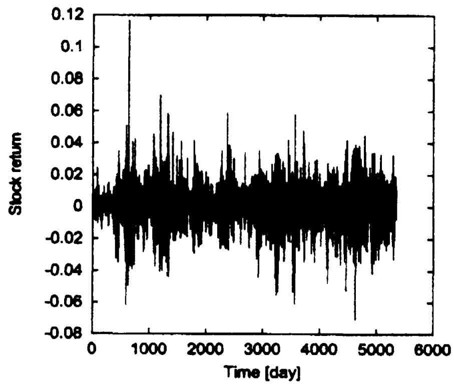

Figure 2: NIKKEI stock return.

calculated by

$$
\begin{aligned}
H(U) &= -\sum_{i=1}^{N} p_U(u_i)\log_2 p_U(u_i) \ , &(3)\\
H(V) &= -\sum_{i=1}^{N} p_V(v_i)\log_2 p_V(v_i) \ , &(4)\\
H(U,V) &= -\sum_{i,j=1}^{N} p(u_i,v_j)\log_2 p(u_i,v_j) \ . &(5)
\end{aligned}
$$

The average mutual information $I(V;U)$ is given by

$$
I(V;U) = H(U) + H(V) - H(U,V) \ . \tag{6}
$$

$I(V;U)$ represents the amount of information that one can know about V given U. In this work, the time series $\{x(t)\}_{t=1}^{N-\Delta t}$ is substituted into U and its time-delayed counterpart $\{x(t+\Delta t)\}_{t=1}^{N-\Delta t}$ is substituted into V. Then, the mutual information is a function of the time lag Δt, denoted by $I(\Delta t)$, and represents the spontaneous loss of information caused by the time evolution of the system in terms of Δt. The optimal value for Δt in embedding is estimated as the first minimum of $I(\Delta t)$ or the Δt corresponding to $I(\Delta t)/e$.

3.2 Wayland test

The algorithm developed by Wayland et al. captures visible determinism underlying a time series in terms of the parallelness of neighboring trajectories reconstructed in the phase space [2]. The central idea of the algorithm is that visible determinism achieves smooth trajectories of time evolution in the phase space according to the definition of determinism that similar causes give rise to similar effects. Hence, for deterministic processes, neighboring trajectories will point in similar directions. In contrast, stochastic processes will give rise to large diversity in the directions of the trajectories. Thus, the degree of diversity depends on how visible determinism is in a time series. Practical procedures of this algorithm are as follows.

We randomly choose a vector $\mathbf{x}(t_0)$ and find its K nearest neighbors $\mathbf{x}(t_k)$, for instance, in the sense of Euclidean distance. Next, we map each vector to $\mathbf{x}(t_k+T)$ $(k=0,\cdots,K)$ with an appropriately chosen time interval T. The diversity of directions of neighboring trajectories can be measured in terms of the translation error:

$$E_{trans} = \frac{1}{K+1}\sum_{k=0}^{K}\frac{\| \mathbf{v}(t_k)-\bar{\mathbf{v}} \|}{\| \bar{\mathbf{v}} \|}\ , \tag{7}$$

$$\mathbf{v} = \frac{1}{K+1}\sum_{k=0}^{K}\mathbf{v}(t_k)\ , \tag{8}$$

$$\mathbf{v}(t_k) = \mathbf{x}(t_k+T)-\mathbf{x}(t_k)\ . \tag{9}$$

The difference vectors $\mathbf{v}(t_k)$ approximate tangential vectors of the trajectories. The more visible determinism is, the smaller E_{trans} will be. To reduce stochastic error in estimating E_{trans}, we use the following procedures. We seek the medians of E_{trans} for S sets of M randomly chosen $\mathbf{x}(t_0)$ and estimate the mean over the S medians. The mean is taken as a representative of E_{trans}.

The previous numerical work [9] suggested the following tentative criteria to infer the dynamical nature of a time series from an estimate of E_{trans} as a function of embedding dimension. E_{trans} below 0.1 indicates the visible determinism of the underlying dynamics regardless of how complex a time series appears. For E_{trans} exceeding 0.5, the time series represents stochastic processes. In particular, white noise (temporally uncorrelated random noise) generates $E_{trans} \approx 1$ independently of the embedding dimension. In case E_{trans} lies between 0.1 and 0.5, the time series may represent either temporally correlated stochastic processes or deterministic processes contaminated with observational random noise. In this work, we utilize these criteria to interpret estimates of E_{trans}.

3.3 Surrogate method

The surrogate method is an ingenious way of evaluating the statistical significance of an estimate of a statistic made from a single time series [3, 4]. Given a time series as a single set of realizations of certain dynamics, we make a null hypothesis about the dynamics. We then synthesize surrogate time series sharing the statistical properties relevant to the null hypothesis with the original time series. This procedure may be viewed as if we reobserved the dynamical system under the restriction of the null hypothesis.

In this work, we made the null hypothesis that daily stock return is subject to an independent identically distributed (*i. i. d.*) random process and hence displays no particular weekly pattern. Under the null hypothesis, forty surrogate time series were synthesized by random-shuffling the original stock prices within each week (from Monday to Friday, totally 764 weeks) using *i. i. d.* random numbers followed by the interpolation of the data and the estimation of stock returns or by random-shuffling the original stock returns within each week. Each surrogate time series has the same statistical moments as those of the original time series.

We next ranked the stock returns within each week and gave scores 1 to 5 to the stock return of each day of the week for both the original and the surrogate data. To test the null hypothesis, a statistical t-test was applied to the mean rank score on each day of the week. Let us denote the mean rank scores for the original time series as s_0, those for the surrogate time series as s, and their mean and variance as $\bar{s}$ and σ_s^2, respectively. Then the two-sided t-test statistic is given as

$$t = \frac{| s_0 - \bar{s} |}{\sigma_s} . \tag{10}$$

In the case of forty surrogate time series, the critical value of t is 2.02 at a two-sided level of 5% reliability. If $t > 2.02$, the null hypothesis is rejected, which means that the original time series is unlikely to be realized by the dynamics of the null hypothesis.

4 Results

We estimated the statistical moments of the first (mean), second (variance), third and fourth order for the stock returns ($N = 5347$). Similar estimates were made for the stock returns on each day of the week ($N = 764$). Results are summarized in Table 1. The stock return data are expressed as annual rate in percentage, i.e., $r(t) \times 360 \times 100$.

Mean stock return takes a negative large value with the smallest variance on Monday, and it is the lowest with a large variance on Friday. It turns out that this statistical tendency is not due to interpolation, from time series analysis based on the Wayland test.

Table 1 Statistical moments of annual stock returns (in %)

Days of the week	Mean	Variance	3rd order	4th order
All days	−9.363	392.368	0.219	6.412
Monday	−16.914	198.951	−0.114	2.480
Tuesday	12.546	465.867	0.813	8.653
Wednesday	−5.656	508.713	−0.043	2.002
Thursday	5.530	491.773	−0.066	2.156
Friday	−30.634	490.552	0.124	2.978

We next estimated the mutual information as a function of time lag for the stock returns. The probability density function and the joint probability density function were approximately estimated from a one-dimensional histogram with 32 partitions and a two-dimensional histogram with 32×32 partitions, respectively. Estimates are shown in Fig. 3. The mutual information takes the first minimum at $\Delta t = 4$, which corresponds to the optimal time lag in embedding. Small peaks appearing every 7 days of time lag are artifacts due to the interpolation of the data. No such artifacts appear in autocorrelation function, as shown in Fig. 4.

The Wayland test was applied to the stock returns with $\Delta t = 4$, 7, $K = 4$, $T = 1$, 5, 6, 7, 8, $M = 301$ and $S = 20$. $\Delta t = 4$ is the optimal time lag in terms of the mutual information and $\Delta t = 7$ corresponds to one week shift. Results are shown in Figs. 5 and 6. In the case of $\Delta t = 4$, determinism in stock return is as invisible as that of Gaussian white noise at $T \leq 7$. At $T = 8$, determinism in stock return becomes as visible as those in temporally correlated stochastic processes, unlike those at $T = 1$, 5, 6, 7, and takes the minimum at $D = 6$. Such a tendency is also observed for $\Delta t = 7$, $T = 7$ and $D = 4$, as shown in Fig. 6.

We next applied the Wayland test to the stock returns on each day of the week. Each time series of stock return consists of 764 data points. Results under $K = 4$, $T = 5$, $M = 51$ and $S = 10$ are presented in Fig. 7. Standard deviation in E_{trans} was estimated to be $0.04 - 0.14$ at $D \geq 8$. We chosen $T = 5$ because of the smallest variance in E_{trans} with respect to the time translation T. The results in Fig. 7 should be compared with those at $T = 5$ in Fig. 6. These results tell us that determinism is the most invisible on Monday. As discussed in the next section, this tendency is not due to the interpolation of data.

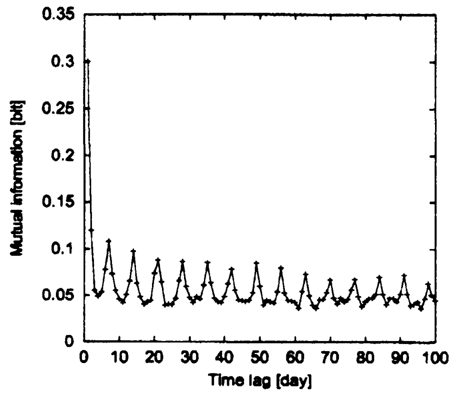

Figure 3: Mutual information of NIKKEI return as a function of time lag.

Estimates of the two-sided t-test statistic for the rank scores of stock return on each day of the week are summarized in Table 2. The critical t-value at 5 % reliability is 2.02. The estimated t-values exceed 2.02 on Monday and Friday. Hence, the null hypothesis can be rejected. The weekly pattern in stock return variation is unlikely to be ascribed to an *i. i. d.* stochastic process.

Table 2 Two-sided t-test statistics on the rank scores for the original and forty surrogate data. Upper row: Estimates for the stock returns calculated from randomly shuffled stock prices within each week. Lower row: estimates for randomly shuffled stock returns within each week. The critical t-value for rejecting the null hypothesis is 2.02.

Shuffling	Mon	Tue	Wed	Thu	Fri
Price	2.04	1.03	0.14	0.37	2.46
Return	1.36	0.70	0.16	0.33	2.44

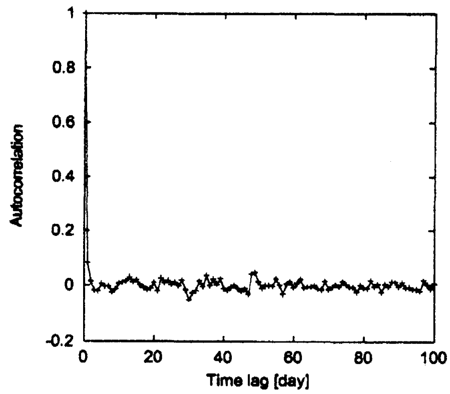

Figure 4: Autocorrelation function of NIKKEI return.

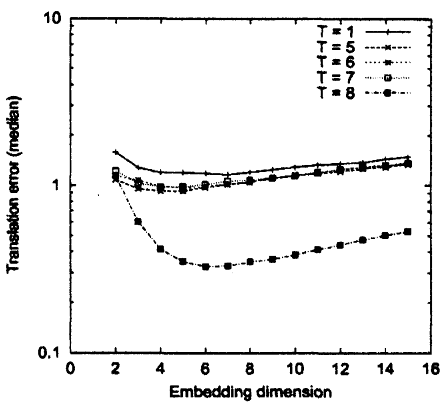

Figure 5: Wayland test for NIKKEI return. $T = 1, 5, 6, 7, 8$ and $\Delta t = 4$.

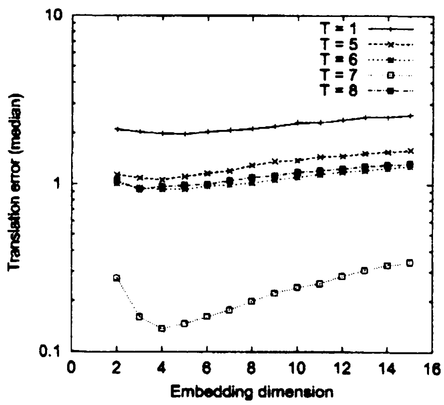

Figure 6: Wayland test for NIKKEI return. $T = 1, 5, 6, 7, 8$ and $\Delta t = 7$.

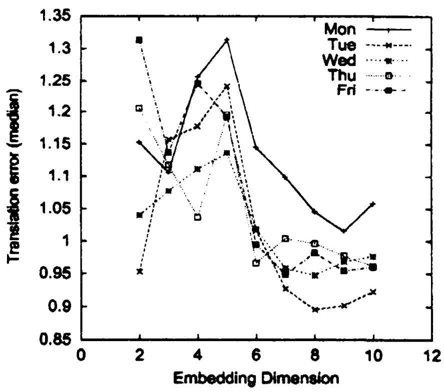

Figure 7: Wayland test for NIKKEI returns on each day of the week ($T = 5$).

5 Discussion

The Wayland test indicates that NIKKEI stock return displays the most complex dynamical behavior on Monday. This is not due to linear interpolation applied to the data, because linear interpolation generates smooth sequences of data points and should lower E_{trans}. These results suggest that the significantly low stock return with a small variance on Monday is likely to represent a particular weekly pattern in stock return variation.

This interpretation is supported by the surrogate analysis. In fact, the null hypothesis that the stock return variation represents an *i. i. d.* stochastic process has been rejected by statistical t-test on the rank order of the stock returns on Monday and Friday. The surrogate analysis tells us that there may be a Friday effect on stock return, as well.

In conclusion, the present analysis suggests that a particular weekly pattern in stock return variation, i.e., the Monday effect as well as the Friday effect, is likely to exist in the Japanese stock market. The economic origin of such an anomaly is not unclear. However, we conjecture one possible mechanism such that the dynamics governing stock return variation might be altered or be different on Saturday and Sunday because of missing stock index as the benchmark for trading. This would be investigated in our future work.

References

[1] W. Tong, "International Evidence on Weekend Anomalies", Journal of Financial Research, **23**, 495 522, 2000.

[2] R. Wayland, D. Bromley, D. Pickett and A. Passamante, "Recognizing Determinism in a Time Series", Phys. Rev. Lett., **70**, 580–582, 1993.

[3] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian and J. D. Farmer, "Testing for nonlinearity in time series: The method of surrogate data", Physica D, **58**, 77–94, 1992.

[4] J. Theiler and D. Prichard, "Constrained-realization Monte-Carlo method for hypothesis testing", Physica D, **94**, 221–235, 1996.

[5] F. Takens, "Detecting strange attractors in turbulence", *Lecture Notes in Mathematics* (Springer–Verlag), **898**, 366–381, 1981.

[6] T. Sauer, J. A. Yorke and M. Casdagli, "Embedology", J. Stat. Phys., **65**, 579–616, 1991.

[7] A. M. Fraser and H. L. Swinney, "Independent Coordinates for Strange Attractors from Mutual Information", Phys. Rev. A, **33**, 1134–1140, 1986.

[8] A. M. Fraser, "Information and Entropy in Strange Attractors", IEEE Trans. Information Theory, **35**, 245–262, 1989.

[9] T. Miyano, "Time Series Analysis of Complex Dynamical Behavior Contaminated with Observational Noise", Int. J. Bifurcation and Chaos, **6**, 2031–2045, 1996.