

Progress of Theoretical Physics Supplements

PTP is predecessor journal of Progress of Theoretical and Experimental Physics (PTEP)

Econophysics 2011:

The Hitchhiker's Guide to the Economy

Proceedings of the YITP Workshop on Econophysics

EDITED BY

Hideaki AOUAMA, Yoshi FUJIWARA, Hiroshi IYETOMI, and Aki-Hiro SATO

Prog. Theor. Phys. Suppl. **194** (2012).

Proceedings of the YITP Workshop on Econophysics

Kyoto, Japan

July 15–16, 2011

Preface

Econophysics is a relatively new discipline that applies concepts and methodologies that originated in physics to the problem of obtaining a better understanding of a wide variety of complex phenomena of a socio-econo-techno nature, including economic, sociological and financial behavior. The growth of this field in recent years has been extremely rapid. It has been driven by both the large quantities and the high quality of socio-econo-techno data that are now available due to the spread of information and communication technology (ICT) in our society.

Econophysics was developed by researchers in several fields. In the 1960s, Mandelbrot began to investigate price movements of financial markets and established the important concept of fractals through inspiration provided by the observation of their fluctuations. In the 1980s, physicists began to investigate statistical properties of human behavior on the basis of socio-economic data. However, the availability of data was limited, and their resolution and coverage were bounded. In spite of these difficulties, the topics investigated by physicists eventually expanded to cover various aspects of financial markets and business activities. They further attempted to develop models to explain collective behavior observed in socio-econo-techno systems in terms of physical concepts, such as scaling, clustering, correlations, and more complicated concepts.

At that time, the conceptual framework proposed by researchers could not be falsified through empirical investigations. However, with the commercialization of the Internet in 1995, the last restrictions on the use of the Internet for commercial purposes were removed. Since the mid-1990s, the Internet has had an unprecedented impact on our society. Today, ICT allows us to collect, store, and analyze huge quantities of data on economic activities. At the end of the 20th century, the availability of data on socio-econo-techno systems allowed the possibility for the development of methods to falsify theories by use of empirical studies.

The Yukawa Institute for Theoretical Physics (YITP), Kyoto University, has



Participants at the YITP workshop held during the period of July 15–16, 2011.

led this emergent interdisciplinary field with a series of workshops, starting in 2003. This institutional location could not be more appropriate. Professor Hideki Yukawa, the first Japanese citizen to be honored with a Nobel Prize, encouraged his students and colleagues to extend their interests into a wide variety of subjects, including elementary particle theory, cosmology, biophysics, and beyond.

The Econophysics Workshop held in 2011 was the fifth workshop. This workshop had 90 participants and was held on July 15 and 16, 2011. The program consisted of 4 invited talks, 18 contributed talks, and 10 contributed posters. In addition, there was a panel session with members of the Public Interest Capitalism Research Division of the Alliance Forum Foundation on the second day of the workshop.

To promote the Japanese econophysics community during the next decade and communication with foreign researchers, we publish this special issue on econophysics containing the papers of authors who participated in the workshop. The heart of econophysics has been transmitted from the previous generation to the present generation, and it will be communicated from the present generation to the next generation. We hope that this special issue will serve as a guide to researchers of the next generation.

In editing this volume, we have received invaluable advice and assistance from our colleagues in the organizing committee, and we wish to take this opportunity to express our especial thanks to Yuichi Ikeda, Jun-ichi Maskawa, Wataru Souma, Hiromi Yokoyama, and Hiroshi Yoshikawa. Finally, without the administrative and financial support of the Yukawa Institute, neither the workshop nor this publication would have been possible. We also received financial support from the Kyoto University Global COE program “Informatics Education and Research Center for Knowledge-circulating Society” and the Public Interest Capitalism Research Division of the Alliance Forum. On behalf of all those involved, we would like to express our gratitude to the Yukawa Institute, the Global COE program, and the Alliance Forum for their invaluable assistance.

April 2012

The Editorial Board:

Hideaki Aoyama, Kyoto University
Yoshi Fujiwara, University of Hyogo
Hiroshi Iyetomi, The University of Tokyo
Aki-Hiro Sato, Kyoto University

List of Presenters in the YITP Workshop (July 15–16, 2011)*)

Invited Talks

1. Toshiyuki Masuda (Kyoto Shinkin Bank, Chairman of the Board of Directors)
2. Toshinori Tanabe (CRD Association, Director)
3. Kiyoshi Kabaya (Tokyo Shoko Research, LTD., Managing Director)
4. Sigehiro Kato (Recruit Co., LTD., Executive Producer)

Panel Discussion

1. Hiroshi Iyetomi (Department of Physics, Niigata University)
2. George Hara (Alliance Forum Foundation, Chairman of the Board of Directors)
3. Hiroshi Yoshikawa (Graduate School of Economics, The University of Tokyo)
4. Hideaki Aoyama (Graduate School of Science, Kyoto University)
5. Jun-ichi Maskawa (Department of Economics, Seijo University)

Contributed Talks

1. Jun-ichi Maskawa (Department of Economics, Seijo University)
2. Kouji Kuroda (Graduate School of Integrated Basic Sciences, Nihon University)
3. Ryoji Minami (Graduate School of Information Science and Technology, The University of Tokyo)
4. Takatoshi Tasaki
5. Yukie Sano (College of Science and Technology, Nihon University)
6. Xin Yang (Graduate School of Engineering, Tottori University)
7. Zeyu Zheng (Department of Environmental Information, Tokyo University of Information Sciences)
8. Kazuko Yamasaki (Department of Environmental Information, Tokyo University of Information Sciences)
9. Akira Ishii (Department of Applied Mathematics and Physics, Tottori University)
10. Makoto Mizuno (School of Commerce, Meiji University)
11. Aki-Hiro Sato (Department of Applied Mathematics and Physics, Kyoto University)
12. Hiroshi Iyetomi (Department of Physics, Niigata University)
13. Atsushi Ishikawa (Department of Informatics and Business, Kanazawa Gakuin University)
14. Shouji Fujimoto (Department of Informatics and Business, Kanazawa Gakuin University)
15. Takayuki Mizuno (Faculty of Engineering, Information and Systems, University

*) The numbers are assigned in the order of presentations.

of Tsukuba)

16. Takashi Iino (Department of Physics, Niigata University)
17. Yoshi Fujiwara (Graduate School of Simulation Studies, University of Hyogo)
18. Yuichi Ikeda (Institute of Industrial Science, The University of Tokyo)

Contributed Posters

1. Lisa Uechi (Bioinformatics Center, Institute for Chemical Research, Kyoto University)
2. Christopher A. Zapart (Institute of Statistical Mathematics)
3. Hidehiro Matsumoto (Department of Applied Mathematics and Physics, Tottori University)
4. Minoru Noda (Department of Applied Mathematics and Physics, Kyoto University)
5. Yuta Arai (Graduate School of Sciences, Niigata University)
6. Takeo Yoshikawa (Graduate School of Sciences, Niigata University)
7. Hisashi Arakaki (Graduate School of Engineering, Tottori University)
8. Tetsuya Takaiishi (Hiroshima University of Economics)
9. Wataru Souma (College of Science and Technology, Nihon University)
10. Kenji Tsumura (Graduate School of Science and Technology, Nihon University)

Long Memory in Trade Signs and Short Memory in Stock Prices

Koji KURODA,¹ Jun-ichi MASKAWA² and Joshin MURAI³

¹*Graduate School of Integrated Basic Sciences, Nihon University,
Tokyo 156-8550, Japan*

²*Department of Economics, Seijo University, Tokyo 157-8511, Japan*

³*Graduate School of Humanities and Social Sciences, Okayama University,
Okayama 700-8530, Japan*

We consider a mathematical model for stock markets and derive a signed volume process having a long memory property and a stock price process having a short memory property. Using the method of cluster expansion developed in the study of phase transitions, we describe our results about scale limits of the processes by using Brownian motion and fractional Brownian motion, which is known as a stochastic process having a long memory property.

§1. Introduction

The dynamics of stock prices has been one of the main research subjects of econophysics.^{3),17)} First of all, we give a brief description of price formation in stock markets. The continuous double auction method is adopted for price formation in all the main world stock markets including New York, London and Tokyo. Orders to buy or sell stocks are classified into two categories: market orders and limit orders. Regarding limit orders we can specify both trade price and trade volume, while we can only specify the trade volume for market orders and the trade price is determined as the best possible price. A market order is transacted immediately; however, a limit order often fails to be transacted immediately and is stored in a queue called a limit order book.

Here, and throughout this paper, we describe a market price as the mid price of the best ask price (the lowest sell price) and the best bid price (the highest buy price) in a limit order book. When a buy (sell) market order comes and clears all volumes at the best ask (bid) price, the best ask (bid) price moves to the second best price and causes a rise (fall) in the mid price; this is called the price impact of transaction. If the volume of a buy (sell) market order is less than the volume at the best ask (bid) price, then the mid price does not change. The mid price also moves by the limit orders placed in a spread that is the clearance between ask and bid, and also by the cancellation of all the best limit orders at the time.

Empirical studies^{1),11),15)} on datasets recording all the transactions in stock markets have revealed the long memory of supply and demand fluctuations, which appear as the auto-correlation in the sign of market orders (i.e., $s = +1$ for buys and $s = -1$ for sells). In fact, the auto-correlation $\rho(\tau)$ between trade signs s_u at time u and $s_{u+\tau}$ at time $u + \tau$ decays very slowly as an inverse power law of time lag τ ,

$$\rho(\tau) \sim \frac{c_0}{\tau^\alpha} \quad (t \rightarrow \infty) \quad (1.1)$$

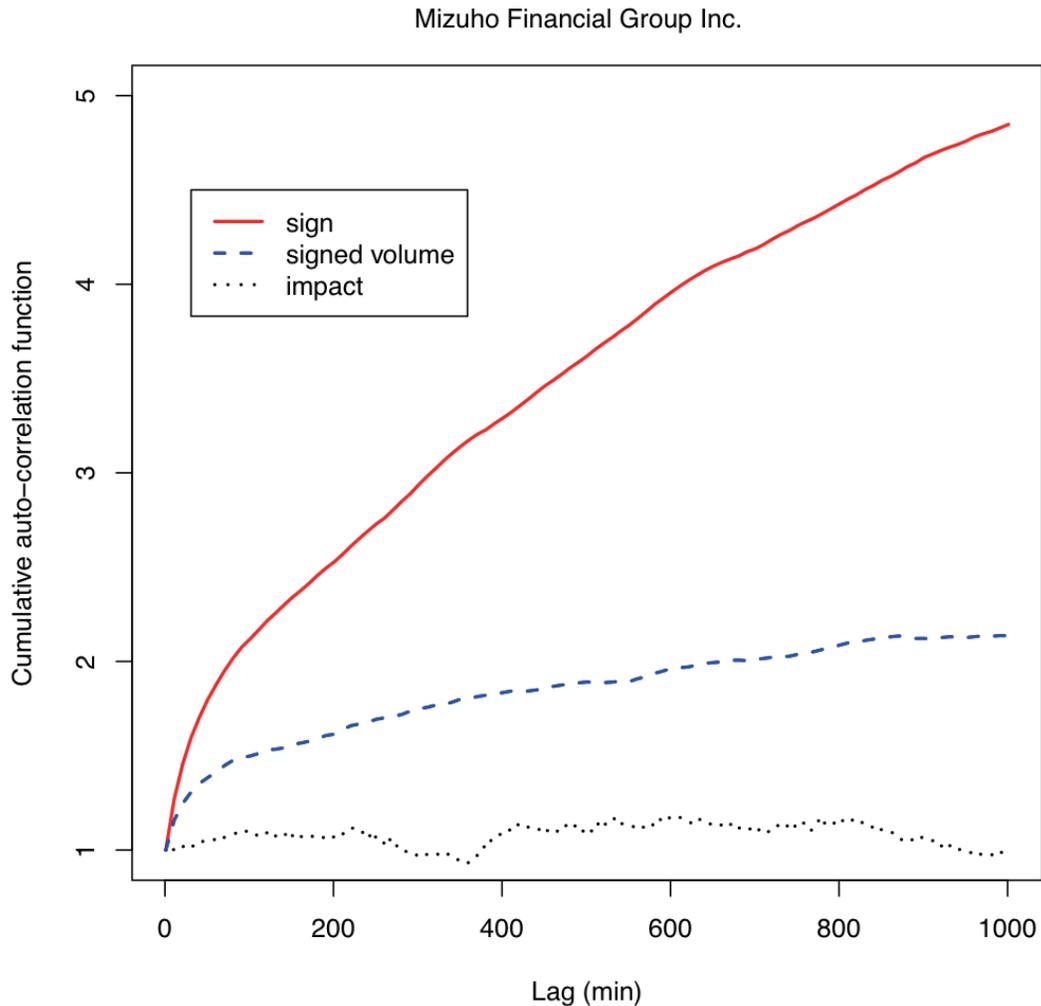


Fig. 1. Cumulative auto-correlation functions of sequences of trade signs, signed transaction volumes and price impacts per minute. Trade signs and signed volumes are defined by the sign of ($\#$ of transactions executed at ask - $\#$ of transactions executed at bid) per minute and (transaction volumes traded at ask - transaction volumes traded at bid) per minute respectively. The midday transactions of liquid securities are analysed. Here we show results for Mizuho Financial Group Inc. for the period from January to December in 2008.

with $0 < \alpha < 1$ and some constant $c_0 > 0$. The long memory of trade signs suggests that price impact should also have long memory, and thus it should be predictable. However, we share the common understanding that a price change of the present time has no linear correlation with a past one and that price is not predictable.^{5),16)} Typical examples of cumulative auto-correlation functions of sequences of trade signs, signed transaction volumes and price impacts are shown in Fig. 1. The cumulative auto-correlation functions of trade signs and signed volumes increase without bounds, showing the long memory nature of those stochastic processes, while price impact has no significant auto-correlation.

How can we resolve this paradox and satisfy both empirical facts? Two different

ideas to resolve the paradox have been proposed. Bouchaud et al.^(1),2) proposed a model where price is the result of the impact of all past trade, mediated by a time varying propagator that describes the response of the market to a past single trade. They argue that the mean-reverting limit orders provided after each market order counter the impact of past trade, thus the price impact of a single trade shows a power-law decays in time and has no auto-correlation with long time scales. In contrast, Lillo and Farmer^(6),15) argue that the price impact of each trade does not decay in time; instead it strongly depends on liquidity at the time when the market order comes, where liquidity is defined as the price response to a transaction of a given size. In Lillo and Farmer's opinion, liquidity fluctuation that is not taken into consideration in the work of Bouchaud et al. correlates with order flows and suppresses the impact of market orders; thus, price itself has no auto-correlation.

In this article, we present a mathematical model describing stock market trades. We construct signed volume process X_u , which describes the time evolution of the imbalance between buy and sell market orders, and stock price process Y_u under some assumptions, and derive a scale limit of the coupled processes of X_u and Y_u . Amongst explanations of the long memory in trade signs, one dominant explanation is that long memory results from traders' strategy to split their orders into small pieces. When traders want to place large orders, they try to keep their demand as secret as possible by splitting their orders into smaller pieces. For the purpose of compatibility with the long memory of sign process and the randomness of price process, we postulate a 'carefulness' of traders when they place a portion of a large market order. We assume that they place market orders very carefully and the portions of large orders are small enough not to move the market price.

We set up a model of stock trades in a discrete time set $\Lambda_n = \{1, \dots, n\}$ taking this explanation into account, and derive a signed volume process W_u and a stock price process H_u . In this model we assume that a trader can divide their order into at most m_0 pieces. Then, we consider a scale process $(X_t^{(n)}, Y_t^{(n)})$ given by

$$X_t^{(n)} = \frac{X_{[nt]}}{c(n)} \quad \text{and} \quad Y_t^{(n)} = \frac{Y_{[nt]}}{c(n)}, \quad (t \in [0, 1])$$

where $c(n)$ is a scale function which will be given explicitly in a sequel.

Readers are reminded that the Brownian motion B_t is obtained as a scale limit of a random walk,

$$B_t = \lim_{n \rightarrow \infty} \frac{S_{[nt]}}{\sqrt{n}},$$

where $\{S_m\}$ is a random walk and is described as the sum of m independent identically distributed random variables.

In a similar way to Brownian motion, we prove that $(X_t^{(n)}, Y_t^{(n)})$ converges to a process (X_t, Y_t) given by

$$X_t = c_1 B_t + \mu_B \sum_{\ell=0}^{m_0-2} a_\ell B_t^{H_\ell},$$

$$Y_t = c_2 \tilde{B}_t + \gamma_B \sum_{\ell=0}^{m_0-2} a_\ell B_t^{H_\ell},$$

where $\{B_t, \tilde{B}_t\}$ are correlated Brownian motions, $\{B_t^{H_\ell}; \ell = 0, \dots, m_0 - 2\}$ are independent fractional Brownian motions with Hurst indices $\{H_\ell\}$, and (B_t, \tilde{B}_t) and $(B_t^{H_0}, \dots, B_t^{H_{m_0-2}})$ are independent. Precise value of H_ℓ will be given in Theorem 3.1.

Let us remind that correlated Brownian motions are given by linear combinations of independent Brownian motions (see Theorem 3.1 for detail).

A fractional Brownian motion B_t^H with Hurst index $H \in (\frac{1}{2}, 1)$ is known as a stochastic process whose increments have a long memory property. It is defined as a centered Gaussian process with a covariance given by

$$\text{Cov}(B_t^H, B_s^H) = \frac{1}{2} (t^{2H} + s^{2H} - |t - s|^{2H}).$$

A covariance between increments ΔB_t^H and ΔB_s^H is given by

$$\text{Cov}(\Delta B_t^H, \Delta B_s^H) \sim H(2H - 1) \cdot \frac{1}{|t - s|^{2-2H}} \Delta t \Delta s,$$

where $\Delta B_t^H = B_{t+\Delta t} - B_t$ and $\Delta B_s^H = B_{s+\Delta s} - B_s$.

From the assumption of the model which is compatible with empirical data, we show that γ_B is small enough. This implies that the long memory part of the stock price process is also small enough.

§2. Description of the model

Let $A_n = \{1, 2, \dots, n\}$ be a period of time to observe trades by traders. We consider one type of security and a model for trades by market orders and limit orders placed in a spread.

It is commonly known that a stock price changes if all volumes at the best price are cleared by a market order or a limit order is placed in the spread. We only consider a sequence of transactions by market orders and limit orders placed in a spread.

To describe transactions by one trader who can split his order into at most m_0 pieces, we introduce the notion of *polymer* \mathbf{p} given by

$$\mathbf{p} = (s, m, \mathbf{t}, \mathbf{v}, \boldsymbol{\ell}) \in \{+1, -1\} \times \{1, \dots, m_0\} \times \mathcal{T}_m \times V^m \times \{0, 1\}^m, \quad (2.1)$$

where $m_0 \geq 2$ is a positive integer,

$$\mathcal{T}_m = \{\mathbf{t} = (t_1, \dots, t_m) \in \mathbf{Z}^m; 0 < t_{i+1} - t_i \leq n \quad (i = 1, \dots, m-1)\}, \quad (2.2)$$

and V is a finite set of positive numbers. When polymer \mathbf{p} is given as above, $s \in \{+1, -1\}$ is a *trade sign*, $\mathbf{t} = (t_1, \dots, t_m) \in \mathcal{T}_m$ is a set of times the trader places orders, and $\mathbf{v} = (v_1, \dots, v_m) \in V^m$ is a set of *trade volumes* in \mathbf{t} . We put $b(\mathbf{p}) = \{t_1, \dots, t_m\}$ and $|\mathbf{p}| = m$.

We assume that all orders are market orders if a trader splits their order into more than two pieces. To describe a price change by an order at time t we define $\ell_t \in \{0, 1\}$ by

$$\ell_t = \begin{cases} 0 & \text{if it is a market order and the trade volume at } t \text{ is less than a volume} \\ & \text{at the best price,} \\ 1 & \text{otherwise.} \end{cases}$$

If it is a limit order placed in a spread or a market order with trade volume equal or more than volume at the best price, then $\ell_t = 1$.

When $s_t = +1$ (-1) and $\ell_t = 1$, a stock price rises (falls) at time t .

We describe all transactions observed in Λ_n by a configuration of mutually disjoint polymers. A *configuration space* Ω_n is defined by

$$\Omega_n = \bigcup_{k=0}^n \{ \omega = \{ \mathbf{p}_1, \dots, \mathbf{p}_k \} ; b(\mathbf{p}_i) \cap \Lambda_n \neq \emptyset \ (i = 1, \dots, k) \\ b(\mathbf{p}_i) \cap b(\mathbf{p}_j) = \emptyset \ (1 \leq i < j \leq k) \}.$$

We define a polymer weight functional $\varphi(\mathbf{p})$ by

$$\varphi(\mathbf{p}) = \begin{cases} d_n \cdot P_A(\ell, v), & (|\mathbf{p}| = 1, \mathbf{p} = (s, t, v, \ell)) \\ d_n^m \cdot \frac{1}{(t_2 - t_1)^\alpha \cdots (t_m - t_{m-1})^\alpha} \prod_{k=1}^m P_B(\ell_k, v_k), & (|\mathbf{p}| \geq 2) \end{cases}$$

where $0 < \alpha < 1$, $d_n = \frac{c}{n^{1-\alpha}}$ ($c > 0$), and $P_A(\ell, v)$ and $P_B(\ell, v)$ are probability measures on $\{0, 1\} \times V$. We consider two types of traders: *type A traders* ($|\mathbf{p}| = 1$) who place orders at one time and do not split them; and *type B traders* ($|\mathbf{p}| \geq 2$) who split their orders into more than two pieces.

A probability measure P_n on Ω_n is defined by

$$P_n(\omega) = \frac{1}{Z_n} \prod_{\mathbf{p} \in \omega} \varphi(\mathbf{p})$$

for each $\omega \in \Omega_n$, where Z_n is a normalization constant.

Furthermore we define γ_A , η_A , μ_A and σ_A^2 by

$$\gamma_A = \sum_{v \in V} P_A(1, v), \quad \eta_A = \sum_{v \in V} v P_A(1, v), \\ \mu_A = \sum_{v \in V} \sum_{\ell \in \{0,1\}} v \cdot P_A(\ell, v), \quad \sigma_A^2 = \sum_{v \in V} \sum_{\ell \in \{0,1\}} v^2 \cdot P_A(\ell, v).$$

In the same way, we also define γ_B , η_B , μ_B and σ_B^2 by

$$\gamma_B = \sum_{v \in V} P_B(1, v), \quad \eta_B = \sum_{v \in V} v P_B(1, v), \\ \mu_B = \sum_{v \in V} \sum_{\ell \in \{0,1\}} v \cdot P_B(\ell, v), \quad \sigma_B^2 = \sum_{v \in V} \sum_{\ell \in \{0,1\}} v^2 \cdot P_B(\ell, v).$$

§3. Statement of results

For any polymer $\mathbf{p} = (s, m, t_1, \dots, t_m, v_1, \dots, v_m, \ell_1, \dots, \ell_m)$, we define $X_u(\mathbf{p})$ and $Y_u(\mathbf{p})$ by

$$X_t(\mathbf{p}) = \sum_{1 \leq k \in b(\mathbf{p}); k \leq t} s(\mathbf{p}) \cdot v_k \quad \text{and} \quad Y_t(\mathbf{p}) = \sum_{1 \leq k \in b(\mathbf{p}); k \leq t} s(\mathbf{p}) \cdot \ell_k.$$

Furthermore, for any configuration $\omega \in \Omega_n$, we define a signed volume process $X_t(\omega)$ and a stock price process $Y_t(\omega)$ by

$$X_t(\omega) = \sum_{\mathbf{p} \in \omega} X_t(\mathbf{p}) \quad \text{and} \quad Y_t(\omega) = \sum_{\mathbf{p} \in \omega} Y_t(\mathbf{p}).$$

Scale process

A coupled scale process $(X_t^{(n)}, Y_t^{(n)})$ of a stock price process and a signed volume process are defined by

$$X_t^{(n)}(\omega) = \frac{X_{[nt]}(\omega)}{c(n)} \quad \text{and} \quad Y_t^{(n)}(\omega) = \frac{Y_{[nt]}(\omega)}{c(n)}$$

for any $\omega \in \Omega_n$ and any $t \in [0, 1]$, where $c(n)$ is a scale function and $[x]$ is an integral part of $x \in \mathbb{R}$.

Theorem 3.1. *Assume that $\frac{m_0-2}{m_0-1} < \alpha$. Let $c(n) = n^{\frac{\alpha}{2}}$. Then the finite dimensional distribution of $(X_t^{(n)}, Y_t^{(n)})$ converges to the corresponding distribution of (X_t, Y_t) given by*

$$\begin{aligned} X_t &= c_1 B_t^1 + c_2 B_t^2 + \mu_B \sum_{k=0}^{m_0-2} a_k B_t^{H_k}, \\ Y_t &= \hat{c}_1 B_t^1 + \hat{c}_2 B_t^2 + \gamma_B \sum_{k=0}^{m_0-2} a_k B_t^{H_k}, \end{aligned}$$

where B_t^1 and B_t^2 are independent Brownian motions, $\{B_t^{H_k}; k = 0, \dots, m_0 - 2\}$ are independent fractional Brownian motions with Hurst indices $H_k = \frac{1}{2}\{(1 - \alpha)(k + 1) + 1\}$, and (B_t, \tilde{B}_t) and $(B_t^{H_0}, \dots, B_t^{H_{m_0-2}})$ are independent.

Here constants c_1, c_2, \hat{c}_1 , and \hat{c}_2 are determined by

$$\begin{cases} c_1^2 + c_2^2 = 2c(\sigma_A^2 + \sigma_B^2 \hat{E}), \\ \hat{c}_1^2 + \hat{c}_2^2 = 2c(\gamma_A + \gamma_B \hat{E}), \\ c_1 \hat{c}_1 + c_2 \hat{c}_2 = 2c(\eta_A + \eta_B \hat{E}), \end{cases}$$

a_k is given by

$$a_k = \sqrt{\frac{4c^{k+2} E_k \Gamma(1 - \alpha)^{k+1}}{\Gamma((1 - \alpha)(k + 1) + 2)}}$$

and E_k is given by

$$E_k = \sum_{u=0}^{m_0-(k+2)} (u+1) \left(\frac{c}{1-\alpha} \right)^u.$$

Remark. The Brownian motion parts of X_t and Y_t are considered to be correlated Brownian motions. The assumption of m_0 is necessary to ensure that Hurst indices are less than 1.

Since the number of price impacts is much smaller than that of transactions, γ_B is much smaller than μ_B . Hence, the signed volume process \bar{X}_t has a long memory property, but the long memory part of the stock price process \bar{Y}_t is small.

§4. Cluster expansion

The method of cluster expansion has been developed in the mathematical theory of phase transitions, and applied to various problems such as stochastic behavior of interface between two distinct phases.^{4),7),8)} In this section we summarize the method of cluster expansion which plays an important role for proof of our results.

Algebraic formalism

To apply an algebraic method of abstract cluster expansion we introduce a space \mathcal{A}_n of a finite number of polymers with multiplicity given by

$$\mathcal{A}_n = \left\{ A : \mathcal{P}_n \rightarrow \mathbf{N} \cup \{0\} ; |A| = \sum_{\mathbf{p} \in \mathcal{P}_n} A(\mathbf{p}) < \infty \right\},$$

where \mathcal{P}_n is a set of all polymers.

Each $A \in \mathcal{A}_n$ is regarded as a configuration of polymers permitting intersections and multiplicities. For $A_1, A_2 \in \mathcal{A}_n$, a sum $A_1 + A_2$ is defined by

$$(A_1 + A_2)(\mathbf{p}) = A_1(\mathbf{p}) + A_2(\mathbf{p}) \quad \text{for all } \mathbf{p} \in \mathcal{P}_n.$$

For each $A \in \mathcal{A}_n$, we put $A! = \prod_{\mathbf{p} \in \mathcal{P}_n} A(\mathbf{p})!$,

$$\text{supp}(A) = \{\mathbf{p} \in \mathcal{P}_n ; A(\mathbf{p}) \geq 1\},$$

$$\bar{A} = \bigcup_{\mathbf{p} \in \text{supp}(A)} \underbrace{\{\mathbf{p}, \dots, \mathbf{p}\}}_{A(\mathbf{p})},$$

$$b(A) = \bigcup_{\mathbf{p} \in \text{supp}(A)} b(\mathbf{p}).$$

We say $A \in \mathcal{A}_n$ is *admissible*, if $A! = 1$ and $b(\mathbf{p}_1) \cap b(\mathbf{p}_2) = \emptyset$ for all $\mathbf{p}_1, \mathbf{p}_2 \in \text{supp}(A)$ ($\mathbf{p}_1 \neq \mathbf{p}_2$). We note that if A is admissible, then \bar{A} is regarded as an element of Ω_n .

For $A \in \mathcal{A}_n$, let

$$\alpha(A) = \begin{cases} 1 & \text{if } A \text{ is admissible,} \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore we consider a functional space \mathfrak{L}_n on \mathcal{A}_n ,

$$\mathfrak{L}_n = \left\{ \psi : \mathcal{A}_n \rightarrow \mathbb{R} ; \sup_{|A|=k} |\psi(A)| < \infty \quad \text{for all } k \geq 1 \right\},$$

and define a product $\psi_1 * \psi_2$ of $\psi_1, \psi_2 \in \mathfrak{L}_n$ as follows,

$$\psi_1 * \psi_2(A) = \sum_{(A_1, A_2); A_1 + A_2 = A} \frac{A!}{A_1! A_2!} \psi_1(A_1) \psi_2(A_2),$$

where the sum runs over all ordered (A_1, A_2) such that $A_1 + A_2 = A$. With this product $*$, the functional space \mathfrak{L}_n becomes a commutative algebra with identity $\mathbf{1}$ defined by

$$\mathbf{1}(A) = \begin{cases} 1 & \text{if } A = \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathfrak{L}_{n,0} = \{\psi \in \mathfrak{L}_n; \psi(\emptyset) = 0\}$ and $\mathfrak{L}_{n,1} = \{\psi \in \mathfrak{L}_n; \psi(\emptyset) = 1\}$. We define an exponential mapping $\text{Exp} : \mathfrak{L}_{n,0} \rightarrow \mathfrak{L}_{n,1}$ by

$$\text{Exp } \psi(A) = \mathbf{1}(A) + \sum_{k=1}^{\infty} \frac{1}{k!} \underbrace{\psi * \cdots * \psi}_k(A).$$

As an inverse mapping of Exp a mapping $\text{Log} : \mathfrak{L}_{n,1} \rightarrow \mathfrak{L}_{n,0}$ is defined by

$$\text{Log } \psi(A) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \underbrace{\psi_0 * \cdots * \psi_0}_k(A),$$

where $\psi_0 = \psi - \mathbf{1}$.

The following Lemma is a fundamental Lemma of cluster expansion.

Lemma 4.1. ^{14), 18)} Let D be a domain in the complex plane \mathbb{C} . Let functionals $a : \mathcal{P}_n \rightarrow [0, \infty)$, $d : \mathcal{P}_n \rightarrow [0, \infty)$ and $\Phi : \mathcal{P}_n \times D \rightarrow \mathbb{C}$ be such that

$$\sum_{\substack{\mathbf{p} \in \mathcal{P}_n \\ b(\mathbf{p}) \cap b(\mathbf{q}) \neq \emptyset}} e^{a(\mathbf{p}) + d(\mathbf{p})} |\Phi(\mathbf{p}, \xi)| \leq a(\mathbf{q})$$

for each $\mathbf{q} \in \mathcal{P}_n$ and $\xi \in D$. Then

$$\sum_{A \in \mathcal{A}_n} \Phi(A, \xi) \alpha(A) = \exp \left[\sum_{A \in \mathcal{A}_n} \Phi(A, \xi) \frac{\alpha^T(A)}{A!} \right]$$

and

$$\sum_{\substack{A \in \mathcal{A}_n \\ b(A) \cap b(\mathbf{q}) \neq \emptyset}} \left| \Phi(A, \xi) \frac{\alpha^T(A)}{A!} \right| e^{d(A)} \leq a(\mathbf{q}) \quad (4.1)$$

for any $\mathbf{q} \in \mathcal{P}_n$ and $\xi \in D$, where $\alpha^T(A) = \text{Log } \alpha(A)$, $\Phi(A, \xi) = \prod_{\mathbf{p} \in \mathcal{P}_n} \Phi(\mathbf{p}, \xi)^{A(\mathbf{p})}$ and $d(A) = \sum_{\mathbf{p} \in \mathcal{P}_n} d(\mathbf{p})A(\mathbf{p})$.

We note that if $\Phi(\mathbf{p}, \xi)$ is analytic in $\xi \in D$, then

$$\sum_{A \in \mathcal{A}_n} \Phi(A, \xi) \frac{\alpha^T(A)}{A!}$$

is also analytic in $\xi \in D$.

Remark that $\Phi(A, \xi)$ is multiplicative with respect to A , that is,

$$\Phi(A_1 + A_2, \xi) = \Phi(A_1, \xi) \cdot \Phi(A_2, \xi)$$

for all $A_1, A_2 \in \mathcal{A}_n$.

When there exists a chain of polymers $\mathbf{p}_{i_0}, \dots, \mathbf{p}_{i_k} \in \text{supp } A$ satisfying

$$\mathbf{p}_{i_0} = \mathbf{p}_i, \mathbf{p}_{i_k} = \mathbf{p}_j, b(\mathbf{p}_{i_q}) \cap b(\mathbf{p}_{i_{q+1}}) \neq \emptyset$$

for any $\mathbf{p}_1, \mathbf{p}_j \in \text{supp } A = \{\mathbf{p}_1, \dots, \mathbf{p}_m\}$, we say A is a *cluster*.

We have a property of a cluster that $\alpha^T(A) = 0$ unless A is a cluster. This property will play an important role in a sequel.

§5. Proof of Theorem 3.1

Let us denote the characteristic function of $(X_{t_1}^{(n)}, \dots, X_{t_r}^{(n)}, Y_{t_1}^{(n)}, \dots, Y_{t_r}^{(n)})$ by

$$\phi_n(z_1, \dots, z_r, \zeta_1, \dots, \zeta_r) = E \left[\exp \left\{ \frac{i}{c(n)} \sum_{k=1}^r (z_k X_{[nt_k]} + \zeta_k H Y_{[nt_k]}) \right\} \right]$$

for fixed $0 < t_1 < \dots < t_r < 1$.

Introducing functionals $Z^{(n)}(\mathbf{p})$ and $Z^{(n)}(A)$ defined by

$$Z^{(n)}(\mathbf{p}) = \frac{1}{c(n)} \sum_{k=1}^r (z_k X_{[nt_k]}(\mathbf{p}) + \zeta_k Y_{[nt_k]}(\mathbf{p}))$$

and

$$Z^{(n)}(A) = \sum_{\mathbf{p} \in \mathcal{P}_n} Z^{(n)}(\mathbf{p}) A(\mathbf{p}),$$

we have

$$\phi_n(z_1, \dots, z_r, \zeta_1, \dots, \zeta_r) = \frac{\sum_{A \in \mathcal{A}_n} e^{iZ^{(n)}(A)} \varphi(A) \chi_n(A) \alpha(A)}{\sum_{A \in \mathcal{A}_n} \varphi(A) \chi_n(A) \alpha(A)},$$

where

$$\chi_n(\mathbf{p}) = \begin{cases} 1 & \text{if } b(\mathbf{p}) \cap \Lambda_n \neq \emptyset, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad \chi_n(A) = \prod_{\mathbf{p} \in \text{supp}(A)} \chi_n(\mathbf{p}).$$

First, we obtain the following inequality,

$$\left| Z^{(n)}(\mathbf{p}) \right| \leq \frac{C_1}{c(n)}, \quad (5.1)$$

where $C_1 = m_0 \sum_{k=1}^r \{\max(V) |z_k| + |\zeta_k|\}$, and $\max(V)$ is the maximal element of the finite set V .

We define a domain in the complex plane by

$$D = \{\xi \in \mathbb{C}; |\operatorname{Re} \xi| < 2h(n)\},$$

where $h(n) = \frac{c(n)}{2C_1}$, and $\operatorname{Re} \xi$ is the real part of a complex number ξ . For any $\xi \in D$, we have

$$\left| e^{\xi Z^{(n)}(\mathbf{p})} \right| \leq e. \quad (5.2)$$

To apply Lemma 4.1 to our model, we prepare the following inequality.

Lemma 5.1. *For any n , we have*

$$\sum_{\substack{\mathbf{p} \in \mathcal{P}_n \\ 1 \in b(\mathbf{p})}} \varphi(\mathbf{p}) \leq C_2 d_n,$$

where

$$C_2 = 2 \left\{ 1 + \sum_{m=2}^{m_0} m \left(\frac{c}{1-\alpha} \right)^{m-1} \right\}.$$

Proof. Since $0 < \alpha < 1$, it is easily seen that

$$\sum_{\tau=1}^n \tau^{-\alpha} \leq \int_0^n x^{-\alpha} dx = \frac{n^{1-\alpha}}{1-\alpha}.$$

Using this inequality, we have

$$\begin{aligned} \sum_{\substack{\mathbf{p} \in \mathcal{P}_n \\ 1 \in b(\mathbf{p})}} \varphi(\mathbf{p}) &= 2d_n \left\{ 1 + \sum_{m=2}^{m_0} m \cdot d_n^{m-1} \left(\sum_{\tau=1}^n \frac{1}{\tau^\alpha} \right)^{m-1} \right\} \\ &\leq 2d_n \left\{ 1 + \sum_{m=2}^{m_0} m \left(\frac{c}{1-\alpha} \right)^{m-1} \right\}. \end{aligned}$$

□

Let us take any $\varepsilon \in (0, 1)$ and fix it. For any $\mathbf{p} \in \mathcal{P}_n$, $A \in \mathcal{A}_n$ and $\xi \in D$, we consider functionals

$$\begin{aligned}\Phi(\mathbf{p}, \xi) &= e^{\xi Z^{(n)}(\mathbf{p})} \chi_n(\mathbf{p}) \varphi(\mathbf{p}), & \Phi(A, \xi) &= \sum_{\mathbf{p} \in \mathcal{P}_n} \Phi(\mathbf{p}, \xi) A(\mathbf{p}), \\ a(\mathbf{p}) &= a_0 e^{m_0+1} d_n^\varepsilon |\mathbf{p}|, & d(\mathbf{p}) &= -(1-\varepsilon) \log d_n.\end{aligned}$$

From now on we assume that n is large enough to satisfy

$$e^{a_0 e^{m_0+1} d_n^\varepsilon} \leq e. \quad (5.3)$$

Under the above assumption on n , we have

$$\begin{aligned}\sum_{\substack{\mathbf{p} \in \mathcal{P}_n \\ b(\mathbf{p}) \cap b(\mathbf{q}) \neq \emptyset}} e^{a(\mathbf{p})+d(\mathbf{p})} |\Phi(\mathbf{p}, \xi)| &\leq e \sum_{\substack{\mathbf{p} \in \mathcal{P}_n \\ b(\mathbf{p}) \cap b(\mathbf{q}) \neq \emptyset}} e^{a_0 e^{m_0+1} d_n^\varepsilon |\mathbf{p}|} \cdot e^{-(1-\varepsilon) \log d_n} \varphi(\mathbf{p}) \\ &\leq e d_n^{-(1-\varepsilon)} \sum_{\substack{\mathbf{p} \in \mathcal{P}_n \\ b(\mathbf{p}) \cap b(\mathbf{q}) \neq \emptyset}} e^{|\mathbf{p}|} \varphi(\mathbf{p}) \\ &\leq a_0 e^{m_0+1} d_n^\varepsilon |\mathbf{q}| = a(\mathbf{q})\end{aligned} \quad (5.4)$$

provided that $\xi \in D$. It follows from Lemma 4.1 and (5.4) that

$$\log \phi_n(z_1, \dots, z_r, \zeta_1, \dots, \zeta_r) = \sum_{A \in \mathcal{A}_n} \left(e^{iZ^{(n)}(A)} - 1 \right) \chi_n(A) \varphi(A) \frac{\alpha^T(A)}{A!}.$$

We note that if $|A| = 1$, then $\alpha^T(A) = 1$ and $A! = 1$. Applying Taylor expansions, we have

$$\log \phi_n(z_1, \dots, z_r, \zeta_1, \dots, \zeta_r) = iI_1 - \frac{1}{2} (I_2 + I_3) - \frac{i}{3!} I_4$$

for some $\theta \in (0, 1)$, where

$$\begin{aligned}I_1 &= \sum_{A \in \mathcal{A}_n} Z^{(n)}(A) \chi_n(A) \varphi(A) \frac{\alpha^T(A)}{A!}, \\ I_2 &= \sum_{\mathbf{p} \in \mathcal{P}_n} Z^{(n)}(\mathbf{p})^2 \chi_n(\mathbf{p}) \varphi(\mathbf{p}), \\ I_3 &= \sum_{A \in \mathcal{A}_n; |A| \geq 2} Z^{(n)}(A)^2 \chi_n(A) \varphi(A) \frac{\alpha^T(A)}{A!}, \\ I_4 &= \sum_{A \in \mathcal{A}_n} Z^{(n)}(A)^3 e^{i\theta Z^{(n)}(A)} \chi_n(A) \varphi(A) \frac{\alpha^T(A)}{A!}.\end{aligned}$$

The first term I_1 is equal to 0 due to symmetry with respect to trade sign. Now we shall consider the second term I_2 and decompose I_2 as follows,

$$I_2 = I_{21} + I_{22} + I_{23},$$

where

$$\begin{aligned} I_{21} &= \sum_{i=1}^r \sum_{j=1}^r z_i z_j \frac{1}{c(n)^2} \sum_{\mathbf{p} \in \mathcal{P}_n} X_{[nt_i]}(\mathbf{p}) X_{[nt_j]}(\mathbf{p}) \chi_n(\mathbf{p}) \varphi(\mathbf{p}), \\ I_{22} &= \sum_{i=1}^r \sum_{j=1}^r z_i \zeta_j \frac{1}{c(n)^2} \sum_{\mathbf{p} \in \mathcal{P}_n} X_{[nt_i]}(\mathbf{p}) Y_{[nt_j]}(\mathbf{p}) \chi_n(\mathbf{p}) \varphi(\mathbf{p}), \\ I_{23} &= \sum_{i=1}^r \sum_{j=1}^r \zeta_i \zeta_j \frac{1}{c(n)^2} \sum_{\mathbf{p} \in \mathcal{P}_n} Y_{[nt_i]}(\mathbf{p}) Y_{[nt_j]}(\mathbf{p}) \chi_n(\mathbf{p}) \varphi(\mathbf{p}). \end{aligned}$$

First, we consider

$$I_{21}(t_i, t_j) = \frac{1}{c(n)^2} \sum_{\mathbf{p} \in \mathcal{P}_n} X_{[nt_i]}(\mathbf{p}) X_{[nt_j]}(\mathbf{p}) \chi_n(\mathbf{p}) \varphi(\mathbf{p})$$

and decompose $I_{21}(t_i, t_j)$ into the sum of $I_{21}^A(t_i, t_j)$ and $I_{21}^B(t_i, t_j)$ given by

$$I_{21}^A(t_i, t_j) = \frac{1}{c(n)^2} \sum_{\substack{\mathbf{p} \in \mathcal{P}_n \\ |\mathbf{p}|=1}} X_{[nt_i]}(\mathbf{p}) X_{[nt_j]}(\mathbf{p}) \chi_n(\mathbf{p}) \varphi(\mathbf{p})$$

and

$$I_{21}^B(t_i, t_j) = \frac{1}{c(n)^2} \sum_{\substack{\mathbf{p} \in \mathcal{P}_n \\ |\mathbf{p}| \geq 2}} X_{[nt_i]}(\mathbf{p}) X_{[nt_j]}(\mathbf{p}) \chi_n(\mathbf{p}) \varphi(\mathbf{p}).$$

It is easily seen that

$$I_{21}^A(t_i, t_j) \sim 2c\sigma_A^2(t_i \wedge t_j) \cdot \frac{n^\alpha}{c(n)^2},$$

where $a_n \sim b_n$ means that $\lim_{n \rightarrow \infty} a_n/b_n = 1$.

Changing the order of sums, we have

$$I_{21}^B(t_i, t_j) = \frac{1}{c(n)^2} \sum_{u=1}^{[nt_i]} \sum_{s=1}^{[nt_j]} \sum_{\substack{\mathbf{p} \in \mathcal{P}_n; |\mathbf{p}| \geq 2 \\ u, s \in b(\mathbf{p})}} v_u(\mathbf{p}) v_s(\mathbf{p}) \varphi(\mathbf{p}) \chi_n(\mathbf{p}).$$

We decompose $I_{21}^B(t_i, t_j)$ into the sum of $I_{211}^B(t_i, t_j)$ and $I_{212}^B(t_i, t_j)$ given by

$$\begin{aligned} I_{211}^B(t_i, t_j) &= \frac{1}{c(n)^2} \sum_{u=1}^{[n(t_i \wedge t_j)]} \sum_{\substack{\mathbf{p} \in \mathcal{P}_n; |\mathbf{p}| \geq 2 \\ u \in b(\mathbf{p})}} v_u(\mathbf{p})^2 \varphi(\mathbf{p}) \chi_n(\mathbf{p}), \\ I_{212}^B(t_i, t_j) &= \frac{1}{c(n)^2} \sum_{\substack{1 \leq u \leq [nt_i] \\ 1 \leq s \leq [nt_j], u \neq s}} v_u(\mathbf{p}) v_s(\mathbf{p}) \varphi(\mathbf{p}) \chi_n(\mathbf{p}). \end{aligned}$$

It follows from the translation invariance of $\varphi(\mathbf{p})$ that

$$\begin{aligned} I_{211}^B(t_i, t_j) &= \frac{2\sigma_B^2}{c(n)^2} [n(t_i \wedge t_j)] \sum_{\substack{\mathbf{p} \in \mathcal{P}_n \\ 0 \in b(\mathbf{p}) \\ |\mathbf{p}| \geq 2}} \varphi(\mathbf{p}) \\ &= \frac{2\sigma_B^2}{c(n)^2} [n(t_i \wedge t_j)] \sum_{m=2}^{m_0} m d_n^m \left(\sum_{t=1}^n \frac{1}{t^\alpha} \right)^{m-1} \\ &\sim \frac{2\sigma_B^2 c(t_i \wedge t_j) n^\alpha}{c(n)^2} \sum_{m=2}^{m_0} m \left(\frac{c}{1-\alpha} \right)^{m-1}. \end{aligned}$$

To handle the second term $I_{212}^B(t_i, t_j)$ we consider polymers \mathbf{p} satisfying the following conditions: (i) $u, s \in b(\mathbf{p})$, (ii) there exist k points in $b(\mathbf{p})$ between u and s , and (iii) there exist q points in $b(\mathbf{p})$ outside of $(u \wedge s, u \vee s)$.

Then we rewrite $I_{212}^B(t_i, t_j)$ as

$$\begin{aligned} I_{212}^B(t_i, t_j) &= \frac{2\mu_B^2}{c(n)^2} \sum_{\substack{1 \leq u \leq [nt_i] \\ 1 \leq s \leq [nt_j] \\ u \neq s}} \sum_{k=0}^{m_0-2} d_n^{k+2} \sum_{q=0}^{m_0-(k+2)} (q+1) \left(d_n \sum_{t=1}^n \frac{1}{t^\alpha} \right)^q \\ &\times \sum_{u \wedge s < t_1 < \dots < t_k < u \vee s} \frac{1}{(t_1 - u \wedge s)^\alpha (t_2 - t_1)^\alpha \dots (t_k - t_{k-1})^\alpha (u \vee s - t_k)^\alpha}. \end{aligned}$$

From the standard argument of calculus we have

$$\begin{aligned} I_{212}^B(t_i, t_j) &\sim \frac{2\mu_B^2}{c(n)^2} \sum_{k=0}^{m_0-2} d_n^{k+2} \frac{n^{k+2}}{n^{(k+1)\alpha}} \sum_{q=0}^{m_0-(k+2)} (q+1) \left(\frac{c}{1-\alpha} \right)^q \\ &\times \left(\frac{1}{n^{k+2}} \sum_{\substack{1 \leq u \leq [nt_i] \\ 1 \leq s \leq [nt_j] \\ u \neq s}} \sum_{u \wedge s < t_1 < \dots < t_k < u \vee s} \frac{1}{\left(\frac{t_1}{n} - \frac{u \wedge s}{n} \right)^\alpha \left(\frac{t_2}{n} - \frac{t_1}{n} \right)^\alpha \dots \left(\frac{u \vee s}{n} - \frac{t_k}{n} \right)^\alpha} \right) \\ &\sim \frac{2\mu_B^2 n^\alpha}{c(n)^2} \sum_{k=0}^{m_0-2} c^{k+2} E_k \int_0^{t_i} du \int_0^{t_j} ds \\ &\times \int_{u \wedge s < x_1 < \dots < x_k < u \vee s} dx_1 \dots dx_k \frac{1}{(x_1 - u \wedge s)^\alpha (x_2 - x_1)^\alpha \dots (x_k - x_{k-1})^\alpha (u \vee s - x_k)^\alpha} \\ &= \frac{2\mu_B^2 n^\alpha}{c(n)^2} \sum_{k=0}^{m_0-2} c^{k+2} J_k E_k \int_0^{t_i} du \int_0^{t_j} ds \frac{1}{|u - s|^{(k+1)\alpha - k}} \\ &= \frac{4\mu_B^2 n^\alpha}{c(n)^2} \sum_{k=0}^{m_0-2} \frac{c^{k+2} J_k E_k}{(1-\alpha)(k+1)((1-\alpha)(k+1)+1)} \cdot \frac{1}{2} \left(t_i^{2H_k} + t_j^{2H_k} - |t_i - t_j|^{2H_k} \right), \end{aligned}$$

where

$$J_k = \int_{0 < w_1 < \dots < w_k < 1} dw_1 \cdots dw_k \frac{1}{w_1^\alpha (w_2 - w_1)^\alpha \cdots (w_k - w_{k-1})^\alpha (1 - w_k)^\alpha},$$

$$E_k = \sum_{q=0}^{m_0 - (k+2)} (q+1) \left(\frac{c}{1-\alpha} \right)^q,$$

$$H_k = \frac{1}{2} ((k+1)(1-\alpha) + 1).$$

The integral J_k is expressed by the Gamma function:

$$J_k = \frac{\Gamma(1-\alpha)^{k+1}}{\Gamma((1-\alpha)(k+1))}.$$

Putting the results of $I_{21}^A(t_i, t_j)$, $I_{211}^B(t_i, t_j)$ and $I_{212}^B(t_i, t_j)$ together we have

$$I_{21}(t_i, t_j) \sim \left(2c\sigma_A^2 + 2\sigma_B^2 c \hat{E} \right) (t_i \wedge t_j) + \mu_B^2 \sum_{k=0}^{m_0-2} a_k^2 \cdot \frac{1}{2} \left(t_i^{2H_k} + t_j^{2H_k} - |t_i - t_j|^{2H_k} \right) \quad (5.5)$$

provided that $c(n) = n^{\frac{1}{2}\alpha}$, where

$$\hat{E} = \sum_{m=2}^{m_0} m \left(\frac{c}{1-\alpha} \right)^{m-1},$$

$$a_k^2 = 4 \frac{c^{k+2} E_k \Gamma(1-\alpha)^{k+1}}{\Gamma((1-\alpha)(k+1) + 2)}.$$

In the same way as $I_{21}(t_i, t_j)$ we define $I_{22}(t_i, t_j)$ and $I_{23}(t_i, t_j)$ with the following results:

$$I_{22}(t_i, t_j) \sim \left(2c\eta_A + 2\eta_B c \hat{E} \right) (t_i \wedge t_j) + \mu_B \gamma_B \sum_{k=0}^{m_0-2} a_k^2 \cdot \frac{1}{2} \left(t_i^{2H_k} + t_j^{2H_k} - |t_i - t_j|^{2H_k} \right), \quad (5.6)$$

$$I_{23}(t_i, t_j) \sim \left(2c\gamma_A + 2\gamma_B c \hat{E} \right) (t_i \wedge t_j) + \gamma_B^2 \sum_{k=0}^{m_0-2} a_k^2 \cdot \frac{1}{2} \left(t_i^{2H_k} + t_j^{2H_k} - |t_i - t_j|^{2H_k} \right). \quad (5.7)$$

Proposition 5.2.

$$\lim_{n \rightarrow \infty} I_2$$

$$\begin{aligned}
&= \sum_{i=1}^r \sum_{j=1}^r z_i z_j \left\{ \left(2c\sigma_A^2 + 2\sigma_B^2 c\hat{E} \right) (t_i \wedge t_j) + \mu_B^2 \sum_{k=0}^{m_0-2} \frac{a_k^2}{2} \left(t_i^{2H_k} + t_j^{2H_k} - |t_i - t_j|^{2H_k} \right) \right\} \\
&+ \sum_{i=1}^r \sum_{j=1}^r z_i \zeta_j \left\{ \left(2c\eta_A + 2\eta_B c\hat{E} \right) (t_i \wedge t_j) + \mu_B \gamma_B \sum_{k=0}^{m_0-2} \frac{a_k^2}{2} \left(t_i^{2H_k} + t_j^{2H_k} - |t_i - t_j|^{2H_k} \right) \right\} \\
&+ \sum_{i=1}^r \sum_{j=1}^r \zeta_i \zeta_j \left\{ \left(2c\gamma_A + 2\gamma_B c\hat{E} \right) (t_i \wedge t_j) + \gamma_B^2 \sum_{k=0}^{m_0-2} \frac{a_k^2}{2} \left(t_i^{2H_k} + t_j^{2H_k} - |t_i - t_j|^{2H_k} \right) \right\}.
\end{aligned}$$

In order to prove Theorem 3.1, it remains to show that I_3 and I_4 converge to zero as $n \rightarrow \infty$. Remarking that

$$\sum_{\substack{A \in \mathcal{A}_n \\ b(A) \cap b(\mathbf{q}) \neq \emptyset}} \left| \Phi(A\xi) \frac{\alpha^T(A)}{A!} \right| \leq a(\mathbf{q})$$

and that $\alpha^T(A) = 0$ unless A is a cluster, we prove that $I_3 \rightarrow 0$ as $n \rightarrow \infty$.

To prove that $I_4 \rightarrow 0$ ($n \rightarrow \infty$), we introduce a functional:

$$F(\xi, \mathbf{q}) = \sum_{\substack{A \in \mathcal{A}_n, |A| \geq 2 \\ b(A) \cap b(\mathbf{q}) \neq \emptyset}} e^{\xi Z^{(n)}(A)} \chi_n(A) \varphi(A) \frac{\alpha^T(A)}{A!}$$

for $\xi \in D$ and $\mathbf{q} \in \mathcal{P}_n$. Obviously, $F(\xi, \mathbf{q})$ is analytic in $\xi \in D$.

Lemma 5.3. *For any $\xi \in D$ and $\mathbf{q} \in \mathcal{P}_n$, we have*

$$|F(\xi, \mathbf{q})| \leq \frac{C_3}{n^{(2-\varepsilon)(1-\alpha)}},$$

where $C_3 = a_0 m_0 e^{m_0+1} c^{2-\varepsilon}$.

Proof. From Lemma 4.1 and (4.1), we have

$$\sum_{\substack{A \in \mathcal{A}_n \\ b(A) \cap b(\mathbf{q}) \neq \emptyset}} \left| e^{\xi Z^{(n)}(A)} \chi_n(A) \varphi(A) \frac{\alpha^T(A)}{A!} \right| e^{d(A)} \leq a(\mathbf{q}). \quad (5.8)$$

Using this inequality, we have

$$\begin{aligned}
|F(\xi, \mathbf{q})| &\leq \sum_{\substack{A \in \mathcal{A}_n, |A| \geq 2 \\ b(A) \cap b(\mathbf{q}) \neq \emptyset}} \left| e^{\xi Z^{(n)}(A)} \chi_n(A) \varphi(A) \frac{\alpha^T(A)}{A!} \right| d_n^{-(1-\varepsilon)(|A|-2)} \\
&\leq d_n^{2(1-\varepsilon)} \sum_{\substack{A \in \mathcal{A}_n \\ b(A) \cap b(\mathbf{q}) \neq \emptyset}} \left| e^{\xi Z^{(n)}(A)} \chi_n(A) \varphi(A) \frac{\alpha^T(A)}{A!} \right| e^{d(A)} \\
&\leq d_n^{2(1-\varepsilon)} a(\mathbf{q}) = \frac{C_3}{n^{(2-\varepsilon)(1-\alpha)}}.
\end{aligned}$$

□

We decompose I_4 as follows,

$$I_4 = I_{41} + I_{42},$$

where

$$I_{41} = \sum_{\substack{\mathbf{p} \in \mathcal{P}_n \\ b(\mathbf{p}) \cap A_n \neq \emptyset}} Z^{(n)}(\mathbf{p})^3 e^{i\theta Z^{(n)}(\mathbf{p})} \varphi(\mathbf{p})$$

and

$$I_{42} = \sum_{\substack{A \in \mathcal{A}_n, \\ |A| \geq 2}} Z^{(n)}(A)^3 e^{i\theta Z^{(n)}(A)} \chi_n(A) \varphi(A) \frac{\alpha^T(A)}{A!}.$$

From Lemma 5.1 and (5.1) we have

$$|I_{41}| \leq C_1^3 n^{1-\frac{3}{2}\alpha} \sum_{\substack{\mathbf{p} \in \mathcal{P}_n \\ 1 \in b(\mathbf{p})}} \varphi(\mathbf{p}) \leq a_0 c C_1^3 n^{-\frac{\alpha}{2}} \rightarrow 0. \quad (n \rightarrow \infty)$$

Lemma 5.4.

$$|I_{42}| \leq 48 C_1^3 C_3 n^{-(1-\alpha)(1-\varepsilon)-\frac{\alpha}{2}} \rightarrow 0. \quad (n \rightarrow \infty)$$

Proof. Since $F(\xi, \mathbf{q})$ is analytic in $\xi \in D$, using the Cauchy formula, we have

$$\begin{aligned} |I_{42}| &\leq n \cdot \max_{\mathbf{q} \in \mathcal{P}_n} \left| \sum_{\substack{A \in \mathcal{A}_n, |A| \geq 2 \\ b(A) \cap b(\mathbf{q}) \neq \emptyset}} Z^{(n)}(A)^3 e^{i\theta Z^{(n)}(A)} \chi_n(A) \varphi(A) \frac{\alpha^T(A)}{A!} \right| \\ &= n \cdot \max_{\mathbf{q} \in \mathcal{P}_n} \left| \frac{\partial^3}{\partial \xi^3} F(i\theta, \mathbf{q}) \right| = n \cdot \max_{\mathbf{q} \in \mathcal{P}_n} \left| \frac{3!}{2\pi} \int_{C(i\theta)} \frac{F(\xi, \mathbf{q})}{(\xi - i\theta)^4} d\xi \right|, \end{aligned}$$

where $C(i\theta) = \{\xi \in \mathbb{C}; |\xi - i\theta| = h(n)\} \subset D$.

Letting $\xi - i\theta = h(n)e^{ix}$ and using Lemma 5.3, we have

$$\begin{aligned} \left| \int_{C(i\theta)} \frac{F(\xi, \mathbf{q})}{(\xi - i\theta)^4} d\xi \right| &= \left| \int_0^{2\pi} \frac{F(i\theta + h(n)e^{ix}, \mathbf{q})}{h(n)^4 \cdot e^{4ix}} i h(n) e^{ix} dx \right| \\ &\leq 16\pi C_1^3 C_3 n^{-(1-\alpha)(2-\varepsilon)-\frac{3}{2}\alpha}. \end{aligned}$$

Hence, we have our assertion. □

In the same way as Lemma 5.4, we have $|I_3| \rightarrow 0$ as $n \rightarrow \infty$.

Hence, we complete the proof of Theorem 3.1.

Acknowledgements

The authors thank the Yukawa Institute for Theoretical Physics at Kyoto University. Discussions during the YITP workshop YITP-W-11-04 on “Econophysics 2011—The Hitchhiker’s Guide to the Economy” were useful to complete this work.

This research was partially supported by a Grant-in-Aid for Scientific Research (C) (No. 21510146).

References

- 1) J.-P. Bouchaud, Y. Gefen, M. Potter and M. Wyart, *Quant. Finance* **4** (2004), 176.
- 2) J.-P. Bouchaud, J. Kockelkoren and M. Potter, *Quant. Finance* **6** (2006), 115.
- 3) J.-P. Bouchaud and M. Potter, *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*, 2nd Edition (Cambridge University Press, 2003).
- 4) R. L. Dobrushin and O. Hryniv, *Commun. Math. Phys.* **189** (1997), 395.
- 5) E. F. Fama, *J. Finance* **25** (1970), 383.
- 6) J. D. Farmer, A. Gerig, F. Lillo and S. Mike, *Quant. Finance* **6** (2006), 107.
- 7) Y. Higuchi, *Z. Wahrsch. Verv. Gebiete* **50** (1979), 287.
- 8) Y. Higuchi, J. Murai and J. Wang, *Adv. Stud. Pure Math.* **39** (2004), 233.
- 9) K. Kuroda and J. Murai, “Stock price process and the long-range percolation”, in *Practical Fruits of Econophysics*, ed. H. Takayasu (Springer-verlag, Tokyo, 2005), p. 163.
- 10) K. Kuroda and J. Murai, *Physica A* **383** (2007), 28.
- 11) K. Kuroda and J. Murai, “A probabilistic model on the long memory property in stock market”, in *International conference 2008 in Okayama, Rising Economies and Regional Cooperation in the East Asia and Europe* (2008), p. 1.
- 12) K. Kuroda and J. Murai, *Prog. Theor. Phys. Suppl. No.* **179** (2009), 26.
- 13) K. Kuroda, J. Maskawa and J. Murai, *Adv. in Mathematical Economics* **14** (2011), 69.
- 14) R. Kotecky and D. Preiss, *Commun. Math. Phys.* **103** (1986), 491.
- 15) F. Lillo and J. D. Farmer, *Stud. Nonlinear Dyn. Econometr.* **8** (2004), 1.
- 16) F. Lillo, S. Mike and J. D. Farmer, *Phys. Rev. E* **71** (2005), 066122.
- 17) R. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 1999).
- 18) C. E. Pfister, *Helv. Phys. Acta.* **64** (1991), 953.

OpenCL Implementation of NeuroIsing

Christopher A. ZAPART^{*)}

The Institute of Statistical Mathematics, Tachikawa 190-8562, Japan

Recent advances in graphics card hardware combined with an introduction of the OpenCL standard promise to accelerate numerical simulations across diverse scientific disciplines. One such field benefiting from new hardware/software paradigms is econophysics. The paper describes an OpenCL implementation of a selected econophysics model: NeuroIsing, which has been designed to execute in parallel on a vendor-independent graphics card. Originally introduced in the paper [C. A. Zapart, “Econophysics in Financial Time Series Prediction”, PhD thesis, Graduate University for Advanced Studies, Japan (2009)], at first it was implemented on a CELL processor running inside a SONY PS3 games console. The NeuroIsing framework can be applied to predicting and trading foreign exchange as well as stock market index futures.

§1. Introduction

Recently there has been an increase in interest in econophysics from physicists, statisticians, economists as well as practitioners (investment professionals, quantitative analysts and traders). Computer-minded investors and hedge fund managers alike could benefit from being able to enhance their trading with new modelling techniques emerging from the field of econophysics. The Efficient Markets Hypothesis (EMH), which forms a cornerstone of the mainstream economics, has in recent years come under criticism from behavioural finance^{2),3)} and econophysics.⁴⁾ The idea of perfectly rational investors acting correctly upon new information does not properly account for “irrational exuberance” or herding⁵⁾ and critical events.⁶⁾ Classical economics assumes that large changes in stock prices are driven by a random flow of news. However, according to a recent econophysics study,⁷⁾ most jumps are not directly linked to the arrival of news items. Instead large movements in financial time series seem to be associated with an increase in herding (imitation) — collective speculative behaviour — in other words internal dynamics of the market. The Efficient Markets Hypothesis and the Random Walk Theory were not intended to provide tools necessary for accurate modelling of such internal market dynamics. Changes in stock prices arise as a result of interactions between a large number of traders, each belonging to different groups, for example producers, speculators, “noise traders”, pension funds, hedge funds, each following different investment objectives/risk profiles that change over time. Trying to capture in a computer simulation all aspects of a real financial market is a difficult task. In contrast with the mainstream economics, computational agent-based models^{8),9)} coming from econophysics attempt to recreate a speculative aspect of these interactions. A positive example can be provided by a study showing how to predict large changes in stock prices directly from an internal dynamics of minority games.¹⁰⁾ In general, behavioural finance, econophysics,

^{*)} E-mail: czapart@ism.ac.jp

information theory etc. provide an alternative¹¹⁾ to the orthodox economics, at least when it comes to modelling a speculative subset of many-trader interactions. Author's current research attempts to offer an econophysics-based computational agents perspective on predicting short-term fluctuations in financial time series.

§2. OpenCL implementation

A full description of the NeuroIsing model as well as forecasting simulations can be found in Refs. 1) and 12). As the model is fairly computationally-expensive, at first experiments were carried out on a SONY PlayStation 3 containing a multi-core CELL processor, as illustrated in Fig. 1. Since IBM has officially put on hold releases of future CELL processors, a decision has been made to make a transition to GPGPU computing, at first utilising NVIDIA CUDA. Subsequently the NeuroIsing code base has been re-implemented in vendor-independent OpenCL, hence ensuring it can be run on whichever future OpenCL-supporting hardware is fastest, for example the yet-unreleased Intel Larrabee.

In the NeuroIsing model individual traders are represented by simple computational agents in the form of multilayer perceptron (MLP) artificial neural networks (non-linear regression models). Although these agents attempt to make trading decisions based on predictions inferred from the past history of share price movements, they are also equally influenced by action taken by other agents.¹³⁾ This is enforced by placing computational agents in nodes of a two-dimensional grid that behaves like a probabilistic Ising model. The agents S_{ij} indexed by a pair of integers $\{i, j\}$ on a 2D $size \times size$ lattice assume binary states $S_k \in \{-1, +1\}$ with a *state probability* $P(S_{ij} = S_k | x_t)$, where x_t symbolises given financial time series $\{x_i\}$ (for example

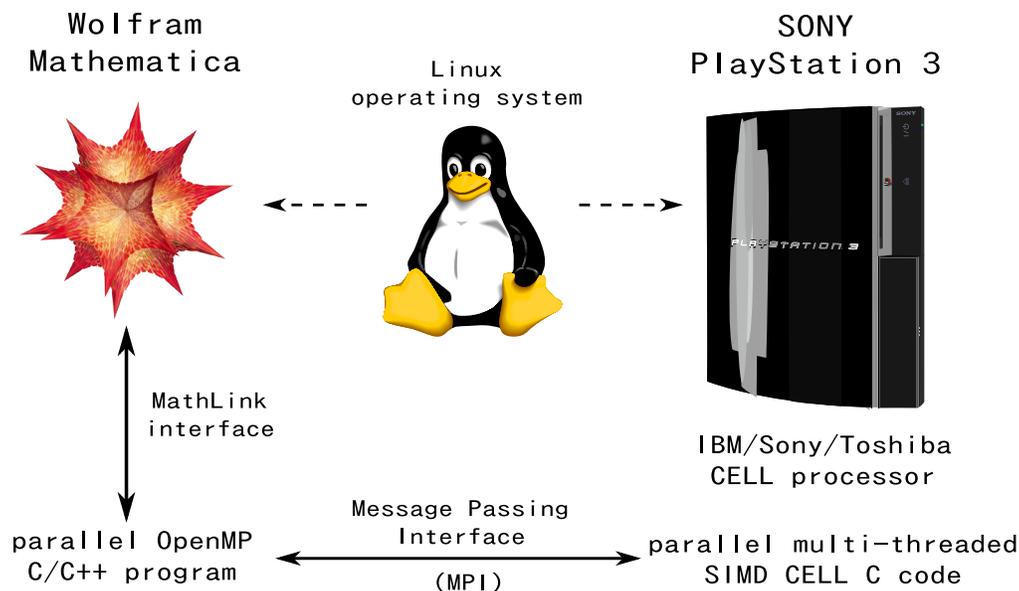


Fig. 1. Early software architecture linking Wolfram Mathematica running on a desktop Linux workstation with Sony PlayStation 3 also running a Gentoo Linux operating system.

the NIKKEI 225 Japanese stock market index) observed up to the current time step t . The overall state probability is modelled by a non-linear function f_{ann} of two variables: the probabilities $P(y_{ij} = S_k | x_t)$ of a non-linear regression model y_{ij} and the Ising spin state $P(S_k)$:

$$P(S_{ij} = S_k | x_t) = f_{\text{ann}}(P(y_{ij} = S_k | x_t), P(S_k)), \quad (1)$$

where y_{ij} is an output from a small artificial neural network associated with the node S_{ij} .

2.1. Regression neural networks

Neural networks are trained to predict the sign of the one-step-ahead next price return x_{t+1} based on m previous returns:

$$y_{ij} = y_{ij}(x_t, x_{t-1}, x_{t-2}, \dots, x_{t-m+1}). \quad (2)$$

The model output y_{ij} is then interpreted as the probability of the next move x_{t+1} being “+1”: $P(y_{ij} = +1 | x_t) = y_{ij}$. Furthermore, in order to introduce a certain degree of diversity to the population of agents, each model is randomly assigned a training dataset of a varying length $T_{ij} \in [T_{\text{min}}, T_{\text{max}}]$. A custom multilayer perceptron (MLP) neural network implementing the non-linear regression function $y_{ij}(x_t, x_{t-1}, x_{t-2}, \dots, x_{t-m+1})$ for $m = 4$ is shown in Fig. 2. By conveniently setting the number of hidden neurons to four one can take advantage of the OpenCL `float4` data type to store hidden layer neuron activations in one variable. Four scalar inputs $\{x_{t-3}, x_{t-2}, x_{t-1}, x_t\}$ are also stored in one `float4` input variable x . Internal neural biases and weights are all stored in a `float4` data type too, as shown in a sample OpenCL code below:

```
typedef struct
{
    //input layer
    float4 x ;

    //hidden layer (four neurons + bias)
    float4 b0 ;
    float4 w1, w2, w3, w4 ;

    float4 db0 ;
    float4 dw1, dw2, dw3, dw4 ;

    float4 hidden, deltaH ;

    //output layer (one neuron)
    float b, db ;
    float4 w, dw ;

    float output, target, error ;
```

```
// (...)
} NeuralForecast ;
```

A hyperbolic tangent activation function is used in the output and hidden neurons, implemented through a `tanh` OpenCL call.

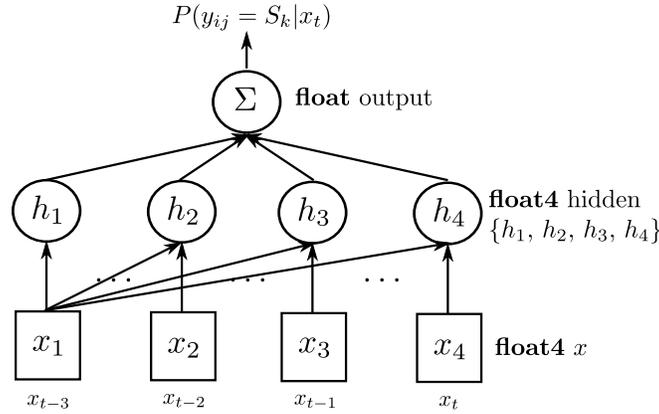


Fig. 2. A regression neural network used to estimate $P(y_{ij} = S_k | x_t)$. For clarity not all neuron connections are shown.

2.2. Extended Ising spin model

The probability $P(S_k)$ of the Ising spin node S_{ij} assuming a state S_k is provided by the Ising model:¹⁴⁾

$$P(S_k) = \frac{\exp(\beta S_k \theta_{ij})}{\exp(-\beta S_k \theta_{ij}) + \exp(\beta S_k \theta_{ij})}, \quad (3)$$

where β is a constant parameter governing the dynamics of the model and θ_{ij} expresses the level of interactions between the node S_{ij} and all the other nodes belonging to the local neighbourhood N_{ij} of S_{ij} . In an extended Ising model used by NeuroIsing multiple-level neighbourhoods are assumed, as illustrated in Fig. 3, which may resemble spin models used in 15). Furthermore, θ used in Eq. (3) is defined as

$$\theta_{ij} = \sum_{k=1}^K \left(\frac{\Phi(k)}{k} \sum_{l \in N_{ij,k}} J_{ij,l} S_l \right), \quad (4)$$

with the neighbourhood function $\Phi(\cdot)$ defined by

$$\Phi(k) = -\tanh\left(\frac{k-\omega}{\rho}\right) \exp\left(-\frac{k}{\tau}\right). \quad (5)$$

The term k seen in the denominator of Eq. (4) stems from the observation that the number of nodes in each successive neighbourhood is directly proportional to the neighbourhood index k . The function $\Phi(\cdot)$, illustrated in Fig. 4, expresses various

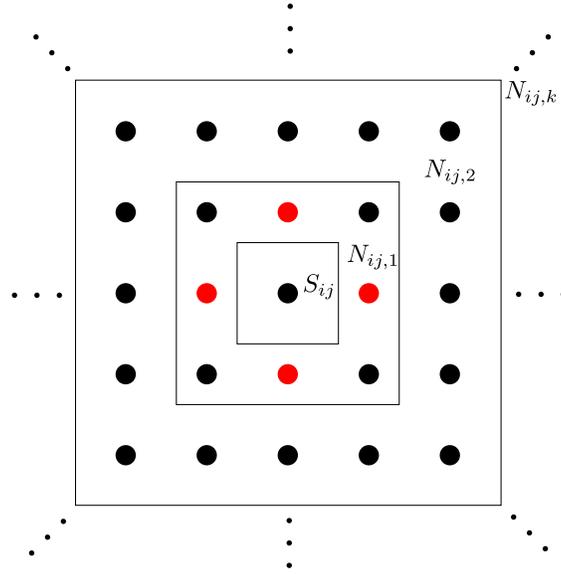


Fig. 3. Multiple-level neighbourhoods $N_{ij,k}$ for a node S_{ij} in an extended Ising model. The neighbouring nodes from the original 2D Ising model are shown in red.

forms of imitation ($\Phi > 0$) and anti-imitation ($\Phi < 0$) for near and far neighbourhoods. For given financial time series $\{x_i\}$ appropriate behaviour regimes are obtained through finding optimum parameters ω , ρ and τ using genetic algorithms. The connection strengths $J_{ij,l}$ between the current node S_{ij} and its neighbours S_l from the neighbourhood $l \in N_{ij,k}$, initially all set to $J_{ij,l} = +1$, can also be made dependent on the recent trading performance of the node S_l :

$$J_{ij,l} = \tanh(\mu c_l), \quad \mu > 0, \quad (6)$$

$$c_l = \kappa \operatorname{sgn}(x_t) \operatorname{sgn}(S_l) + (1 - \kappa) c_l, \quad (7)$$

where $\mu > 0$ and κ is an exponential smoothing coefficient. Calculating the *confidence to trade* c_l is equivalent to assigning virtual scores to binary strategy tables used in realistic minority game models.¹⁶⁾ Varying the connection strength J_{ij} as a function of other agents' trading performance is akin to implementing herding behaviour observed in real markets.¹³⁾

Due to using multiple neighbourhoods $N_{ij,k}$, calculating θ_{ij} involves computing interactions between a given node $\{i, j\}$ and all remaining nodes in the grid, separated by increasing distances indexed by k . These interactions are partly governed by the terms $\frac{\Phi(k)}{k}$ unique to each node $\{i, j\}$. To avoid repeatedly computing the same values, for a given set of parameters ω , ρ and τ , values $\frac{\Phi(k)}{k}$ are pre-calculated and cached in a global memory GPU buffer `ising_matrix_buf` as a long floating point vector of the length $size^2 \times size^2$, where $size$ denotes the dimension of the Ising grid (in actual simulations $size = 30$).

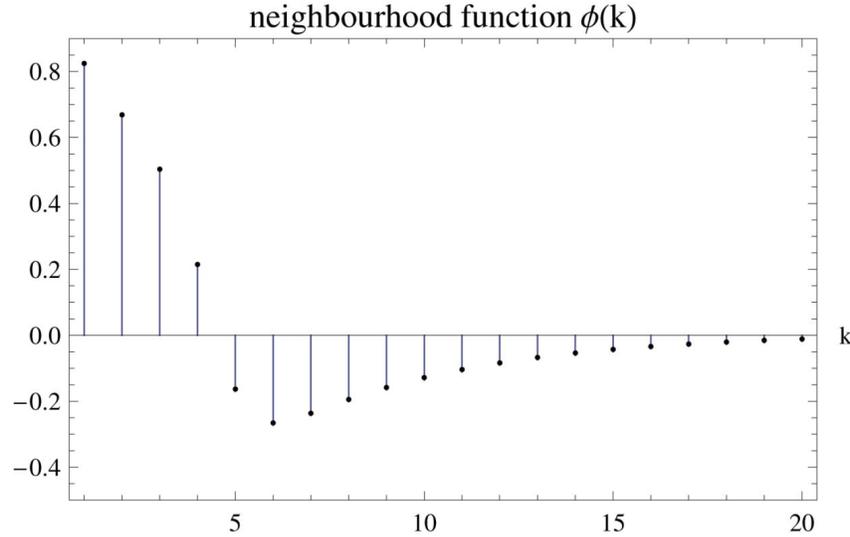


Fig. 4. An example hyperbolic tangent neighbourhood function $\Phi(k)$, $\omega = 4.5$, $\rho = 1$, $\tau = 5.0$ for a node S_{ij} used in an extended Ising model with an asymptotic exponential decay to 0 for large k .

2.3. Parallel genetic algorithms

Choosing the right set of parameters adequate for modelling given financial time series is not straightforward. A numerical optimisation procedure is needed to fine-tune model settings. Gradient-based maximum-likelihood techniques cannot be easily applied because of the difficulty in calculating analytically gradients of any assumed error cost function. To train NeuroIsing on particular financial time series, genetic algorithms were first used in Ref. 1). There are several advantages of evolutionary programming methods. One is their ability to optimise at the same time both the model parameters as well as the model architecture (for example the range $[T_{min}, T_{max}]$). Another is their inherent parallelism facilitating a straightforward GPGPU implementation, as illustrated schematically in Fig. 5.

The NeuroIsing model uses a $size \times size$ 2D grid with nodes indexed by an integer pair $\{i, j\}$. Instead of storing them in $size \times size$ matrices, all state variables for nodes $\{i, j\}$ are stored in long vectors of the length $size^2$. Evolutionary programming involves maintaining a population of N solutions (different sets of NeuroIsing parameters). At each iteration genetic algorithms need to evaluate a fitness objective function which involves scheduling N parallel NeuroIsing runs, each executed with a distinct set of parameters. Furthermore, given the probabilistic nature of the model, for every parameter set the resulting outputs are averaged from M independent NeuroIsing runs, each started with different initial conditions. In terms of data requirements, a parallel NeuroIsing implementation hence requires storing its internal states (node state probabilities etc.) in floating-point vectors of the length $size^2 \times M \times N$. Each OpenCL thread operates on a single item of the NeuroIsing state vectors leading to a total number of OpenCL threads equal to $30^2 \times 10 \times 25 = 225,000$, which may be large enough to mitigate global memory access latencies.

prices of the NIKKEI 225 Japanese stock market index. In Ref. 1) the daily NIKKEI 225 log-returns were re-scaled using the Volatility Index Japan (VXJ) implied options volatility in order to keep daily returns within the constant range $[-1, +1]$:

$$x_t = \operatorname{erf} \left(\frac{100\sqrt{252}(\log p_t - \log p_{t-1})}{VXJ_t\sqrt{2}} \right), \quad (8)$$

where x_t denotes transformed (re-scaled) returns used as inputs to regression neural networks in §2.1, p_t and p_{t-1} denote raw Nikkei 225 index values measured at times t and $t - 1$, and the Volatility Index Japan time series VXJ_t can be downloaded freely from <http://www-csfi.sigmath.es.osaka-u.ac.jp/>. However, the implied volatility values, listed on the Osaka University, Center for the Study of Finance and Insurance website, are only made available publicly once a week during the weekend. Therefore using Eq. (8) in real trading carries a risk of using incorrect data since during the following week the most up-to-date implied volatilities are not available. The other issue that occurs in real trading is the need to diversify risk, not to rely on a single instrument, for example by trading multiple foreign exchange currency pairs. However, the author does not have access to implied options volatility data from foreign exchange markets nor individual Japanese equities, which would be necessary for initial pre-processing using Eq. (8).

In recent markets dominated by high frequency trading, relying on data sampled once a day means losing potentially invaluable intra-day price action. For example, the so-called “2010 Flash Crash” took place within a 30-minute timespan. However, simply sampling the time series at ever decreasing fixed intervals (for example one hour or one minute) does not offer a complete solution. The problem is that on an intra-day basis the time flows in a non-linear fashion. Trading activity slows down during night, only to pick up around important news releases and other noteworthy events. To some extent this problem has been addressed in Ref. 12) by employing a volatility-adjusted intrinsic trading time.¹⁷⁾ Experimental results published in 12) showed a clear improvement in simulated forex trading in comparison with a fixed-interval sampling scheme. However, the intrinsic trading time still relies on calibrating the relationship between price and volatility using historical data.

In contrast, recently the author has been attempting to avoid at all cost having to fit time series models to past data as this carries a high risk of over-fitting models to historical data. Therefore with the recent move away from using once-a-day closing quotes towards high-frequency high-volume tick data time series, the model is currently undergoing re-engineering to meet the challenges of processing in real-time incoming foreign exchange currency tick data. The goal is to produce a fully computerised forecasting/trading system capable of automatically placing trades in the foreign exchange markets 24 hours per day, without any human intervention, and without resorting to fitting any time series.

To this end, instead of using genetic algorithms to maximise a single global fitness function (in other words fitting time series models, still employed in Ref. 1)), the Simulated Annealing Monte Carlo method will be used to minimise individual fitness (utility) functions (lower spin-interaction energy levels of each node in the 2D

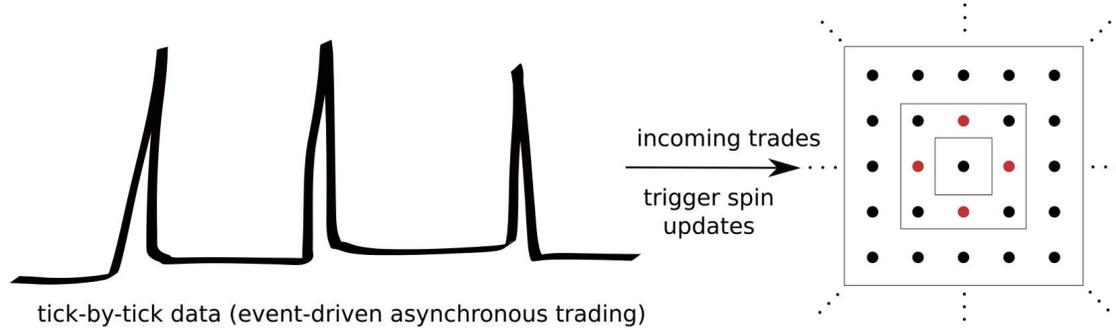


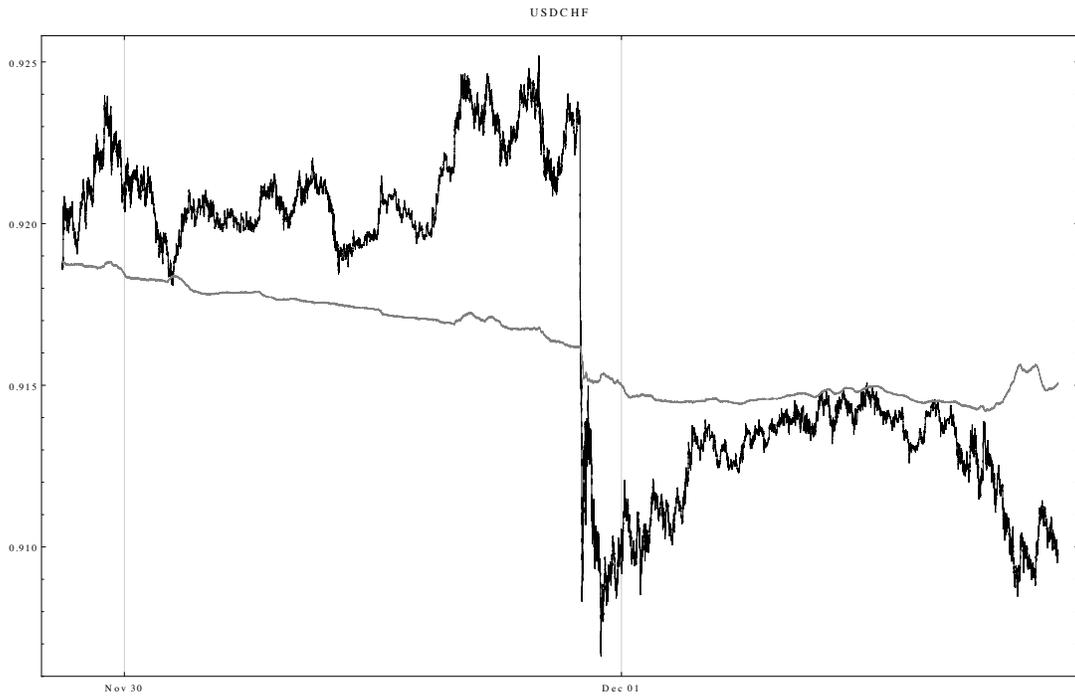
Fig. 7. Event-based trading where updates to the model (Ising spin flips) are driven by incoming new trades that are spaced irregularly in time, as opposed to fixed time intervals (for example once-a-day, once-hourly) commonly used by econometrics time series models.

Ising model). The revised research calls for encoding traders' behaviour in the form of an energy function shown in Eq. (9):

$$h_i = \underbrace{-\alpha \sum_{j \in N_i} S_i S_j J_{ij}}_{\text{copy successful traders}} - \underbrace{\beta \min(p_i, 0)}_{\text{cut losses}} + \underbrace{\gamma \max(p_i, 0)}_{\text{take profits}}, \quad (9)$$

where N_i denotes the local neighbourhood of the i th Ising node (the i th trader) distributed on a 2D square lattice, $\alpha, \beta, \gamma > 0$, $\beta > \gamma$ are constant model parameters, $S_i, S_j \in \{-1, +1\}$ denote spins (traders' buy/sell decisions), p_i, p_j are vector elements holding current profits/losses that are updated after receiving each tick data and J_{ij} expresses the interaction strength between nodes i and j . As an added bonus, the Monte Carlo spin dynamics of the Ising model provides a natural way to model the non-linear flow of time observed in 24-hour foreign exchange markets. Instead of adopting an intrinsic trading time, an alternative way to model the flow of time is to associate the number of spin updates per one Monte Carlo step with the number of forex transactions (or bid/ask changes) occurring within a predefined fixed time interval. Hence the new model facilitates realistic event-based trading, as illustrated in Fig. 7.

Setting J_{ij} in Eq. (9) to the recent trading performance appearing in Eq. (7) causes traders (spins) to exhibit a herding behaviour through imitating recently successful neighbouring traders. Figures 8(a) and (b) show representative Monte Carlo simulations performed for two currency pairs: USD/CHF and AUD/JPY. In particular cumulative profits/losses, summed over all 30×30 agents, are plotted as a function of time together with the performance of the underlying currency pairs. It can be observed that, notwithstanding brief periods of positive spikes, overall agents tend to steadily lose money over time. Losses made by agents are not due to the bid/ask spread (transaction costs) since mid-point prices were used in the simulation. Rather, the losses can be attributed to traders attempting to copy behaviour of recently successful other traders. Since over the long term financial markets are said to be mean-reverting, it can be speculated that recently successful traders may become prone to encountering losses after long successful runs.

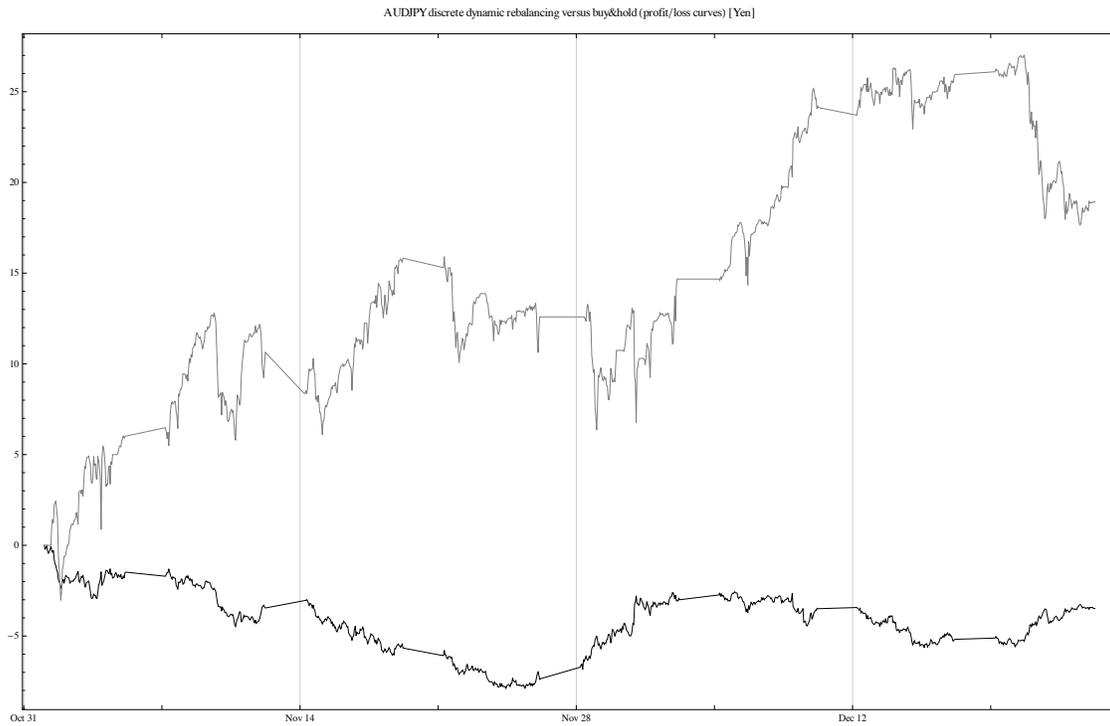


(a)

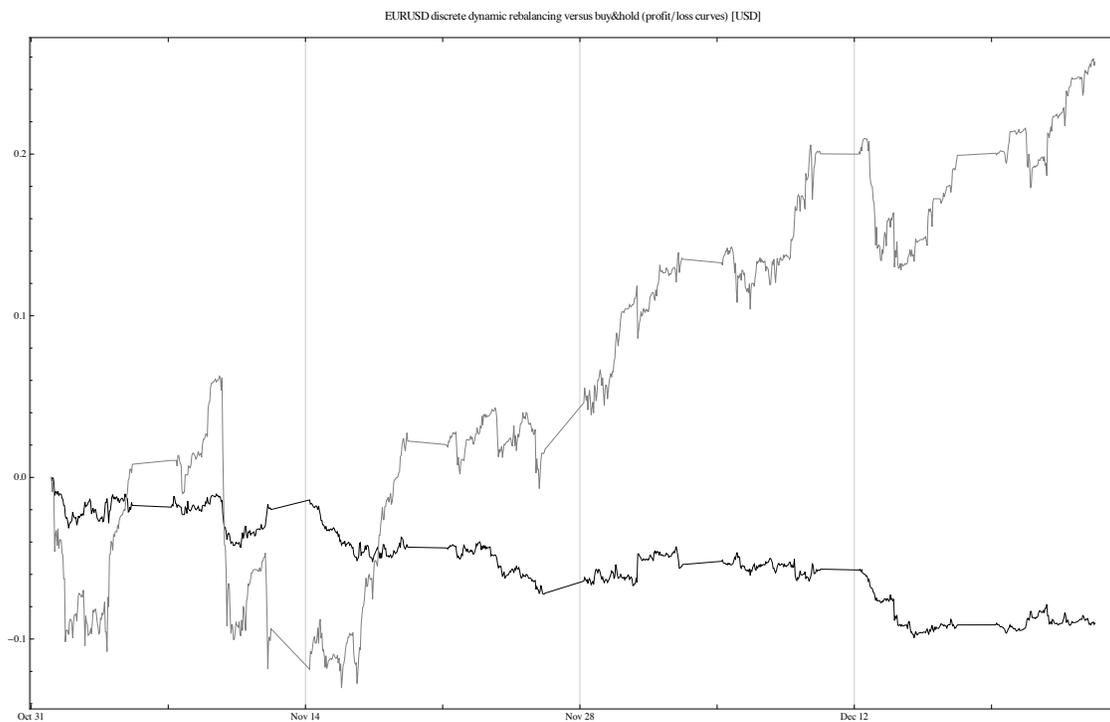


(b)

Fig. 8. Cumulative profit/loss curves (grey line) for all the agents (Ising spins) plotted together with the underlying foreign currency pairs (black line) for the period between 2011-11-29 21:00:00 and 2011-12-01 21:05:00. Examples for two currency pairs are shown: (a) - USD/CHF and (b) - AUD/JPY.

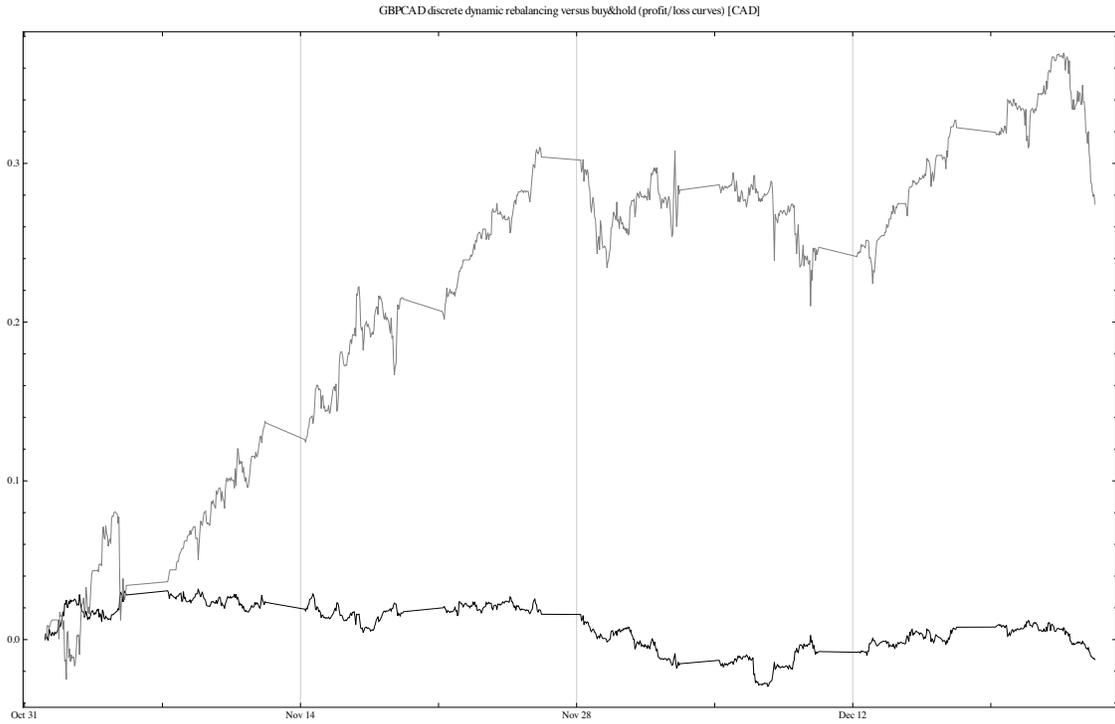


(a)

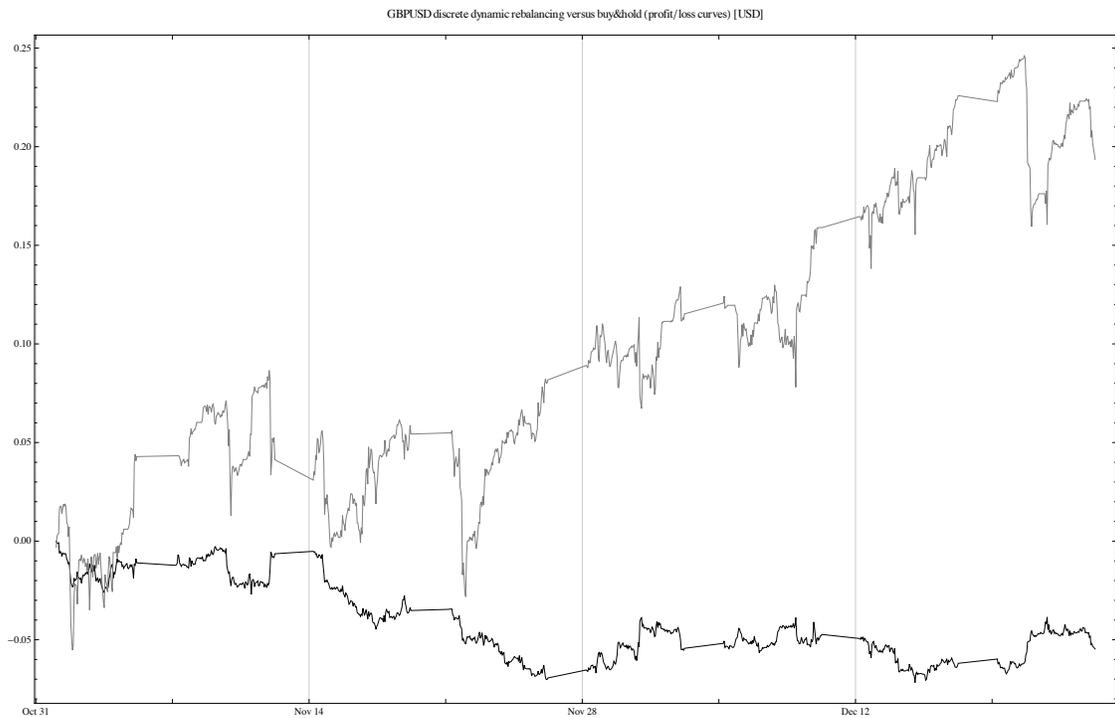


(b)

Fig. 9. Simulated trading performances (grey line) of a contrarian trading system for a range of currency pairs (black line): (a) - AUD/JPY and (b) - EUR/USD.

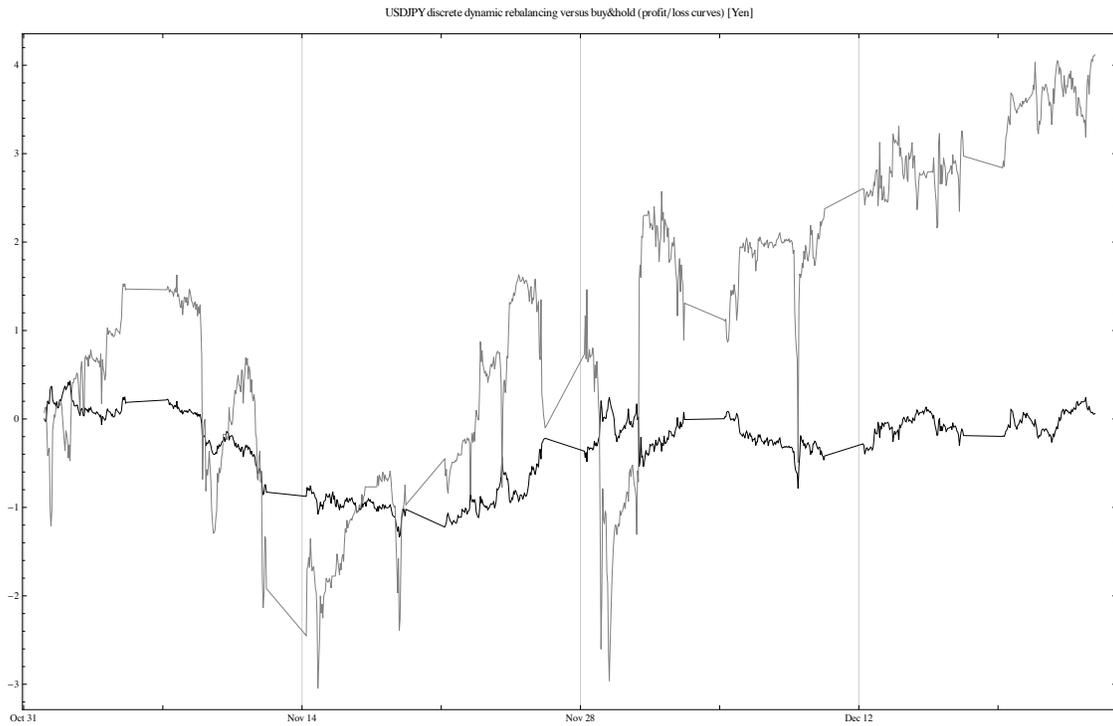


(a)

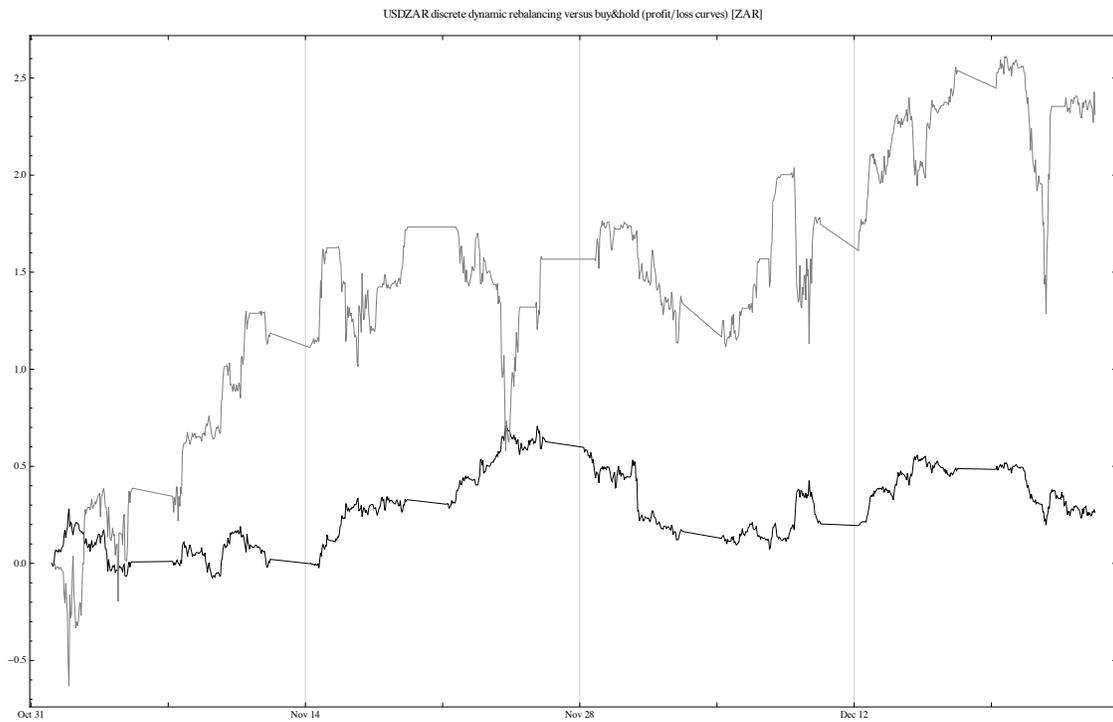


(b)

Fig. 10. Simulated trading performances (grey line) of a contrarian trading system for a range of currency pairs (black line): (a) - GBP/CAD and (b) - GBP/USD.



(a)



(b)

Fig. 11. Simulated trading performances (grey line) of a contrarian trading system for a range of currency pairs (black line): (a) - USD/JPY and (b) - USD/ZAR.

The knowledge that on average most agents (Ising spins) tend to lose money can form the foundation of a contrarian trading system. One can simply count the numbers of UP and DOWN spins, thus measure the NET difference between buyers and sellers, and then place contrarian trades with the foreign exchange dealer, i.e. whenever there are significantly more UP than DOWN spins, one can place a SELL order. Conversely, if DOWN spins are predominant then one can initiate a BUY order. Simulated trading performances of such a contrarian trading system, carried out for a range of currency pairs, are shown in Figs. 9–11.

§4. Conclusions

In the first part the paper outlined basics of implementing the NeuroIsing model using the OpenCL GPGPU parallel programming paradigm. Direct like-for-like benchmark comparisons between the prior CELL and current GPGPU implementations are difficult. Whilst moving the code from CELL SPU Altivec first to Intel SSE, then to NVIDIA CUDA, and subsequently followed by the vendor-independent OpenCL, the actual implementation was changed (re-implemented) at each step. This reflects the way real software development often proceeds. As a rough estimate, compared with what used to take about two hours to run on SONY PS3 several years ago, a functionally equivalent OpenCL code currently only takes a few minutes to run on NVIDIA GTX 260. The goal of the research is not to compare/benchmark the code across different hardware. Instead, the goal is to produce a fast working implementation of NeuroIsing — matched to the available hardware — that can be applied in practice to predict financial time series. The adoption of OpenCL instead of NVIDIA CUDA helps to future-proof the existing NeuroIsing code base to ease transitions to newer, faster hardware platforms whilst avoiding being tied to any particular vendor.

In the final section the paper showed preliminary results from recent research developments which adopt an event-driven asynchronous model architecture. The ideas outlined in this paper have already been put into practice. Following an integration within the FXCM trading API, a fully automated trading system uses the Ising model to execute short-term trades in the foreign exchange markets using real money.

References

- 1) C. A. Zapart, “Econophysics in Financial Time Series Prediction”, PhD thesis, Graduate University for Advanced Studies, Japan (2009).
- 2) A. Lo, *J. of Portfolio Management* **30(5)** (2004), 15.
- 3) A. Lo, *J. of Investment Consulting* **7** (2005), 21.
- 4) C. Eom, S. Choi, G. Oh and W.-S. Jung, *Physica A* **387** (2008), 4630.
- 5) R. Shiller, *Irrational Exuberance*, 2nd ed. (Princeton University Press, 2005).
- 6) D. Sornette, *Why Stock Markets Crash: Critical Events in Complex Financial Systems* (Princeton University Press, 2003).
- 7) A. Joulin, A. Lefevre, D. Grunberg and J. P. Bouchaud, arXiv:0803.1769v1.
- 8) D. Challet, M. Marsili and Y. Zhang, *Minority Games: Interacting Agents in Financial Markets (Oxford Finance)* (Oxford University Press, 2005).
- 9) D. Challet and Y. Zhang, *Physica A* **246** (1997), 407.

- 10) D. Lamper, S. Howison and N. F. Johnson, *Phys. Rev. Lett.* **88** (2001), 017902.
- 11) A. Lo and A. C. MacKinlay, *A Non-Random Walk Down Wall Street* (Princeton University Press, 1999).
- 12) C. A. Zapart, *Predicting Foreign Exchange Time Series with an Extended Ising Model, a chapter in Computer Games: Learning Objectives, Cognitive Performance and Effects on Development* (NOVA Science Publishing, 2010).
- 13) J. Apesteguia, S. Huck and J. Oechssler, *J. of Economic Theory* **136(1)** (2007), 217.
- 14) L. E. Reichl, *A modern course in statistical physics* (John Wiley & Sons, Inc., 1998).
- 15) E. Burgos, H. Ceva and R. P. J. Perazzo, *Physica A* **354** (2005), 518.
- 16) P. Jeffries, M. L. Hart, P. M. Hui and N. F. Johnson, *Eur. Phys. J. B* **20** (2001), 493.
- 17) U. A. Muller, R. Dacorogna, R. D. Davé, O. V. Pictet, R. B. Olsen and J. R. Ward, “Fractals and Intrinsic Time – A Challenge to Econometricians”, Opening address of the XXXIXth International Conference of the Applied Econometrics Association (AEA), Real Time Econometrics – Submonthly Time Series, Ascona, Switzerland (1993).

Analysis of Realized Volatility in Two Trading Sessions of the Japanese Stock Market

Tetsuya TAKAISHI,^{1,*} Ting Ting CHEN² and Zeyu ZHENG³

¹*Hiroshima University of Economics, Hiroshima 731-0192, Japan*

²*Faculty of Integrated Arts and Sciences, Hiroshima University,
Higashi-Hiroshima 739-8521, Japan*

³*Department of Environmental Sciences, Tokyo University of Information Sciences,
Chiba 265-8501, Japan*

We analyze realized volatilities constructed using high-frequency stock data on the Tokyo Stock Exchange. In order to avoid non-trading hours issue in volatility calculations we define two realized volatilities calculated separately in the two trading sessions of the Tokyo Stock Exchange, i.e. morning and afternoon sessions. After calculating the realized volatilities at various sampling frequencies we evaluate the bias from the microstructure noise as a function of sampling frequency. Taking account of the bias to realized volatility we examine returns standardized by realized volatilities and confirm that price returns on the Tokyo Stock Exchange are described approximately by Gaussian time series with time-varying volatility, i.e. consistent with a mixture of distributions hypothesis.

§1. Introduction

In recent years statistical properties of asset price returns have been extensively studied in econophysics.^{1)–5)} One of the pronounced properties is that the probability distribution of returns exhibits a fat-tailed distribution which is not Gaussian.^{6)–9)} It is revealed that the tail distributions of returns exhibit the power law^{10)–12)} and the distributions of returns at short time scale are well fitted by Student's t-distributions^{13)–17)} which is also known as the q-Gaussian distributions in nonextensive statistical mechanics.^{18),19)} It is crucial to understand the price dynamics which results in the fat-tailed distributions. A possible explanation for the origin of the fat-tailed distributions is that the price dynamics follows the Gaussian process with time-varying volatility, which is called the mixture of distributions hypothesis (MDH).²⁰⁾ With this hypothesis each return r_t at time t is described by $r_t = \sigma_t \epsilon_t$, where σ_t^2 is a variance of the Gaussian distribution and ϵ_t is a Gaussian random variable with mean 0 and variance 1. From the MDH the return distributions are derived as the superposition of the conditional return distribution with σ_t and the distribution of volatility σ_t^2 . This view is closely related to the superstatistics²¹⁾ where the statistics of physical systems are separated by two time scales and unconditional distributions are obtained by the superposition of these statistics.

In the MDH the shape of return distributions is determined by the volatility distributions which are not known a priori. Empirical studies suggested that the volatility distributions are described by the inverse gamma distribution or log-normal distribution^{13),17),22)–25)} with which the unconditional return distributions

^{*}) E-mail: tt-taka@hue.ac.jp

result in fat-tailed distributions. Especially the inverse gamma distribution gives the Student's t-distribution (q-Gaussian distribution) for the unconditional return distribution.

The MDH itself does not derive the shape of volatility distributions. The volatility dynamics which accounts for the shape of volatility distributions may depend on microscopic features of the markets such as volume, transactions, information arrival etc.^{26)–29)} which we do not address here. Toward the complete dynamics of price, it is important to check whether the MDH holds for the real market price data.

The consistency between the MDH and the real markets data can be checked by the standard normality of the returns standardized by volatility. Namely under the MDH the standardized returns given by r_t/σ_t should show Gaussian-distributed variables with variance (or standard deviation) 1 and kurtosis 3. Note that it is crucial to check both properties of variance 1 and kurtosis 3. In order to perform this check we need to estimate the value of volatility since volatility is not a direct observable from the market data. The conventional and popular approach for volatility estimation is to use parametric models, such as GARCH-type^{30),31)} and stochastic volatility models^{32)–34)} which are designed to capture the relevant properties observed in financial time series, e.g. volatility clustering and fat-tailed distribution for returns. Each model is constructed under specific assumptions for the time-varying volatility process. The volatility measures obtained from those models could differ each other. Thus the validity for volatility measures is also another issue which has to be considered.

In this paper in order to deal with more accurate volatility we focus on realized volatility constructed as a sum of intraday squared returns^{23),24),35)} and examine the view of the MDH for stock returns on the Tokyo Stock Exchange. The recent availability of high-frequency financial data enables us to easily access to calculations of realized volatility and an advantage over parametric models is that realized volatility is a model-free estimate. Furthermore if there is no measurement error this estimate provides an unbiased volatility measure for the integrated volatility and converges to the integrated volatility in the limit of infinite high sampling frequency. However actual market prices suffer from measurement errors caused by microstructure noise such as bid-ask spread etc.³⁶⁾ Therefore in practical calculation realized volatility is considered to be biased by such unwanted microstructure noise. Provided that the log-price observed in the market is contaminated with an independent noise³⁷⁾ we expect that this bias will be especially crucial at high frequency and increases with increasing the sampling frequency. Such behavior has been actually observed and can be depicted in the so-called “volatility signature plots”.³⁸⁾ In order to reduce this bias and also to maintain accuracy of the measurement typically the returns used for realized volatility calculations are sampled at a moderate frequency, e.g. 5 min. frequency.^{23),24),39)–41)}

When we deal with the daily volatility estimated as realized volatility another problem arises, that is non-trading hours issue in measuring daily volatility. In the Tokyo Stock Exchange, there are two trading sessions in a day, i.e. Morning Session (MS) 9:00-11:00 and Afternoon Session (AS) 12:30-15:00. Thus available stock price data are taken only in the two sessions and outside of these sessions no available data

exist. If the realized volatility is calculated with data only from trading hours, the resulting measure could be biased and unreliable as daily volatility. To circumvent this problem Hansen & Lunde⁴²⁾ proposed to introduce an adjustment factor which corrects the realized volatility so that the average of the realized volatility matches the variance of daily returns.

The standard normality of standardized return time series has been examined by using realized volatility^{23)-25),37),43)} and it is found that the distributions of the standardized return series are approximately described by Gaussian distributions. However from the analysis of stocks in the Dow Jones Industrial Average (DJIA) the standard deviations of the standardized return series turned out to be still apart from 1 and all the values are less than 1. Moreover the values of the kurtosis also vary significantly around 3.

Instead of using log-price returns Ref. 44) examines the price change for U.S. stocks and finds that while the standardized price changes are well described by Gaussian-distributed random variables the kurtosis exceeds 3. In addition the variance of the the standardized price changes has not been examined. Thus it has not been concluded that the MDH exactly holds for the price changes.

From previous studies it is turned out that the return dynamics is approximately consistent with the MDH. However the origin of the deviations from the MDH which appear in variance and kurtosis are not known yet. In order to reveal the origin and examine whether the MDH precisely holds for the returns it is important to control the bias in measuring volatility. In this paper we carefully evaluate the bias from the microstructure noise remaining in the realized volatility and examine the standard normality of standardized returns and the consistency with the MDH.

Under the MDH the standardized returns should exhibit properties of variance 1 and kurtosis 3, i.e. standard normality. In order to examine both properties on the Tokyo Stock Exchange we have to avoid “non-trading hours” issue, i.e. introduction of the Hansen & Lunde (HL) adjustment factor to the realized volatility. Otherwise the realized volatility with the HL adjustment factor can change the values of variance and we cannot check the property of variance 1 under the MDH. In Ref. 25) standardized daily returns on the Tokyo Stock Exchange have been analyzed and it is found that the standardized daily returns exhibit the Gaussianity, i.e. kurtosis 3. However due to the “non-trading hours” issue the property of variance 1 which is needed for the consistency check of the MDH has not been confirmed yet. To avoid “non-trading hours” issue here we do not deal with a whole day volatility. Instead we calculate two realized volatilities separately in the two trading periods, i.e. one measured in the MS and the other in the AS. In this way we obtain two realized volatilities RV_{MS} and RV_{AS} without the HL adjustment factor. By investigating standardized returns in the MS and AS separately we are able to examine both properties of variance 1 and kurtosis 3, i.e. the MDH.

The outline of the paper is as follows. Section 2 describes the realized volatility, the microstructure noise in the realized volatility and the HL adjustment factor. In §3 we explain the data on the Tokyo Stock Exchange analyzed in this study. In §4 we analyze the microstructure noise effects on the realized volatility. In §5 the MDH is examined by the returns standardized by realized volatility. We conclude in §6

with a brief summary.

§2. Realized volatility

The realized volatility is a model-free estimate of volatility constructed as a sum of squared returns.^{23),24),35)} Let us assume that the logarithmic price process $\ln p(s)$ follows a continuous time stochastic diffusion,

$$d \ln p(s) = \tilde{\sigma}(s) dW(s), \quad (2.1)$$

where $W(s)$ stands for a standard Brownian motion and $\tilde{\sigma}(s)$ is a spot volatility at time s . In financial applications our main interest is to measure an integrated volatility which is defined by

$$\sigma_h^2(t) = \int_t^{t+h} \tilde{\sigma}(s)^2 ds, \quad (2.2)$$

where h stands for the interval to be integrated. If we consider daily volatility h takes one day. Since $\tilde{\sigma}(s)$ is latent and not available from market data, Eq. (2.2) can not be evaluated analytically.

Constructing n intraday returns from high-frequency data, the realized volatility RV_t is given by a sum of squared intraday returns,

$$RV_t = \sum_{i=1}^n r_{t+i\Delta}^2, \quad (2.3)$$

where Δ is a sampling period*) defined by $\Delta = h/n$ and returns are given by log-price difference,

$$r_{t+i\Delta} = \ln P_{t+i\Delta} - \ln P_{t+(i-1)\Delta}. \quad (2.4)$$

Without any bias RV_t goes to the integrated volatility of Eq. (2.2) in the limit of $n \rightarrow \infty$. In the presence of bias such as microstructure noise,³⁶⁾ the convergence of RV_t to the integrated volatility is not guaranteed. Following Zhou³⁷⁾ we assume that the log-price observed in financial markets is contaminated with independent noise, i.e.

$$\ln P_t^* = \ln P_t + \xi_t, \quad (2.5)$$

where $\ln P_t^*$ is the observed log-price in the markets which consists of the true log-price $\ln P_t$ and noise ξ_t with mean 0 and variance ω^2 .

Using this assumption the observed return r_t^* is given by

$$r_t^* = r_t + \eta_t, \quad (2.6)$$

where $\eta_t = \xi_t - \xi_{t-\Delta}$. Thus RV_t^* which is actually observed from the market data is obtained as a sum of the squared returns r_t^* ,

$$RV_t^* = \sum_{i=1}^n (r_{t+i\Delta}^*)^2, \quad (2.7)$$

*) Small sampling period corresponds to high sampling frequency.

$$= RV_t + 2 \sum_{i=1}^n r_{t+i\Delta} \eta_{t+i\Delta} + \sum_{i=1}^n \eta_{t+i\Delta}^2. \quad (2.8)$$

With these independent noises the bias appears as $\sum_{i=1}^n \eta_{t+i\Delta}^2$ which corresponds to $\sim 2n\omega^2$. Thus due to the bias the RV_t^* diverges as $n \rightarrow \infty$. Such divergent behavior has been seen in the volatility signature plot³⁸⁾ for liquid assets.

To avoid distortion from microstructure noise a good sampling frequency which reduces the bias but maintains accuracy of the realized volatility measure should be considered. It is suggested that 5 min. sampling frequency is short enough for realized volatility calculation.^{23),24)} In this paper we also use 5 min. sampling frequency to analyze standardized returns.

When we consider daily realized volatility we have to cope with another problem which is that high-frequency data are not available for the entire 24 hours except for some exchange rates. At the Tokyo stock exchange market domestic stocks are traded in the two trading sessions: (1) morning trading session (MS) 9:00-11:00. (2) afternoon trading session (AS) 12:30-15:00. The daily realized volatility calculated without including intraday returns during the non-traded periods can be underestimated.

Hansen and Lunde⁴²⁾ advocated an idea to circumvent the problem by introducing an adjustment factor which modifies the realized volatility so that the average of the realized volatility matches the variance of the daily returns. Let (R_1, \dots, R_N) be N daily returns. The adjustment factor c is given by

$$c = \frac{\sum_{t=1}^N (R_t - \bar{R})^2}{\sum_{t=1}^N RV_t}, \quad (2.9)$$

where \bar{R} denotes the average of R_t . Then using this factor the daily realized volatility is modified to cRV_t . When we analyze the returns standardized by the realized volatility this adjustment factor largely affects the standard deviation of the standardized returns.

In order to avoid the non-trading hours issue and the introduction of the adjustment factor to the realized volatility we consider two realized volatilities: (i) RV_{MS} , realized volatility in the morning session and (ii) RV_{AS} , realized volatility in the afternoon session. Since these realized volatilities are calculated separately and each does not include non-trading hours, no adjustment factor is needed.

§3. Data analyzed

Our analysis is based on data for 5 stocks, 1: Mitsubishi Co., 2: Nomura Holdings Inc., 3: Nippon Steel, 4: Panasonic Co. and 5: Sony Co. These stocks are listed in the Topix core 30 index which includes liquid stocks of the Tokyo Stock Exchange. Our data set begins June 3, 2006 and ends December 30, 2009.

Figure 1 shows daily intraday return time series in the different time zones of the Tokyo Stock Exchange. The figure shows the time series of Mitsubishi Co. as a representative one. Each return is calculated by the log-price difference between the

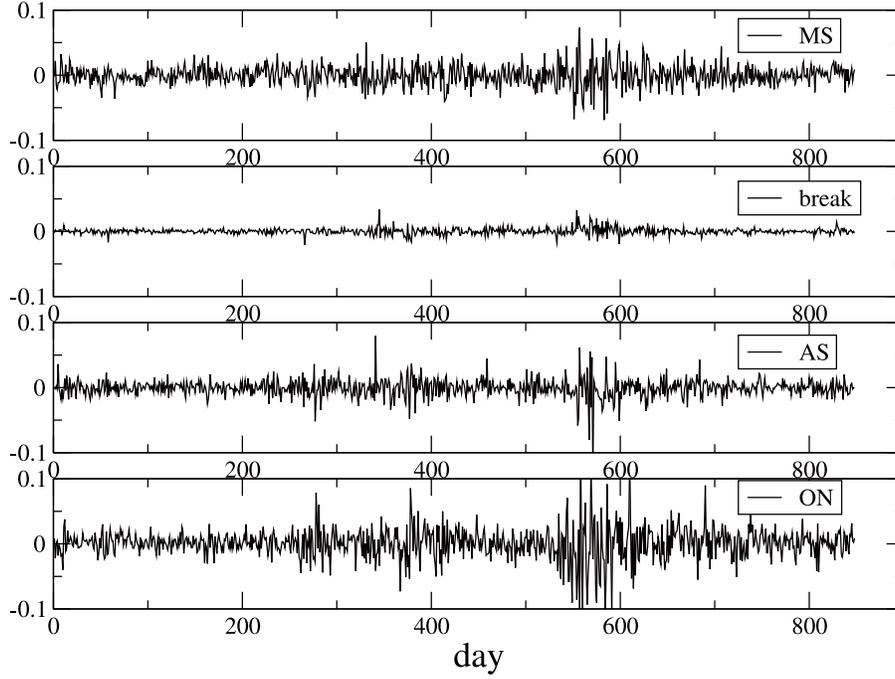


Fig. 1. Return time series in the different time zones for Mitsubishi Co. as representative series.

opening and closing prices of the corresponding zone. For instance $R_{MS,t}$ is given by $\ln P_{MS,t}^o - \ln P_{MS,t}^c$, where $P_{MS,t}^o$ is the opening price of the morning session on the day t and $P_{MS,t}^c$ the closing price of the morning session on the day t . In a similar manner $R_{AS,t}$ is given by $\ln P_{AS,t}^o - \ln P_{AS,t}^c$. Returns in the lunch break are calculated by $R_{break,t} = \ln P_{MS,t}^c - \ln P_{AS,t}^o$. $R_{ON,t}$ is the overnight return given by $\ln P_{MS,t}^o - \ln P_{AS,t-1}^c$. We see that the magnitude of the returns in the lunch break is very small which means that the price change in this zone is small. On the other hand in the overnight zone the magnitude of the returns is big as well as in the trading zones.

We focus on two realized volatilities which are constructed using data from different trading zones. Let us denote $RV_{MS,t}$ ($RV_{AS,t}$) as the realized volatility calculated using data in the MS (AS). For instance $RV_{MS,t}$ is defined by

$$RV_{MS,t} = \sum_{i=1}^n r_{t_{MS}+i\Delta_{MS}}^2, \quad (3.1)$$

where n and Δ_{MS} stand for the sampling number and sampling period respectively, and the relation between n and Δ_{MS} is given by $\Delta_{MS} = h_{MS}/n$, where h_{MS} is the trading time of the MS, i.e. 120 min. at the Tokyo stock exchange market. t_{MS} is the opening time of the MS, i.e. 9:00. $RV_{AS,t}$ is also defined in a similar manner, e.g. $h_{AS} = 150$ min. and $t_{AS} = 12:30$.

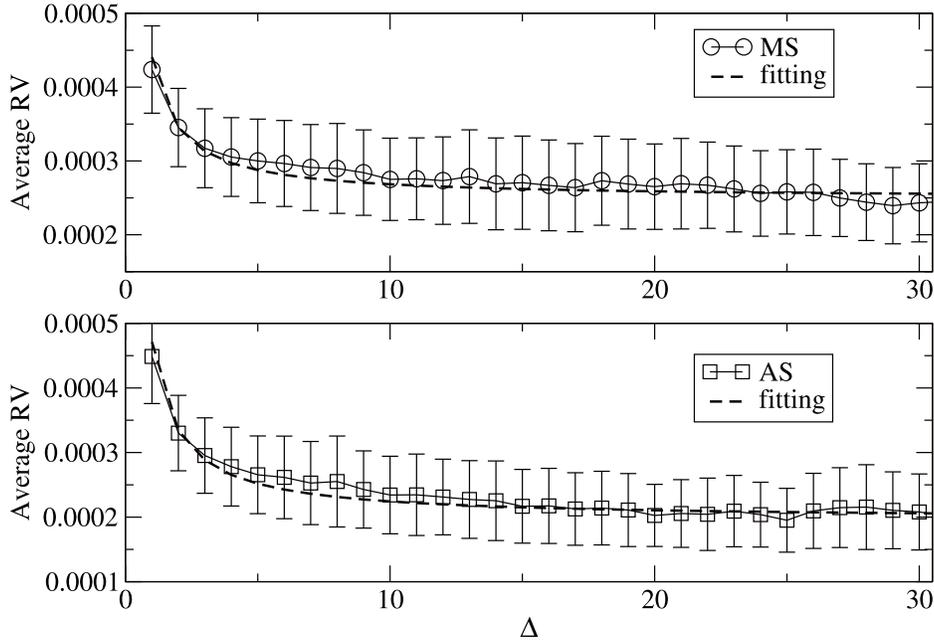


Fig. 2. Average realized volatility in the MS (top) and AS (bottom) for Mitsubishi Co. as a function of sampling period Δ (min.). The dotted lines show the fitting results to the expected formula of $a_0(1 + a_1/\Delta)$.

§4. Microstructure noise

In order to quantify the microstructure noise we measure the realized volatility at various sampling frequencies and make “volatility signature plot”³⁸⁾ to visualize the bias effect from the microstructure noise. Figure 2 shows the volatility signature plot for Mitsubishi Co. as a representative one. The top (bottom) panel of the figure shows the average realized volatility in the MS (AS) as a function of sampling period Δ . As seen in Fig. 2 due to the microstructure noise the average realized volatility increases with increasing of the sampling frequency (with decreasing of the sampling period Δ). We also find quantitatively similar results for volatility signature plots of other stocks. The average realized volatility is well fitted to a functional form of $a_0(1 + a_1/\Delta)$ which is an expected from Eq. (2·8). The dotted lines in Fig. 2 show the fitting results. The true average realized volatility can be evaluated to be a_0 .

The value of $\delta(\Delta) = a_1/\Delta$ corresponds to the bias from the microstructure noise at the sampling period Δ . The fitted parameters a_0, a_1 and the bias values at $\Delta = 5$ min. are summarized in Table I. It is found that the bias in the afternoon session is larger than the one in the morning session. It seems that the average realized volatility stabilizes around 5–10 min. within error bar. However if we compare the bias to the true average volatility which is evaluated as a_1/Δ the bias at 5 min. still contributes largely. Especially in the afternoon session, the bias at 5 min. proceeds 30% for most stocks examined here. These biases result in decreasing the variance of standardized returns. Later we incorporate these biases to calculate the variance of standardized returns.

Table I. Fitted parameters and bias at 5 min. a_0 corresponds to the average RV in the limit of $n \rightarrow \infty$. The bias at $\Delta = 5$ min. is calculated by $\delta = a_1/\Delta$.

Stock		1: Mitsubishi	2: Nomura	3: Nippon St.	4: Panasonic	5: SONY
MS	a_0	2.5×10^{-4}	2.5×10^{-4}	2.0×10^{-4}	1.6×10^{-4}	1.7×10^{-4}
	a_1	0.705	0.366	1.09	0.643	0.780
	$\delta(5 \text{ min.})$	0.141	0.073	0.228	0.126	0.156
AS	a_0	1.8×10^{-4}	1.8×10^{-4}	1.7×10^{-4}	1.1×10^{-4}	1.2×10^{-4}
	a_1	1.58	0.897	1.88	1.59	1.80
	$\delta(5 \text{ min.})$	0.316	0.179	0.376	0.318	0.360

§5. Standard normality of standardized returns

From the MDH return time series, $R_{MS,t}$ and $R_{AS,t}$ are expected to be

$$R_{MS,t} = \sigma_{MS,t}\epsilon_t, \quad (5.1)$$

and

$$R_{AS,t} = \sigma_{AS,t}\epsilon_t, \quad (5.2)$$

respectively. $\sigma_{MS,t}^2$ ($\sigma_{AS,t}^2$) is an integrated volatility in the morning (afternoon) session. ϵ_t is an independent Gaussian random variable with mean 0 and variance 1. Substituting realized volatility for the integrated volatility, i.e. $\sigma_{MS,t} = RV_{MS,t}^{1/2}$ and $\sigma_{AS,t} = RV_{AS,t}^{1/2}$ we expect that returns standardized by those standard deviation exhibit Gaussian-distributed time series.

Table II (III) shows the standard deviations and kurtoses of original and standardized returns in the MS (AS). We find that the kurtoses for the original returns are very high, compared to 3 of Gaussian random variables. On the other hand the kurtoses of the standardized returns come near 3. However it is observed that in the MS all kurtoses are slightly smaller than three. At the present moment we do not know what causes this slight difference between the MS and AS. We also tested the normality by the Anderson-Darling test and the normality was not rejected (p-values are listed in Tables).

We find that the standard deviation of the standardized returns also approaches one. However in most cases the values of the standard deviation are slightly less than one. This difference could be explained by the bias which still remains at 5 min. sampling frequency. In Tables II and III we also list the bias-corrected standard deviation. Let σ be the standard deviation of the standardized returns. The bias-corrected standard deviation is given by $\sigma * (1 + \delta(5 \text{ min.}))^{1/2}$, where $\delta(5 \text{ min.})$ is the bias at 5 min. sampling frequency listed in Table I. It is evident that the bias-corrected standard deviation comes more close to one. Although the values of the bias-corrected standard deviation in the AS for some stocks appear to be slightly less than 1, other results turns out to be satisfactory very close to one, compared to the average value of 0.8 which is obtained from the standardized returns of the DJIA stocks.²⁴⁾

As stylized facts of asset returns it is well known that while linear autocorrelations of returns are not significant except for very small time scale the autocorre-

Table II. Standard deviation and kurtosis of original and standardized returns in the morning session. The values of parentheses show statistical errors estimated by the Jackknife method. AD stands for the Anderson-Darling normality test.

Stock		1: Mitsubishi	2: Nomura	3: Nippon St.	4: Panasonic	5: Sony
R_t	std.dv. ($\times 10^2$)	1.58(20)	1.58(16)	1.53(17)	1.28(13)	1.36(14)
	kurt.	4.99(50)	4.90(71)	6.10(100)	5.53(94)	4.71(39)
$R_t/RV_t^{1/2}$	std.dv. ($= \sigma$)	0.915(32)	0.951(20)	0.909(30)	0.895(26)	0.921(26)
	kurt.	2.75(13)	2.68(11)	2.79(13)	2.66(12)	2.66(12)
	$\sigma * (1 + \delta(5 \text{ min}))^{1/2}$	0.977(35)	0.985(31)	1.007(33)	0.950(28)	0.990(28)
AD	p-value	0.310	0.208	0.115	0.065	0.085

Table III. Same as in Table II but in the afternoon session.

Stock		1: Mitsubishi	2: Nomura	3: Nippon St.	4: Panasonic	5: Sony
R_t	std.dv. ($\times 10^2$)	1.40(22)	1.34(15)	1.50(24)	1.12(16)	1.25(21)
	kurt.	17.3(85)	8.66(25)	24.8(144)	11.9(36)	18.3(61)
$R_t/RV_t^{1/2}$	std.dv. ($= \sigma$)	0.787(27)	0.825(47)	0.837(36)	0.795(33)	0.820(41)
	kurt.	3.18(20)	3.21(15)	3.20(20)	2.91(14)	3.16(18)
	$\sigma * (1 + \delta(5 \text{ min}))^{1/2}$	0.903(31)	0.896(52)	0.982(42)	0.913(38)	0.957(48)
AD	p-value	0.159	0.063	0.087	0.266	0.196

lations of absolute returns decay very slowly. On the other hand the standardized returns are expected to be independent Gaussian variables and thus not only the standardized returns but also absolute standardized returns should show insignificant autocorrelation. Figure 3 compares the autocorrelation function (ACF) between absolute returns and absolute standardized returns.^{*)} The top (bottom) panel shows the ACF in the MS (AS). The solid lines in the figure show 95% confidence limits. We see that while the autocorrelation function of absolute returns decays very slowly the autocorrelation function of absolute standardized returns immediately disappears in the noise level of 95% confidence band. These features also support the view of the MDH.

§6. Conclusion

We constructed two realized volatilities $RV_{MS,t}$ and $RV_{AS,t}$ defined in the two trading sessions of the Tokyo Stock Exchange. Using the two realized volatilities we investigated standardized returns in the MS and AS separately and examine properties of standard deviation 1 and kurtosis 3 under the MDH. By calculating the realized volatilities at various sampling frequencies we quantified the bias from the microstructure noise as a function of the sampling frequency. Taking account of the bias effect we find that the bias-corrected standard deviations of standardized returns in the MS and AS come close to one. Furthermore we also find that the values of the kurtosis in the MS and AS come close to three. In Ref. 25) standardized daily returns on the Tokyo Stock Exchange have been examined and the normality of the

^{*)} We have also verified that the standardized returns show no significant autocorrelation.

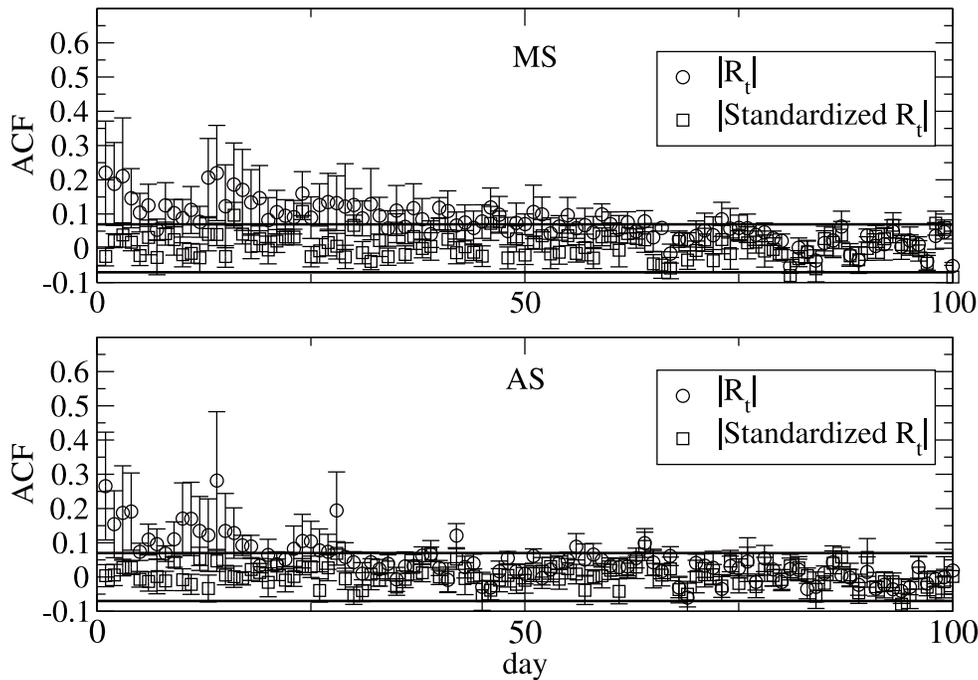


Fig. 3. Autocorrelation functions (ACF) of absolute returns and absolute standardized returns for Mitsubishi Co. The top (bottom) panel shows the ACF in the MS (AS). Solid lines in the figure show 95% confidence limits.

standardized daily returns is confirmed. However due to non-trading hours issue it was not confirmed that the standard deviation of the standardized daily returns comes to one. Our observations confirm that standardized returns in the MS and AS on the Tokyo Stock Exchange show both properties of standard deviation 1 and kurtosis 3, expected from the MDH. Thus we conclude that the price dynamics on the Tokyo Stock Exchange is consistent with the MDH.

While the kurtoses in the AS are consistent with three, we observe slightly smaller kurtoses in the MS. Although we do not understand this slight difference of the kurtoses between the MS and AS this might be caused by the trading duration time difference or trading activity difference between the MS and AS. We also observe slight deviation from one for standard deviations of standardized returns of some stocks. The slight deviation might indicate that there still remain other small biases not considered here. For instance the diffusion assumption of Eq. (2.1) might need an additional term such as the jump component.⁴³⁾ When the jump effect is present it might distort the standard normal property. In order to fully understand the price dynamics, in future it should be clarified whether such slight differences in the standard deviations and kurtoses actually indicate any important effects or not.

Under the MDH it is expected that not only the standardized returns but also the absolute standardized returns show no significant autocorrelation. We verified that the absolute standardized returns have no autocorrelations. These results also support the MDH for the price dynamics of the stocks on the Tokyo Stock Exchange.

Although our findings are consistent with the MDH, the MDH itself does not

explain the volatility dynamics which may account for other relevant properties such as volatility clustering and long-memory of absolute returns. In order to understand the complete price dynamics more studies needed, especially for volatility dynamics. In econophysics absolute returns are often used as a substitute for volatility measures which do not match the integrated volatility.^{45)–49)} Since the realized volatility calculated by using high-frequency data is an accurate measure of the integrated volatility it may serve as a tool for further studies of the volatility dynamics.

Acknowledgements

Numerical calculations in this work were carried out at the Yukawa Institute Computer Facility and the facilities of the Institute of Statistical Mathematics. This work was supported by a Grant-in-Aid for Scientific Research (C) (No. 22500267).

References

- 1) R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 2000).
- 2) J. Voit, *The Statistical Mechanics of Financial Markets* (Springer-Verlag, Heidelberg, 2001).
- 3) J. P. Bouchaud and M. Potters, *Theory of Financial Risks and Derivative Pricing: From Statistical Physics to Risk Management* (Cambridge University Press, Cambridge, 2003).
- 4) S. Sinha, A. Chatterjee, A. Chakraborti and B. K. Chakraborti, *Econophysics* (WILEY-VCH Verlag GmbH, 2011).
- 5) R. Cont, *Quantitative Finance* **1** (2001), 223.
- 6) B. Mandelbrot, *J. of Business* **36** (1963), 394.
- 7) E. F. Fama, *J. of Business* **38** (1965), 34.
- 8) R. N. Mantegna and H. E. Stanley, *Nature* **376** (1995), 46.
- 9) X. Gabaix, P. Gopikrishnan, V. Plerou and H. E. Stanley, *The Quarterly J. of Economics* **121** (2006), 461.
- 10) P. Gopikrishnan, M. Meyer, L. A. N. Amaral and H. E. Stanley, *Eur. Phys. J. B* **3** (1998), 139.
- 11) P. Gopikrishnan, V. Plerou, L. A. N. Amaral, M. Meyer and H. E. Stanley, *Phys. Rev. E* **60** (1999), 5305.
- 12) V. Plerou, P. Gopikrishnan, L. A. N. Amaral, M. Meyer and H. E. Stanley, *Phys. Rev. E* **60** (1999), 6519.
- 13) P. D. Praetz, *J. of Business* **45** (1972), 49.
- 14) C. Tsallis, C. Anteneodo, L. Borland and R. Osorio, *Physica A* **324** (2003), 89.
- 15) S. M. Duarte Queiros, L. G. Moyano, J. de Souza and C. Tsallis, *Eur. Phys. J. B* **55** (2007), 161.
- 16) G. F. Gu, W. Chen and W. X. Zhou, *Physica A* **387** (2008), 495.
- 17) A. Gerig, J. Vicente and M. A. Fuentes, *Phys. Rev. E* **80** (2009), 065102(R).
- 18) C. Tsallis, *J. Stat. Phys.* **52** (1988), 479.
- 19) L. G. Moyano, C. Tsallis and M. Gell-Mann, *Europhys. Lett.* **73** (2006), 813.
- 20) P. K. Clark, *Econometrica* **41** (1973), 135.
- 21) C. Beck and E. G. D. Cohen, *Physica A* **322** (2003), 267.
- 22) E. Van der Straeten and C. Beck, *Phys. Rev. E* **80** (2009), 036108.
- 23) T. G. Andersen, T. Bollerslev, F. X. Diebold and P. Labys, *J. of the American Statistical Association* **96** (2001), 42.
- 24) T. G. Andersen, T. Bollerslev, F. X. Diebold and H. Ebens, *J. of Financial Economics* **61** (2001), 43.
- 25) T. Takaishi, *Evolutionary and Institutional Economics Review* **7** (2010), 89.
- 26) G. E. Tauchen and M. Pitts, *Econometrica* **51** (1983), 485.
- 27) T. G. Andersen, *J. of Finance* **51** (1996), 169.
- 28) C. M. Jones, G. Kaul and M. L. Lipson, *Review of Financial Studies* **7** (1994), 631.

- 29) C. C. Chan and W. M. Fong, *J. of Banking & Finance* **30** (2006), 2063.
- 30) R. F. Engle, *Econometrica* **60** (1982), 987.
- 31) T. Bollerslev, *J. of Econometrics* **31** (1986), 307.
- 32) E. Jacquier et al., *J. of Business & Economic Statistics* **12** (1994), 371.
- 33) S. Kim, N. Shephard, S. Chib, *Review of Economic Studies* **65** (1998), 361.
- 34) N. Shephard and M. K. Pitt, *Biometrika* **84** (1997), 653.
- 35) T. G. Andersen and T. Bollerslev, *International Economic Review* **39** (1998), 885.
- 36) J. Y. Campbell, A. W. Lo and A. C. MacKinlay, *The Econometrics of Financial Markets* (Princeton University Press, 1997).
- 37) B. Zhou, *J. of Business & Economics Statistics* **14** (1996), 45.
- 38) T. G. Andersen, T. Bollerslev, F. X. Diebold and P. Labys, *Risk Magazine* **13** (2000), 105.
- 39) B. J. Blair, S.-H. Poon and S. J. Taylor, *J. of Econometrics* **105** (2001), 5.
- 40) S. J. Koopman, B. Jungbacker and E. Hol, *J. of Empirical Finance* **12** (2005), 445.
- 41) E. Ghysels, P. Santa-Clara and R. Valkanov, *J. of Econometrics* **131** (2006), 59.
- 42) P. R. Hansen and A. Lunde, *J. of Finance* **48** (2005), 1779.
- 43) T. G. Andersen, T. Bollerslev, F. X. Diebold and P. Labys, *Multinational Finance J.* **4** (2000), 159.
- 44) V. Plerou, P. Gopikrishnan, L. A. N. Amaral, X. Gabaix and H. E. Stanley, *Phys. Rev. E* **62** (2000), R3023.
- 45) P. Cizeau, Y. Liu, M. Meyer, C.-K. Peng and H. E. Stanley, *Physica A* **245** (1997), 441.
- 46) Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C. K. Peng and H. E. Stanley, *Phys. Rev. E* **60** (1999), 1390.
- 47) F. Wang, K. Yamasaki, S. Havlin and H. E. Stanley, *Phys. Rev. E* **73** (2006), 026117.
- 48) F. Ren, L. Guo and W. X. Zhou, *Physica A* **388** (2009), 881.
- 49) F. Wang, S. J. Shieh, S. Havlin and H. E. Stanley, *Phys. Rev. E* **79** (2009), 056109.

Observation of Frustrated Correlation Structure in a Well-Developed Financial Market

Takeo YOSHIKAWA,^{1,*} Takashi IINO² and Hiroshi IYETOMI²

¹*Graduate School of Science and Technology, Niigata University,
Niigata 950-2181, Japan*

²*Department of Physics, Niigata University, Niigata 950-2181, Japan*

We approach the correlation structure in the Tokyo Stock Exchange (TSE) market through a concept of community of network. To construct a network, the correlation matrix of stock price changes, purified by random matrix theory, is regarded as an adjacency matrix. The stock correlation network thus constructed has negatively weighted links as well as positively weighted links. By extracting groups in which stocks are mainly interconnected by positive links, we find that the stocks decomposed into four comoving groups forming communities, three of which are strongly anticorrelated to each other, and the remainder is comparatively neutral to the rest of the communities. The conflicting triangle relationship between communities may cause complicated behavior in a well-development market such as TSE. Additionally, it is observed that some industrial sectors form distinctive coherent groups and others are separated to competing communities.

§1. Introduction

Principal component analysis (PCA) is a standard tool to explore multivariable data. Financial correlation matrix is analyzed by PCA in the seminal works,^{1),2)} and these studies revealed correlation structures underlying stock markets. In particular, the Random Matrix Theory (RMT) was used to estimate how many principal components should be retained as being statistically significant; the RMT works as a null hypothesis. Information contained in the eigenvectors of the correlation matrix elucidated existence of collective motion of business sectors or groups in well-developed markets.³⁾⁻⁵⁾ Such clustering of stocks was visualized graphically and also modeled in various ways.

Complex systems has been actively studied from a network viewpoint since the epoch-making study by Watts & Strogatz.⁶⁾ Among these studies, community detection is a powerful tool for understanding complex structure of real networks. In this paper, we take a new way to delve into correlation in markets from an alternative point of view, that is, market structure as a network. For this, we first identify the noise of correlation matrix by using the RMT, in order to extract genuine correlation. The correlation matrix, purified by the RMT, is then regarded as an adjacency matrix to construct a stock correlation matrix. We detect groups in which nodes are strongly correlated as communities in the network. However, the stock correlation network has both negative and positive links, because pairs of stocks can be anticorrelated. Recent works^{7),8)} provide us with a theoretical framework for detecting communities in networks with negative links as well as positive links. Preliminary results of the present work have been reported in Ref. 9).

*) E-mail: tko@phys.sc.niigata-u.ac.jp

§2. Stock correlations

2.1. Correlation matrix

We analyzed the daily prices of $N = 557$ stocks belonging to the Tokyo Stock Exchange (TSE) for the 10-year period 1996–2006 ($T = 2,706$ daily returns). The correlation matrix \mathbf{C} is calculated as

$$\mathbf{C} = \frac{1}{T} \mathbf{G} \mathbf{G}^T, \quad (2.1)$$

and its elements are

$$C_{ij} = \frac{1}{T} \sum_{t=1}^T G_{i,t} G_{j,t}, \quad (i, j = 1, \dots, N) \quad (2.2)$$

where \mathbf{G} is an $N \times T$ matrix, and the time series $\{G_{i,t}\}_{t=1, \dots, T}$ are standardized log-return of the price change of the stock i at time t .

2.2. Random matrix theory

As a null hypothesis for filtering the correlation matrix, we consider a *random correlation matrix* given as

$$\mathbf{R} = \frac{1}{T} \mathbf{H} \mathbf{H}^T, \quad (2.3)$$

where \mathbf{H} is an $N \times T$ matrix composed by N time series of random variables with zero mean and unit variance of length T ; its elements are totally independent. In the limit $N, T \rightarrow \infty$ with fixed $Q \equiv T/N$, the probability density function $\rho(\lambda)$ of eigenvalue λ of the random correlation matrix \mathbf{R} is given by

$$\begin{cases} \rho(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \\ \lambda_{\pm} = \left(1 \pm \frac{1}{\sqrt{Q}}\right)^2, \end{cases} \quad (2.4)$$

where the eigenvalues are bounded by $[\lambda_-, \lambda_+]$.

2.3. Genuine correlations

The correlation matrix may be decomposed^(2),3),10) into

$$\begin{aligned} \mathbf{C} &= \mathbf{C}_{\text{market}} + \mathbf{C}_{\text{group}} + \mathbf{C}_{\text{random}} \\ &= \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \sum_{i=2}^{13} \lambda_i \mathbf{u}_i \mathbf{u}_i^T + \sum_{i=14}^{557} \lambda_i \mathbf{u}_i \mathbf{u}_i^T, \end{aligned} \quad (2.5)$$

where λ_i 's are the eigenvalues of \mathbf{C} sorted in descending order and \mathbf{u}_i 's are the corresponding eigenvectors. If there were no correlation between stock prices, the eigenvalues shall satisfy the distribution of $\rho(\lambda)$. Hence, one cannot distinguish fluctuations ascribed to \mathbf{u}_i with $\lambda_i \leq \lambda_+$ from noise. In the TSE data, the value of $Q \simeq 4.86$,

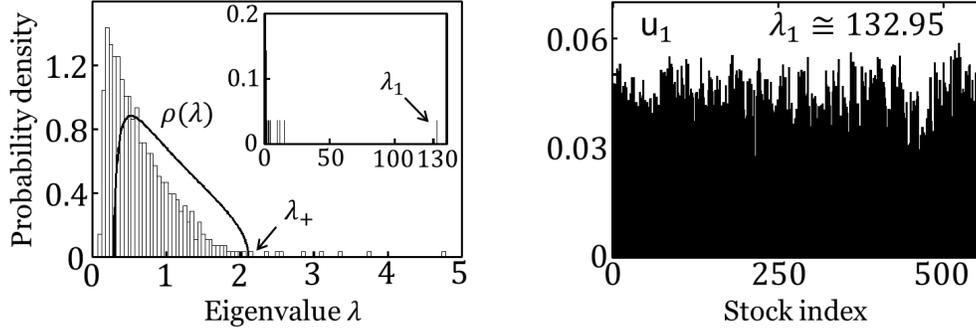


Fig. 1. (Left) Probability density of the eigenvalues of the correlation matrix \mathbf{C} compared with that of the random correlation matrix \mathbf{R} , $\rho(\lambda)$, given by Eq. (2.4). The inset shows the location of the largest eigenvalue λ_1 . (Right) Components of the eigenvector \mathbf{u}_1 associated with λ_1 .

and we obtained $\lambda_+ \simeq 2.11$. The 13th and 14th eigenvalues are $\lambda_{13} > \lambda_+ > \lambda_{14}$. Therefore, the interaction $\sum_{i=14}^{557} \lambda_i \mathbf{u}_i \mathbf{u}_i^T$ is regarded as being random (Fig. 1 (left)). Furthermore, $\lambda_1 (\simeq 132.95)$ is eight times or more as large as $\lambda_2 (\simeq 15.48)$, and the components of \mathbf{u}_1 are all positive (see the right panel of Fig. 1). For these reasons, $\lambda_1 \mathbf{u}_1 \mathbf{u}_1^T$ indicates the market mode.

Now, we adopt $\mathbf{C}_{\text{group}}^{(l)}$ for the adjacency matrix to construct a stock correlation network. We exclude $\mathbf{C}_{\text{market}}$ because it just describes a collective motion of the whole market. To see how much each $\lambda_i \mathbf{u}_i \mathbf{u}_i^T$ ($i = 2, 3, \dots, 13$) contributes to $\mathbf{C}_{\text{group}}$, we define $\mathbf{C}_{\text{group}}^{(l)}$ as

$$\mathbf{C}_{\text{group}}^{(l)} = \sum_{i=2}^l \lambda_i \mathbf{u}_i \mathbf{u}_i^T, \quad l = 2, 3, \dots, 13. \quad (2.6)$$

2.4. Collective behavior of sectors

Collective motion of some sectors appears in the eigenvectors of large eigenvalues.³⁾⁻⁵⁾ To observe this behavior of sectors, we calculated the polarization of stocks in each sector defined by

$$P_s \equiv \frac{\sum \mathbf{u}_i(j)}{\sum |\mathbf{u}_i(j)|}, \quad (2.7)$$

where $\mathbf{u}_i(j)$ represents the j -th component of the \mathbf{u}_i . And \sum sums up $\mathbf{u}_i(j)$ in each sector separately. By definition, $-1 \leq P_s \leq 1$. If a sector is comoving in an eigenmode, absolute value $|P_s|$ of the polarization of the sector is closer to one. As observed in Table I, roughly half of the sectors form groups of comoving stocks in each of the three eigenvectors. These behaviors manifested in \mathbf{u}_2 , \mathbf{u}_3 and \mathbf{u}_4 just confirm the results in Ref. 4).

Table I. Polarization of stocks in u_2 , u_3 and u_4 . The figure in the parenthesis attached to each sector is the number of stocks belonging to it. Values of the polarization with $|P_s| \geq 2/3$ are highlighted in **boldface**.

Sector	mode 2	mode 3	mode 4
Electric Appliances (72)	-0.97	-0.21	0.92
Chemicals (60)	0.10	0.19	0.04
Machinery (54)	-0.34	0.61	0.47
Foods (32)	1.00	-0.46	-0.12
Construction (31)	0.71	0.82	-0.80
Transportation Equipment (31)	-0.14	-0.29	-0.39
Wholesale Trade (21)	0.25	0.71	0.13
Banks (20)	-0.24	-1.00	-1.00
Iron & Steel (20)	0.45	0.99	0.35
Retail Trade (17)	-0.47	0.18	-0.73
Textiles & Apparels (17)	0.94	0.27	-0.27
Land Transportation (17)	0.81	-0.97	-0.82
Glass & Ceramics Products (15)	0.68	-0.36	0.98
Other Products (14)	-0.51	0.00	-0.47
Nonferrous Metals (14)	-0.04	0.22	0.98
Pharmaceutical (13)	0.62	-0.61	-0.62
Precision Instruments (13)	-0.99	-0.08	0.71
Electric Power & Gas (13)	1.00	-1.00	-1.00
Information & Communication (10)	-0.90	-1.00	0.63
Other Financing Business (9)	-0.61	-0.52	-0.56
Metal Products (8)	0.48	0.08	-0.17
Rubber Products (7)	0.47	-0.03	-0.22
Services (6)	0.04	0.04	-0.35
Securities & Commodity Futures (6)	-1.00	0.95	1.00
Real Estate (6)	0.98	0.45	0.47
Insurance (6)	0.11	-0.97	-1.00
Marine Transportation (5)	1.00	0.30	0.49
Warehousing & Harbor Transportation Services (5)	1.00	-0.37	-0.12
Pulp & Paper (5)	1.00	-0.37	-0.36
Fishery, Agriculture & Forestry (4)	1.00	-1.00	0.66
Oil & Coal Products (3)	1.00	-0.43	-0.31
Mining (2)	1.00	1.00	0.27
Air Transportation (1)	1.00	-1.00	-1.00

§3. Community detection

We suppose that the network is partitioned into communities; the community to which node i belong is designated by σ_i . A configuration of the community assignment is thus represented by a set $\{\sigma\}$ of σ_i 's. At a given assignment $\{\sigma\}$, there exist correlation between communities and anticorrelation within communities. Such a situation is referred to as frustration, and it is measured⁸⁾ by

$$F(\{\sigma\}) = - \sum_{ij} A_{ij} \delta(\sigma_i, \sigma_j), \quad (3.1)$$

Table II. Decomposition of stocks into communities. The number of stocks in each community is shown at varied l .

l	Comm. 1	Comm. 2	Comm. 3	Comm. 4
2	279	278	—	—
3	199	192	166	—
4	175	149	119	114
5	175	149	119	114
6	177	149	117	114
7	179	148	117	113
8	179	150	117	111
9	178	146	118	115
10	178	145	119	115
11	173	146	123	115
12	174	146	122	115
13	175	148	118	116
557	177	148	117	115

where $\delta(i, j)$ is the Kronecker delta. Minimizing the frustration $F(\{\sigma\})$ thus means maximizing the number of positive links or the sum of positive weights within communities. The frustration is peculiar to networks with both positive and negative links, so that it does not have its counterpart for networks with only positive links.

§4. Correlation structure in the stock network

We carried out community detection in the stock correlation network within the framework of frustration. We used a simulated annealing method¹¹⁾ to minimize the frustration $F(\{\sigma\})$ for the networks determined by $C_{\text{group}}^{(l)}$.

4.1. Evolution of the community decomposition

Table II shows that complexity of the community structure is enhanced with increasing l . The eigenvector u_2 perfectly separates the stocks into two groups. Stocks within their own communities move collectively and stocks belonging to different communities move oppositely. Then inclusion of u_3 decomposes the two groups into three, which strongly competes against each other as will be shown later. Inclusion of u_4 further decomposes the three groups into four. Remarkably, evolution of the community decomposition ceases at the level of $l = 4$. It has been confirmed that details of the four communities remained almost the same beyond $l = 4$. Furthermore, almost the same structure is obtained from adjacency matrix $\sum_{i=2}^{557} \lambda_i u_i u_i^T$. This fact indicates that the community structure disclosed here is a very fundamental property of the market.

4.2. Polarization of correlation

To elucidate the correlation structure obtained by the frustration optimization, in the same way in §2.4, we calculated the polarization of correlation defined by

$$P_c = \frac{\sum A_{ij}}{\sum |A_{ij}|}, \quad (4.1)$$

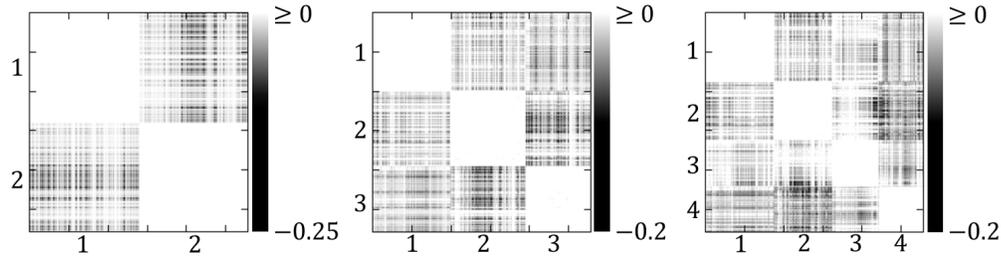


Fig. 2. Pictorial representation of $C_{\text{group}}^{(2)}$ (left), $C_{\text{group}}^{(3)}$ (center) and $C_{\text{group}}^{(4)}$ (right). Anticorrelation is emphasized. Stocks are classified according to the community assignment at each l .

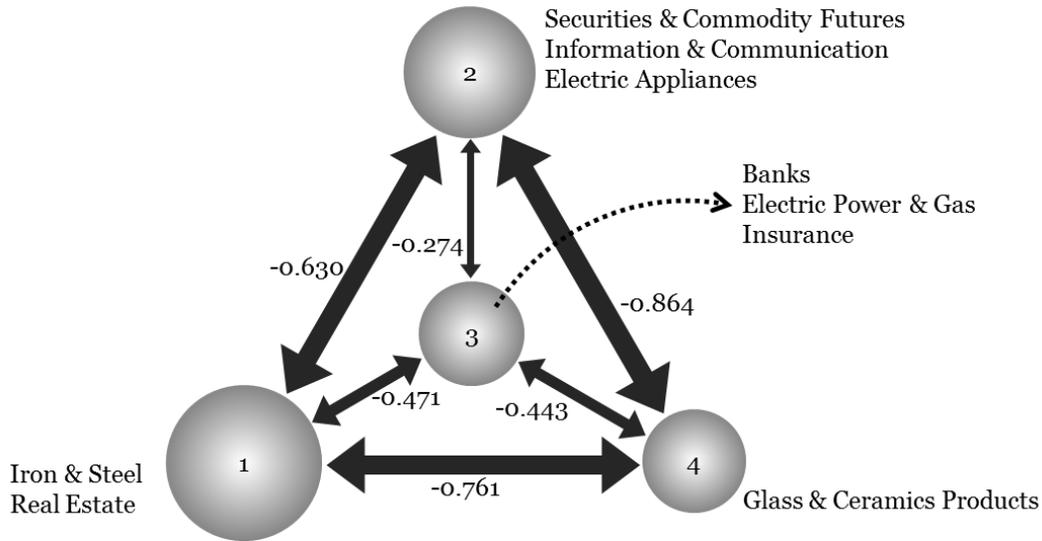


Fig. 3. Correlation structure in communities detected at $l = 4$. The arrows represent the anticorrelation between communities, and the width is proportional to P_c . Circles represent communities with their size proportional to the number of stocks in it.

where \sum sums up weights of links within community or weights of links between communities. Figure 2 shows that all of the communities are full of positive weights and almost perfectly exclusive of negative weights. In return, the communities are interconnected mainly with negative links. The new group (Comm. 3) formed at the level of $l = 4$ is composed of stocks spilled over from the original three groups at $l = 3$ and takes a rather neutral position against the remains that are strongly frustrated with each other (Fig. 3). We thus find that there exists strong frustration in the stock well-developed market.

4.3. Sectors of strongly correlated stocks

We pay attention to only the sectors consisting of at least five stocks, because the stock correlation network has four communities. Table III shows how stocks are distributed over the communities for each sector. As shown in Fig. 3, these are nine sectors of all stocks of which gather together in one community more than $2/3$. As we observed in §2.4, these sectors are composed of strongly correlated stocks in the

Table III. Decomposition of sectors through formation of communities. Relative distributions of stocks in the communities at $l = 4$ are listed with those over 2/3 highlighted in **boldface**.

Sector	Comm. 1	Comm. 2	Comm. 3	Comm. 4
Electric Appliances (72)	0.18	0.79	—	0.03
Chemicals (60)	0.43	0.13	0.18	0.25
Machinery (54)	0.44	0.39	0.02	0.15
Foods (32)	0.28	—	0.28	0.44
Construction (31)	0.48	0.07	0.23	0.23
Transportation Equipment (31)	0.32	0.29	0.20	0.20
Wholesale Trade (21)	0.62	0.19	0.10	0.10
Banks (20)	—	—	1.00	—
Iron & Steel (20)	0.70	0.05	0.10	0.15
Retail Trade (17)	0.24	0.18	0.53	0.06
Textiles & Apparels (17)	0.35	—	0.29	0.35
Land Transportation (17)	—	0.06	0.53	0.41
Glass & Ceramics Products (15)	0.13	0.20	—	0.66
Other Products (14)	0.43	0.29	0.21	0.07
Nonferrous Metals (14)	0.36	0.29	—	0.36
Pharmaceutical (13)	0.23	—	0.54	0.23
Precision Instruments (13)	0.39	0.62	—	—
Electric Power & Gas (13)	—	—	1.00	—
Information & Communication (10)	—	0.90	—	0.10
Other Financing Business (9)	0.11	0.56	0.11	0.22
Metal Products (8)	0.25	0.13	0.38	0.25
Rubber Products (7)	0.43	0.29	—	0.29
Services (6)	0.50	0.17	0.17	0.17
Securities & Commodity Futures (6)	—	1.00	—	—
Real Estate (6)	0.67	—	—	0.33
Insurance (6)	—	—	1.00	—
Marine Transportation (5)	0.40	—	—	0.60
Warehousing & Harbor Transportation Services (5)	0.20	—	0.40	0.40
Pulp & Paper (5)	0.20	—	0.20	0.60
Fishery, Agriculture & Forestry (4)	—	—	0.25	0.75
Oil & Coal Products (3)	0.33	—	—	0.67
Mining (2)	1.00	—	—	—
Air Transportation (1)	—	—	—	1.00

eigenvectors.

Each sector of the strongly correlated stocks has its own characteristics. Electric Power & Gas, Securities & Commodity Futures, Construction and Land Transportation are sectors forming very tight groups in any eigenvectors of u_2 , u_3 and u_4 . As the result of community decomposition, all the stocks of Electric Power & Gas and Securities & Commodity Futures belong to the same communities, i.e., Comm. 3 and Comm. 2, respectively. This is a well-expected result. In contrast, Construction is separated into three communities, Comms. 1, 3 and 4, and Land Transportation is separated into two communities, Comms. 3 and 4.

The analysis at the level of eigenvectors just elucidates the formation of groups in the market, but gives no information on relations among them. By contrast, relations between communities are clear as the result of community detection. For example,

Table IV. Partitioning of stocks in the Machinery sector into the four communities as shown in Fig. 3.

Company Name	
Comm. 1	
Tsugami Corporation	Toshiba Machine Co., Ltd.
Howa Machinery, Ltd.	OKK Corporation
Hitachi Construction Machinery Co., Ltd.	Kioritz Corporation
Kitagawa Iron Works Co., Ltd.	Mitsubishi Kakoki Kaisha, Ltd.
Tokyo Kikai Seisakusho, Ltd.	Aichi Corporation
Sumitomo Precision Products Co., Ltd.	Sakai Heavy Industries, Ltd.
Toyo Kanetsu K.K.	TCM Corporation
Nippon Conveyor Co., Ltd.	Kato Works Co., Ltd.
Yuken Kogyo Co., Ltd.	Tadano Ltd.
Juki Corporation	Janome Sewing Machine Co., Ltd.
Silver Seiko Ltd.	MAX Co., Ltd.
Riken Corporation	Kitz Corporation
Comm. 2	
Takuma Co., Ltd.	Amada Co., Ltd.
Makino Milling Machine Co., Ltd.	OSG Corporation
SMC Corporation	Oiles Corporation
Komatsu Ltd.	Komori Corporation
Ebara Corporation	Daikin Industries, Ltd.
Kurita Water Industries Ltd.	Tsubakimoto Chain Co.
Daifuku Co., Ltd.	Fujitec Co., Ltd.
Sankyo Co., Ltd.	Amano Corporation
Sanden Corporation	NSK Ltd.
THK Co., Ltd.	Hitachi Koki Co., Ltd.
Makita Corporation	
Comm. 3	
Kubota Corporation	
Comm. 4	
The Japan Steel Works, Ltd.	Sumitomo Heavy Industries, Ltd.
Iseki & Co., Ltd.	Nippon Piston Ring Co., Ltd.
NTN Corporation	Jtekt Corporation
Nachi-Fujikoshi Corp.	IHI Corporation

Iron & Steel and Real Estate belong to the same community (Comm. 1), i.e., these two sectors are strongly correlated. Similarly, Information & Communication and Securities & Commodity Futures (Comm. 2), Banks, Insurances and Electric Power & Gas (Comm. 3) are strongly correlated, as well. At the end of this section, we note that stocks in Machinery sector are divided into rather evenly three communities forming the conflicting the triangle as listed in Table IV.

§5. Summary

We revisited correlation structure in the Tokyo Stock Exchange market from a network viewpoint. The correlation matrix of stock price changes, purified by the RMT, was used to construct a network. The stock network thus constructed has

weighted links and furthermore those weights can be even negative. The present study has uncovered that the market had four communities consisting of strongly correlated stocks through minimization of the frustration.

Note that the correlation structure determined the only three eigenvectors u_2 , u_3 and u_4 . By considering a stock market as a network, it becomes easier for us to understand the relations between stocks. Additionally, we can see the shape of the whole market in this way. The stock prices within the groups comove almost perfectly. In contrast, three of the four groups are in a strongly anticorrelated state. Such a frustrated triangle relationship among the groups may give rise to complicated market structure.

Acknowledgements

The authors thank the Yukawa Institute for Theoretical Physics at Kyoto University, where this work was completed during the YITP workshop (YITP-W-11-04) entitled “Econophysics 2011 — The Hitchhikers’s Guide to the Economy”. This work was also supported, in part, through *the Program for Promoting Methodological Innovation in Humanities and Social Sciences by Cross-Disciplinary Fusing* and a Grant-in-Aid for Scientific Research (B), No. 22300080 (2010-12), provided by the Japan Society for the Promotion of Science and by the Ministry of Education, Culture, Sports, Science and Technology, respectively.

References

- 1) L. Laloux, P. Cizeau, J.-P. Bouchaud and M. Potters, Phys. Rev. Lett. **83** (1999), 1467.
- 2) V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral and H. E. Stanley, Phys. Rev. Lett. **83** (1999), 1471.
- 3) V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral and H. E. Stanley, Phys. Rev. E **65** (2002), 066126.
- 4) A. Utsugi, K. Ino and M. Oshikawa, Phys. Rev. E **70** (2004), 026110.
- 5) R. K. Pan and S. Sinha, Phys. Rev. E **76** (2007), 046116.
- 6) D. J. Watts and S. H. Strogatz, Nature **393** (1998), 440.
- 7) S. Gómez, P. Jensen and A. Arenas, Phys. Rev. E **80** (2009), 016114.
- 8) V. A. Traag and J. Bruggeman, Phys. Rev. E **80** (2009), 036115.
- 9) T. Yoshikawa, T. Iino and H. Iyetomi, *Proc. of the 3rd Int’l Conf. on Intelligent Decision Technologies* (2011), p. 511.
- 10) D. H. Kim and H. Jeong, Phys. Rev. E **72** (2005), 046133.
- 11) S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, Science **220** (1983), 671.

Revenue Prediction of a Local Event Using the Mathematical Model of Hit Phenomena

Akira ISHII, Takehiro MATSUMOTO and Shinji MIKI

*Department of Applied Mathematics and Physics, Tottori University,
Tottori 680-8552, Japan*

We propose a theoretical approach to investigate human-human interaction in the society, which uses a many-body theory that incorporates human-human interaction. We treat advertisement as an external force, and include the word of mouth (WOM) effect as a two-body interaction between humans and the rumor effect as a three-body interaction among humans. The parameters to define the strength of human interactions are assumed to be constant values. The calculated result explained well the two local events “Mizuki-Shigeru Road in Sakaiminato” and “the sculpture festival at Tottori” in Japan.

§1. Introduction

Human interaction in the real society can be considered as a many-body one. In particular, after the proliferation of social network systems (SNSes) such as blogs, Twitter, Facebook, and Google+, the interactions between two humans can be stocked as digital data. Though online social networks are not equivalent to the real world, we can assume that communication in SNSes is very similar to that in the real society. Thus, we can use the huge amount of digital data as observation data for the real society.^{1)–4)} Using this observation, we can apply the method of statistical mechanics to social sciences. Since word-of-mouth (WOM) has been pointed out to be very significant, for example, in marketing science,^{5)–8)} such analysis and prediction of digital WOM in the sense of statistical physics is important today.

Previous works^{9)–14)} applied statistical mechanics to human dynamics. In particular, Ishii focused on the Japanese motion picture entertainment industry. In this study, we consider the hit phenomena for movies both experimentally and theoretically. For an experimental viewpoint, we observe daily revenue data and daily blog-mentions for many movies. We also obtain the advertisement cost for the movies. For a theoretical viewpoint, we present a mathematical model for hit phenomena as non-equilibrium, nonlinear, dynamical phenomena. We compare the simulated revenue with the observed revenue to check the model. The calculations for movie revenues presented in previous works^{9)–14)} confirm that our model explains the movies’ revenues and blog-mentions very well.

In this paper, we apply this theory to a local event in Japan to show its applicability not only in the motion picture entertainment industry but also in many other economic fields in the real society.

§2. Theory

2.1. Purchase intention for individual person

Based on the observation of the hit phenomena in the Japanese market, we present a theory to explain and predict hit phenomena. First, instead of the number of visitors $N(t)$, we introduce the integrated purchase intention of an individual customer $J_i(t)$ defined as follows:

$$N(t) = \sum_i J_i(t), \tag{2.1}$$

where the suffix j corresponds to an individual who wishes to go to a local event. $J_i(t)$ is considered here as the sum of the purchase intention of person j from the past to time t .

The daily purchase intention $I_i(t)$ is defined from $J_i(t)$ as follows:

$$\frac{dJ_i(t)}{dt} = I_i(t). \tag{2.2}$$

The total number of visitors can be calculated using the purchase intention as follows:

$$N(t) = \int_0^t \sum_i I_i(\tau) d\tau. \tag{2.3}$$

Since the purchase intention of the individual customer increases due to both advertisement and communication with other people, we construct a mathematical model for hit phenomena as follows:

$$\frac{dI_i(t)}{dt} = \text{advertisement}(t) + \text{communication}(t). \tag{2.4}$$

Usually, rush effects are observed in finite-duration events. Rush effects are also observed in the registration or submission for an academic conference. The local event considered in this paper also has a finite duration. As such, we use the external function $f(t)$ to explain the rush effect. Thus, in this case, the mathematical model for a local event is

$$\frac{dI_i(t)}{dt} = (\text{advertisement}(t) + \text{communication}(t))f(t). \tag{2.5}$$

The function $f(t)$ is assumed to be unity except in the last part of the duration. We use the external function $f(t)$ in the final part of the chapter.

2.2. Advertisement

Advertisement is a very important factor to increase the purchase intention of a customer in the market. Usually, an advertisement campaign is done on TV, print, and other media. We consider the advertisement effect as an external force term $A(t)$ on the purchase intention as follows:

$$\frac{dI_i(t)}{dt} = A(t) + \sum_j D_{ij} I_j(t) + \sum_j \sum_k P_{ijk} I_j(t) I_k(t), \tag{2.6}$$

where D_{ij} is the factor for direct communication and P_{ijk} is the factor for indirect communication. Because of the indirect communication term, this equation is a nonlinear equation.

2.3. Mean field approximation

To solve Eq. (2.6), we introduce the mean field approximation for simplicity. Namely, we assume that every person moves equally so that we can introduce the averaged value of individual purchase intention.

$$I = \frac{1}{N_p} \sum_j I_j(t), \quad (2.7)$$

where the number of potential incoming persons is N_p . We obtain the direct communication term from the persons who do not attend the event as follows:

$$\sum_j D_{ij} I_j(t) = N_p \frac{1}{N_p} \sum_j D_{ij} I_j(t) \Rightarrow \frac{N_p - N(t)}{N_p} (N_p - N(t)) D^{nn} I, \quad (2.8)$$

where D^{nn} is the factor of direct communication between the persons who do not attend the event at time t .

Similarly, we obtain the indirect communication term due to the communication between the persons who do not attend the event at time t :

$$\begin{aligned} \sum_j \sum_k P_{ijk} I_j(t) I_k(t) &= N_p \frac{1}{N_p} \sum_j N_p \frac{1}{N_p} \sum_k P_{ijk} I_j(t) I_k(t) \\ &\Rightarrow \left(\frac{N_p - N(t)}{N_p} \right)^3 N_p^2 P^{nn} I^2 = \frac{(N_p - N(t))^3}{N_p} P^{nn} I^2, \end{aligned} \quad (2.9)$$

where P^{nn} is the factor of indirect communication between the persons who do not attend the event at time t .

Direct communication between the attending and non-attending persons can be written as

$$\sum_j D_{ij} I_j(t) = N_p \frac{1}{N_p} \sum_j D_{ij} I_j(t) \Rightarrow \frac{N(t)}{N_p} (N_p - N(t)) D^{ny} I, \quad (2.10)$$

where D^{ny} is the factor of direct communication between the attending and non-attending persons. For indirect communication, we obtain two more terms corresponding to indirect communication due to communication between the attending persons and communication between attending and non-attending persons as follows:

$$\frac{(N(t))^2 (N_p - N(t))}{N_p} P^{yy} I^2 + \frac{N(t) (N_p - N(t))^2}{N_p} P^{ny} I^2, \quad (2.11)$$

where P^{yy} is the factor of the indirect communication between the attended persons and P^{ny} is the factor of the indirect communication due to the communication between the attended person and the unattended person at the time t .

Finally, we obtain the equation of the mathematical model for hit phenomena within the mean field approximation:

$$\begin{aligned} \frac{I(t)}{dt} = & \left(A(t) + \frac{N(t)}{N_p}(N_p - N(t))D^{ny}I \right. \\ & + \frac{(N(t))^2(N_p - N(t))}{N_p}P^{yy}I^2 + \frac{N(t)(N_p - N(t))^2}{N_p}P^{ny}I^2 \left. \right) f(t) \\ & + \frac{N_p - N(t)}{N_p}(N_p - N(t))D^{nn}I + \frac{(N_p - N(t))^3}{N_p}P^{nn}I^2, \end{aligned} \quad (2.12)$$

where

$$N(t) = N_p \int_0^t I(\tau) d\tau \quad (2.13)$$

and $f(t)$ means the rush effect or the increasing number of attendees during the last day of the event, and where b is the parameter representing the difference between the number of visitors on a weekday and that on a weekend.

$$f(t) = c(t_{end} - t)^{-4}. \quad (2.14)$$

The power -4 is decided by adjusting our calculation to the observed data of the rush effect in the case of the World Sand Sculpture Festival 2009 in Tottori, Japan.

Equations (2.12) and (2.13) are based on the equation we presented in a previous work on the motion picture entertainment industry.^{11)–13)} This equation is, in principle, also derived in our recent paper¹⁴⁾ using the stochastic processes. Thus, the equation (2.12) with (2.13) is the nonlinear integrodifferential equation. However, since we deal with day-based data, the time difference is one day, and we can solve the equation numerically as a difference equation.

§3. Results

3.1. Target local event: Mizuki-Shigeru Road

Tottori prefecture is far from Tokyo and contains Sakaiminato city. Mr. Mizuki Shigeru, who was born in Sakaiminato city, produced many pieces of work based on the motif of Yokais (Demons), such as “Ge Ge Ge no Kitaro”. He was awarded the Shiju Hosho Decoration (Medal with the Purple Ribbon) in 1991, the Minister of Education Award of Japan in 1996, and the Kyokujitu Sho Decoration (Order of the Rising Sun) in 2003, for his excellent achievements and cultural contributions. In Sakaiminato city, there is Mizuki-Shigeru Road where 134 yokai characters were sterically casted into detailed bronze statues that are accessible to the public. These statues are regarded as fine examples of art, and are a major tourist draw.

For the past 10 years, the Tourism Section of the Trading and Tourism Division of the Industry Environment Department of Sakaiminato City has been in charge of setting advertisement costs and recording the daily number of incoming guests to the Mizuki-Shigaru Road. Therefore, we use these data as a test case. The daily number of blog-mentions for the Mizuki-Shigeru Road was collected using a

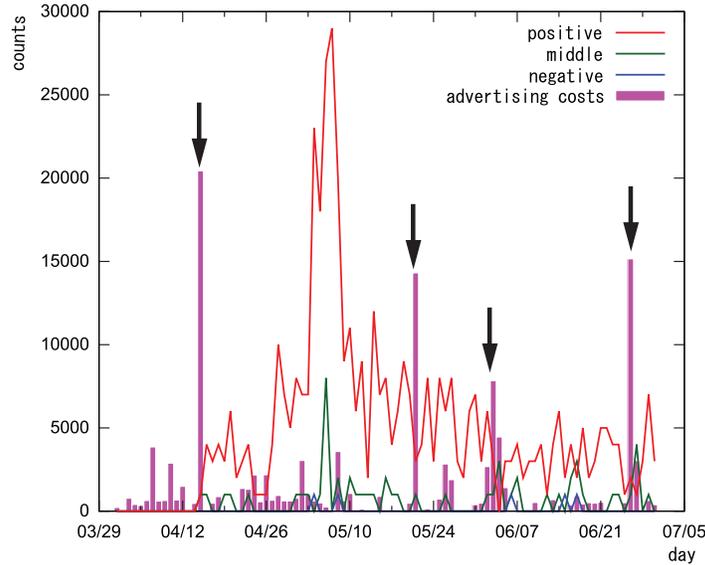


Fig. 1. Daily positive, neutral, and negative blog-mentions for the Sakaiminato Mizuki-Shigeru Road, and daily advertisement cost for the Mizuki-Shigeru Road in April–July, 2008. The horizontal axis denotes time.

website service called “Kizasi”. We found that the number of blog postings behaves similar to the revenue of a movie. The number of blog-mentions can thus be used as quasi-revenue.^{11)–14)}

First, we show the daily positive, neutral, and negative blog-mentions for the Sakaiminato Mizuki-Shigeru Road and the daily advertisement cost for the Mizuki-Shigeru Road in April–July, 2008 in Fig. 1. Here, a positive entry indicates an intent to revisit and a negative one indicates otherwise.

In Fig. 1, it should be noted that in Japan, the holiday season is from April 29 to May 5. Consequently, we witnessed a major peak in early May. We found that most entries are positive. As such, we can use the daily number of blog-mentions as purchase intention $I(t)$.

3.2. Decline of advertisement effect

Moreover, we found that the peaks of daily advertisement cost and of positive blog-mentions have shifted by one or two days. The arrows in Fig. 1 indicate the advertisement cost. There is a delay in the peak of blog-mentions. From these data for 2008 and 2009, we found the decline rate of the advertisement effect to be nearly 0.4/days, where effective advertisement effect $A(t)$ in Eq. (2·12) is defined from the real daily advertisement cost $A_{cost}(t)$ as

$$A(t) = C_{adv} \int_{-\infty}^t A_{cost}(\tau) e^{a\tau} d\tau, \quad (3·1)$$

where the coefficient C_{adv} is the strength of the advertisement effect on the purchase intention discussed in the following section.

3.3. Strength of advertisement effect

From Fig. 1, we can also find the strength of the advertisement effect for the Mizuki-Shigeru Road. The strength of the advertisement effect C_{adv} is defined as the increase in the number of attending persons or in the revenue of the Mizuki-Shigeru Road as a function of the per day advertisement cost. In this paper, we assume the coefficient of the strength of the advertisement effect C_{adv} to be constant during the duration of the event. In actual analysis, we define the actual value of C_{adv} by adjusting the corresponding peak value for the number of visitors using the advertisement cost.

3.4. Results

Using the decline factor of the advertisement effect and the strength of the advertisement effect, we calculate the purchase intention for the Mizuki-Shigeru Road in 2008. The parameters of the direct and indirect communications are adjusted to explain the observed daily number of visitors.

In order to adjust the parameters to fit the calculation with the observed blog data, the adjustment should be reliable. In this regard, we introduce the so-called R -factor (reliable factor), which is well-known in the field of low energy electron diffraction (LEED) experiments.¹⁵⁾ In a LEED experiment, the experimentally observed curve of current vs voltage is compared with the corresponding theoretical curve using the R -factor. For the LEED case, the best parameters of the material structure are those with the lowest R -factor. For our purpose, we define the R -factor as follows:

$$R = \frac{\sum_i (f(i) - g(i))^2}{\sum_i (f(i)^2 + g(i)^2)}, \tag{3.2}$$

where functions $f(i)$ and $g(i)$ correspond to the theoretically calculated and observed daily positive blog-mentions. The smaller the value of R , the better the match between functions f and g . In our calculation, we search for the parameters with the smallest R -factor.

The calculated results are shown in Fig. 2. The adjusted parameters are shown in Table I.

In Fig. 3, we show a similar calculation for the World Sand Sculpture Festival 2009 held in Tottori-City, Tottori, Japan. The rush effect is very important for this event and the rush power is adjusted to be -4 . The rush effect is assumed to be switched on for the last one week so that the parameter c in Eq. (2.13) is adjusted

Table I. Table of parameters for the Mizuki-Shigeru Road. N_p is the number of potential customers; C_{adv} is the strength of the advertisement; a is the decline factor of the advertisement; D_{nn} is the direct communication factor; P_{nn} is the indirect communication factor.

N_p	500,000
C_{adv}	0.03 (1/10 ³ yen)
a	0.4 (1/day)
$N_p D_{nn}$	0.005 (1/day)
$N_p^2 P_{nn}$	0.000013 (1/day)

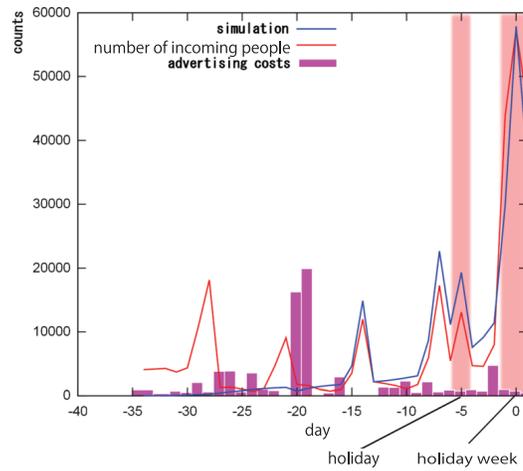


Fig. 2. Theoretical calculation of the daily number of visitors for Sakaiminato Mizuki-Shigeru Road in April–July, 2008. The horizontal axis is the date measured from the day corresponding to the peak number of persons recorded during the 2008 summer vacations.

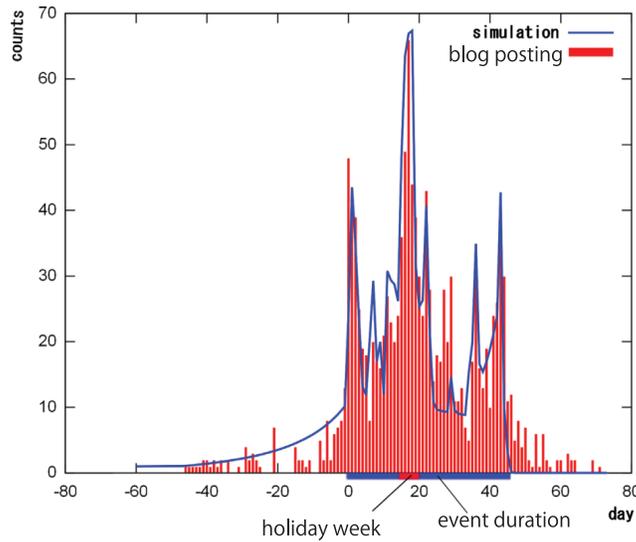


Fig. 3. Theoretical calculation of the daily positive blog-mentions for the World Sand Sculpture Festival 2009 held in Tottori-City, Tottori, Japan. The solid curve is our theoretical calculation and the histogram is the daily number of blog-mentions.

to be $f(t) = 1$ at one week before the last day of the event. We found a good fit for daily positive blog-mentions.

§4. Discussion

From Figs. 2 and 3, we found that our simulation can reproduce the daily counts of incoming persons to a local event. Since our theory is very general, it can be

Table II. Table of parameters for the World Sand Sculpture Festival 2009 held in Tottori-City, Tottori, Japan. N_p is the number of potential customers; C_{adv} is the strength of the advertisement; a is the decline factor of the advertisement; D_{nn} is the direct communication factor; P_{nn} is the indirect communication factor; “before” and “after” indicate the parameters before the event opens and after the event opens, respectively.

N_p	300,000
C_{adv}	0.0013 (1/10 ³ yen)
a	0.01 1/day
$N_p D_{nn}(before)$	0.04 (1/day)
$N_p D_{nn}(after)$	0.037 (1/day)
$N_p D_{ny}$	0.00003 (1/day)
$N_p^2 P_{nn}(before)$	0.003 (1/day)
$N_p^2 P_{nn}(after)$	0.002 (1/day)
$N_p^2 P_{ny}$	0.0001 (1/day)
$N_p^2 P_{yy}$	0.00005 (1/day)

applied not only to the motion picture entertainment industry but also to many other fields including local events. Therefore, our theory (given by Eq. (2·12)) can be used to investigate human-human interaction especially those meant for entertainment. The main assumption underlying our theory is that the price of the target is not expensive enough that the consumer would not be affected by the advertisement. Thus, when producing local events, the event producer can use our theory to estimate the rough number of incoming people from the time distribution of advertisement costs.

We successfully apply our theory to the Mizuki-Shigeru Road and the World Sand Sculpture Festival 2009 because in both these situations, the number of daily incoming guests is counted exactly using photosensors. Since it is very important to check daily revenue and daily advertisement costs as these are necessary to accurately predict near-future revenue.

§5. Conclusion

A mathematical model for hit phenomena was presented that incorporates the advertisement cost and the communication effect. Here, communication effect includes both direct and indirect communications. The results indicated that our model can predict the revenue of corresponding local events very well. The conclusion presented in this paper is very useful even in marketing science.

Acknowledgements

The advertisement cost data and the revenue data were provided by the Tourism Section of the Trading and Tourism Division of the Industry Environment Department of Sakaiminato City. This research is partially supported by the Tottori University Research Project for Rural Sustainability. The authors thank the Yukawa Institute for Theoretical Physics at Kyoto University. Last, the discussions during the YITP workshop YITP-W-11-04 on “Econophysics 2011” were helpful in the

completion of this work.

References

- 1) D. T. Allsop, B. R. Bassett and J. A. Hoskins, *J. of Advertising Research* **47** (2007), 398.
- 2) J. Kostka, Y. A. Oswald and R. Wattenhofer, *Lecture Notes in Computer Science* **5058** (2008), 185.
- 3) E. Bakshy, J. M. Hofman, W. A. Mason and D. J. Watts, *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM'11, 9-12 February 2011, Hong Kong*.
- 4) B. J. Jansen, M. Zhang, K. Sobel and A. Chowdury, *J. of the American Society for Information Science and Technology* **60** (2009), 2169.
- 5) J. Brown and P. Reingen, *J. of Consumer Research* **14** (1987), 350.
- 6) K. Murray, *J. of Marketing* **55** (1991), 10.
- 7) A. Banerjee, *Quarterly J. of Economics* **107** (1992), 797.
- 8) J. Taylor, *Brandweek* (June 2, 2003), 26.
- 9) A. Ishii and N. Yoshida, *Reports of the Faculty of Engineering Tottori University* **36** (2005), 71.
- 10) A. Ishii, N. Yoshida, H. Arakaki and F. Yamazaki, *Reports of the Faculty of Engineering Tottori University* **37** (2007), 107.
- 11) A. Ishii, S. Umemura, T. Hayashi, N. Matsuda, T. Nakagawa, H. Arakaki and N. Yoshida, arXiv:1002.4460.
- 12) N. Yoshida, A. Ishii and H. Arakaki, *Daihitto no Houteishiki (Equation of Big Hit)*, (Discover Twenty One, Tokyo, 2010), ISBN-10: 4887598467, in Japanese.
- 13) A. Ishii, H. Arakaki, S. Umemura, T. Urushidani, N. Matsuda and N. Yoshida, *Proceedings of the 2011 International Conference on Data Engineering and Internet Technology (DEIT 2011)*, Conference Record 17831, IEEE Catalog Number: CFP1113L-CDR, ISBN 978-1-4244-8581-9.
- 14) A. Ishii, H. Arakaki, N. Matsuda, S. Umemura, T. Urushidani, N. Yamagata and N. Yoshida, *New J. Phys.*, in press.
- 15) J. B. Pendry, *J. of Phys. C* **13** (1980), 937.

Testing Randomness by Means of Random Matrix Theory

Xin YANG,^{*)} Ryota ITOI and Mieko TANAKA-YAMAWAKI^{**)}

*Department of Information and Electronics, Graduate School of Engineering,
Tottori University, Tottori 680-8552, Japan*

Random matrix theory (RMT) derives, at the limit of both the dimension N and the length of sequences L going to infinity, that the eigenvalue distribution of the cross correlation matrix with high random nature can be expressed by one function of $Q = L/N$. Using this fact, we propose a new method of testing randomness of a given sequence. Namely, a sequence passes the test if the eigenvalue distribution of the cross correlation matrix made of the pieces of a given sequence matches the corresponding theoretical curve derived by RMT, and fails otherwise. The comparison is quantified by employing the moments of the eigenvalue distribution to its theoretical counterparts. We have tested its performance on five kinds of test data including the Linear Congruential Generator (LCG), the Mersenne Twister (MT), and three physical random number generators, and confirmed that all the five pass the test. However, the method can distinguish the difference of randomness of the derivatives of random sequences, and the initial part of LCG, which are distinctly less random than the original sequences.

§1. Introduction

The random matrix theory¹⁾ can be used to extract the principal components, by subtracting the random part from the time series with high randomness like stock prices.^{2),3)} We propose in this paper a new algorithm of testing the randomness of marginally-random sequences that we encounter in various situations, such as social, economic/financial or medical applications. This method is a straightforward application of the RMT-PCA method⁴⁾ originally developed in order to extract trendy business sectors from a massive database of stock prices. We name this method the ‘RMT-test’ and examine its effect on several examples of pseudo-random numbers including LCG, and MT, as well as the true random number sequences made by Toshiba, Hitachi, and Tokyo-Electron.

§2. Random matrix theory

In applying the random matrix theory, we follow the line of thought that was developed in the course of extracting the principal components of the stock time series in the markets about a decade ago.^{2),3)} Namely, we compare the eigenvalue distribution of the correlation matrix, between N time series of length L , to the corresponding theoretical formula of the eigenvalue distribution^{5),6)} derived from the random matrix theory in the limit of N and L going to infinity, keeping $Q = L/N$ as a constant.

$$P_{\text{RMT}}(\lambda) = \frac{Q}{2\pi\lambda} \sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)} \quad (2.1)$$

^{*)} E-mail: yx0709@ike.tottori-u.ac.jp

^{**)} E-mail: mieko@ike.tottori-u.ac.jp

with

$$\lambda_{\pm} = \left(1 \pm \sqrt{\frac{1}{Q}}\right)^2. \quad (2.2)$$

§3. Procedure of the RMT-test

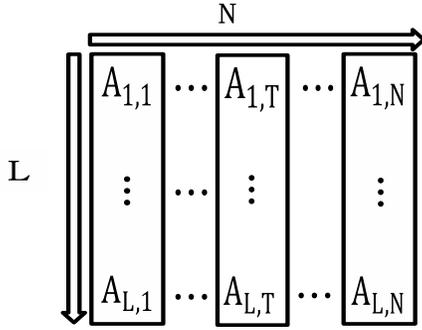


Fig. 1. Data structure.

elements in the 2nd row, etc., by discarding the remainder if the length of the sequence is not divisible by L , as shown in Fig. 1. Then we normalize each column of the matrix to have zero mean and single variance,

$$g_{i,j} = \frac{A_{i,j} - \langle A_j \rangle}{\sqrt{\langle A_j^2 \rangle - \langle A_j \rangle^2}} \quad (3.1)$$

to have the normalized matrix G as follows,

$$G = \begin{pmatrix} g_{11} & \cdots & g_{1N} \\ \vdots & \ddots & \vdots \\ g_{L1} & \cdots & g_{LN} \end{pmatrix}. \quad (3.2)$$

- Compute the correlation matrix,

$$C = \frac{1}{L} G^T G \quad (3.3)$$

which is symmetric

$$C_{i,j} = C_{j,i} \quad (3.4)$$

by definition and

$$C_{i,i} = 1 \quad (3.5)$$

due to normalization.

The method of the RMT is outlined as follows.^{2),3)} We aim to test the randomness of a long 1-dimensional sequence of numerical data.

- Preparing the data

We prepare a long enough sequence (by using the pseudo-random number generators, or downloading physical random numbers from the web site) and cut it into N pieces of equal length L , then shape them in an $L \times N$ matrix, $A_{i,j}$ by placing the first L elements in the first row of the matrix, and the next L elements in the 2nd row, etc., by discarding the remainder if the length of the sequence is not divisible by L , as shown in Fig. 1.

- Obtain the eigenvalues of correlation matrix C by numerical calculation.
- Compare the eigenvalue distribution to the corresponding theoretical formula in Eq. (2.1). If the two lines match, that data passes the RMT-test, and if they do not match, it fails the RMT-test.

We further quantify the test by adding the next step, in order to discriminate tiny differences invisible in the visual comparison of the two curves.

- Quantitative evaluation based on the moment method.
Compare the k -th moment of the obtained eigenvalues

$$m_k = \frac{1}{N} \sum_{i=1}^N x_i^k \quad (3.6)$$

to the corresponding theoretical formula obtained from P_{RMT} .

$$\mu_k = E(\lambda^k) = \int_{\lambda_-}^{\lambda_+} \lambda^k P_{RMT}(\lambda) d\lambda. \quad (3.7)$$

The sample sequence passes the quantitative RMT-test (Quantitative) if the ratio of the moment m_k over its theoretical value μ_k is close to 1. The moments up to 6th can be expressed by the function of Q as follows,

$$\mu_1 = 1, \quad (3.8)$$

$$\mu_2 = 1 + \frac{1}{Q}, \quad (3.9)$$

$$\mu_3 = 1 + \frac{3}{Q} + \frac{1}{Q^2}, \quad (3.10)$$

$$\mu_4 = 1 + \frac{6}{Q} + \frac{6}{Q^2} + \frac{1}{Q^3}, \quad (3.11)$$

$$\mu_5 = 1 + \frac{10}{Q} + \frac{20}{Q^2} + \frac{10}{Q^3} + \frac{1}{Q^4}, \quad (3.12)$$

$$\mu_6 = 1 + \frac{15}{Q} + \frac{50}{Q^2} + \frac{50}{Q^3} + \frac{15}{Q^4} + \frac{1}{Q^5}. \quad (3.13)$$

By using those formulas, we evaluate the errors of k -th moments by the deviation of the ratio of the experimental value over the theoretical formula in Eqs. (3.9) to (3.13) from one, as follows,

$$error = m_k / \mu_k - 1. \quad (3.14)$$

We can choose the optimal level of error $< 5\%$ to judge the randomness of the sequence, based on our experiments.

§4. Applications of the RMT-test on random sequences

4.1. Determining the reasonable range of N and L

In this section, we determine the reasonable range of N and L . Since the theoretical formula of the eigenvalue distribution P_{RMT} is derived at the limit of N and L being infinity, we need to choose large enough N and L in order to justify the test.

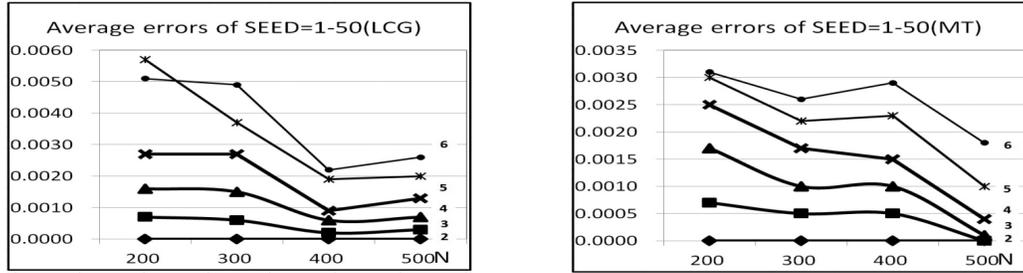


Fig. 2. Errors of the moments ($k = 1, \dots, 6$) for $N = 200, \dots, 500$ for LCG and MT.

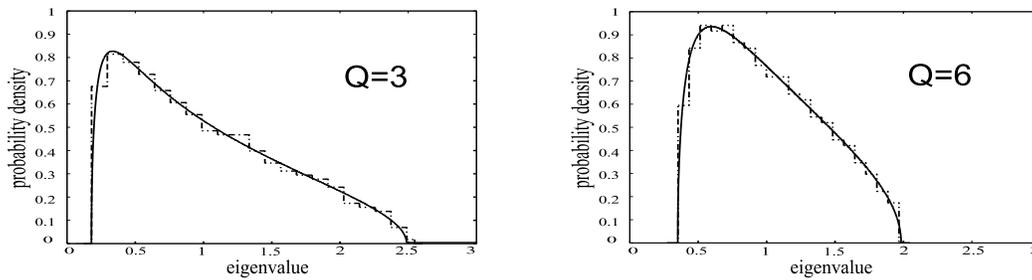


Fig. 3. Examples of pseudo-random sequences by LCG passing the RMT-test.

For this purpose, we have applied the test on the data taken from the two popular pseudo random number generators, LCG and MT, for $N = 200, 300, 400, 500$ at $Q = 3$, and compared the moments up to the 6th order to the corresponding theoretical formula. We have performed the experiment using 50 different values of seed, of $SEED = 1, \dots, 50$, and took the average.

The result is shown in the two figures of Fig. 2, by six lines corresponding the 1st to the 6th moments from the bottom to the top, for LCG (left) and MT (right). In both figures, the errors go down gradually as N increases from 200 to 300, and to 400, then become stable after N reaches the range of 400–500. At $N = 500$, the errors become smaller than 0.3%. Based on this fact, we justify the value $N = 500$ being large enough (at least for $Q = 3$) to apply our RMT-test.

4.2. Qualitative evaluation of randomness of pseudo-random generators

The LCG⁷⁾ is the most popular pseudo-random number generators, in which the random numbers are generated by the following formula,

$$X_{n+1} = (aX_n + b) \bmod M. \quad (4.1)$$

The following parameters are used in the above formula, for $\text{rand}(\)$,

$$a = 1103515245, \quad b = 12345, \quad M = 2147483648. \quad (4.2)$$

We generate a random sequence of length $500 \times 1,500$, then cut it into 500 ($= N$) pieces of length 1,500 ($= L$) each to make LCG ($Q = 3$) data. Although LCG is

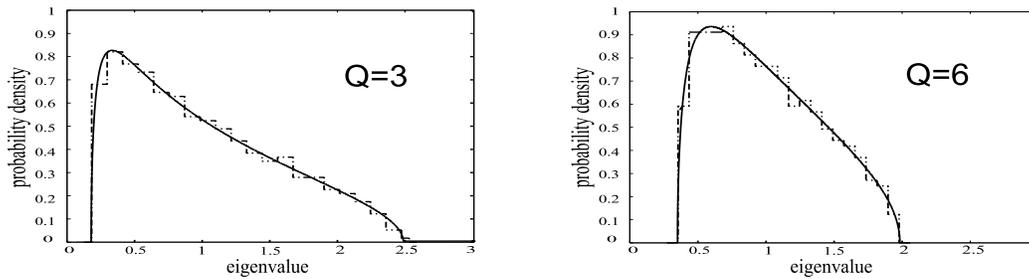


Fig. 4. Examples of pseudo-random sequences by MT passing the RMT-test.

known to have many problems, the RMT-test cannot detect its off-randomness. As is shown in the left figure of Fig. 3, this data passes the RMT-test safely for $Q = 3$ (left) with a wide variety of seeds. The right figure of Fig. 3 is a corresponding result for $Q = 6$ at $N = 500$, $L = 3,000$.

The Mersenne Twister (MT)⁸⁾ is a recently proposed, highly reputed random number generator. The most valuable feature of MT is its extremely long period, $2^{19937} - 1$. We test the randomness of MT in the same procedure as above. The result is shown in Fig. 4 for $Q = 3$, and $Q = 6$. The MT also passes the RMT-test in a wide range of N and L .

So far, we have seen the two popular pseudo-random sequences pass the RMT-test. We need at this moment some sequences of lower randomness in order to discriminate the level of randomness to be detected by using the RMT-test. The first example is a set of the initial part of random numbers of LCG and MT.

4.3. Qualitative evaluation of randomness of physical random number

We test the randomness of some physical random numbers.⁹⁾ Because the physical random numbers have neither regularity nor reproducibility, the forecast of the sequence is impossible (for example: we cannot predict the points when we throw a dice). We use physical random numbers generated by three physical random number generators: Toshiba, Hitachi, and Tokyo-Electron obtained from the homepage of the institute of statistical mathematics, and show the results in Figs. 5, 6 and 7, respectively. The parameters N and L are chosen to be the same as the cases of LCG and MT for the sake of comparison, such that $N = 500$, $L = 1,500$ (left) and $N = 500$, $L = 3,000$ (right). Note that all the examples pass the RMT-test, for the wide variety of seeds.

4.4. Quantitative evaluation of the degree of randomness by means of moments

We have come to the point of discussing the choice of randomness measure. In §4.1, we argued the validity of the RMT-test for the parameter $N > 400$ for $Q = L/N = 3$, and 6, based on the comparison of the experimental value over the theoretical value of the k -th moments for $k = 1-6$, using the last step introduced in Chapter 3. We observed that the errors between the experimental moments and its theoretical counterparts do not converge to zero, but gradually reduce to the

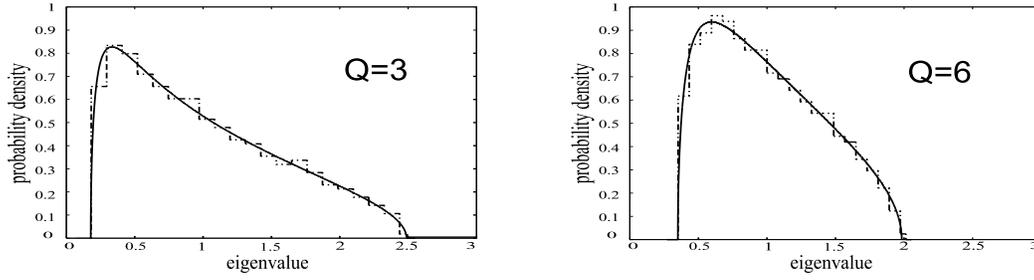


Fig. 5. Examples of physical random numbers by Toshiba passing the RMT-test.

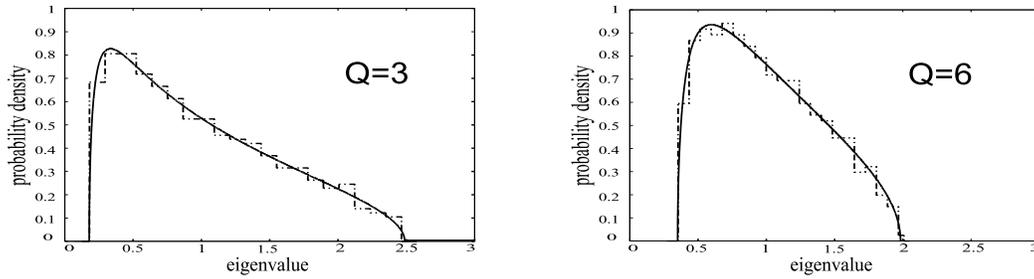


Fig. 6. Examples of physical random numbers by Hitachi passing the RMT-test.

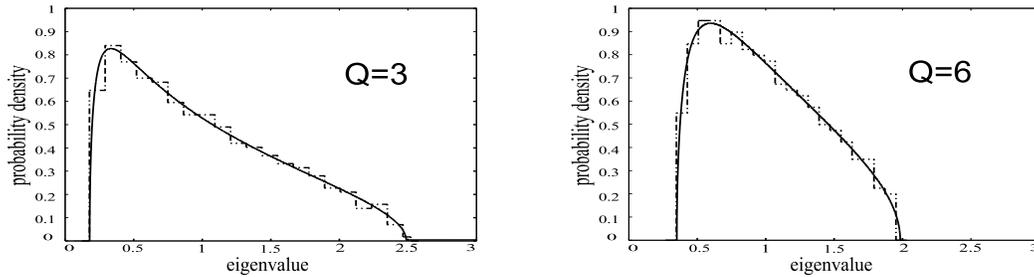


Fig. 7. Examples of physical random numbers by Tokyo-Electron passing the RMT-test.

value less than 0.2–0.3% as N reaches the region of 400–500. Is this because the two pseudo random numbers are not perfectly random, or the RMT formula is not perfectly valid for the finite values of N and L .

In order to answer to this question, we examine the degree of randomness of the physically generated random numbers by the moment method as we introduced in §4.1 and compare the errors to those of the pseudo-random numbers. Since the case of $k = 1$ is trivial because the average of all the eigenvalues thus equal to one by the normalization condition, Eq. (3.5) and the invariance of the trace under similarity transformations used in the process of diagonalization of the correlation

Table I. $Q = 3$ ($L = 1,500$) average (standard deviation) of 50 independent tests.

k	LCG	MT	Toshiba	Hitachi	Tokyo Electron
2	-.0003(.0009)	.0000(.0010)	-.0004(.0010)	-.0004(.0011)	-.0000(.0008)
3	-.0007(.0023)	.0001(.0026)	-.0010(.0027)	-.0010(.0030)	-.0002(.0021)
4	-.0013(.0039)	.0004(.0045)	-.0017(.0048)	-.0017(.0053)	-.0005(.0038)
5	-.0020(.0058)	.0010(.0067)	-.0025(.0073)	-.0022(.0081)	-.0008(.0057)
6	-.0026(.0080)	.0018(.0091)	-.0033(.0101)	-.0027(.0113)	-.0010(.0079)

Table II. $Q = 3$ ($L = 1,500$) [min.:max.] of 50 independent tests.

k	LCG	MT	Toshiba	Hitachi	Tokyo Electron
2	[-.0021 : .0016]	[-.0021 : .0024]	[-.0021 : .0031]	[-.0026 : .0030]	[-.0018 : .0017]
3	[-.0051 : .0040]	[-.0052 : .0061]	[-.0058 : .0077]	[-.0064 : .0081]	[-.0047 : .0048]
4	[-.0086 : .0065]	[-.0086 : .0109]	[-.0111 : .0131]	[-.0107 : .0147]	[-.0080 : .0094]
5	[-.0124 : .0104]	[-.0122 : .0164]	[-.0174 : .0191]	[-.0149 : .0226]	[-.0126 : .0149]
6	[-.0164 : .0152]	[-.0160 : .0227]	[-.0243 : .0257]	[-.0206 : .0316]	[-.0177 : .0211]

Table III. $Q = 6$ ($L = 3,000$) average (standard deviation) of 50 independent tests.

k	LCG	MT	Toshiba	Hitachi	Tokyo Electron
2	-.0001(.0005)	-.0003(.0006)	-.0001(.0006)	-.0004(.0010)	-.0002(.0005)
3	-.0003(.0014)	-.0007(.0016)	-.0004(.0016)	-.0010(.0021)	-.0005(.0013)
4	-.0006(.0026)	-.0012(.0030)	-.0008(.0029)	-.0015(.0035)	-.0009(.0025)
5	-.0008(.0041)	-.0017(.0046)	-.0012(.0045)	-.0021(.0051)	-.0012(.0038)
6	-.0010(.0059)	-.0021(.0065)	-.0016(.0063)	-.0028(.0069)	-.0014(.0053)

Table IV. $Q = 6$ ($L = 3,000$) [min.:max.] of 50 independent tests.

k	LCG	MT	Toshiba	Hitachi	Tokyo Electron
2	[-.0012 : .0009]	[-.0016 : .0010]	[-.0014 : .0011]	[-.0061 : .0011]	[-.0012 : .0008]
3	[-.0033 : .0027]	[-.0042 : .0025]	[-.0036 : .0030]	[-.0101 : .0032]	[-.0032 : .0023]
4	[-.0060 : .0054]	[-.0075 : .0044]	[-.0064 : .0060]	[-.0150 : .0060]	[-.0059 : .0045]
5	[-.0093 : .0088]	[-.0111 : .0072]	[-.0100 : .0098]	[-.0205 : .0092]	[-.0089 : .0078]
6	[-.0128 : .0128]	[-.0150 : .0111]	[-.0144 : .0145]	[-.0265 : .0127]	[-.0121 : .0120]

matrix, we show the results of the k -th moment for $k = 2, \dots, 6$ in the rest of our discussion. In Table I, we summarize the average values (and the standard deviation in parentheses) of 50 independent experiments for three physical random numbers (Toshiba, Hitachi and Tokyo-Electron) and the two pseudo random numbers (LCG, MT).

The k -th moment ratio for three physical random numbers by the three physical random numbers are compared to the corresponding results of two pseudo-random numbers, LCG and MT ($SEED = 1 - 50$, $N = 500$).

Among three physical random generators, Tokyo-Electron generates the most random sequences compared to Hitachi or Toshiba. However, those physical generators are basically less stable compared to the pseudo random generators. In other words, pseudo random generators can produce more uniform sequences with high stability. Moreover, they are deterministic and the entire sequence can be reproduced

once the *SEED* is known together with the algorithm. The results that are shown from Table I to IV tell us that the moment analysis cannot distinguish significant difference among LCG, MT and the three physical random numbers, in accordance with our qualitative test, by which they all pass the RMT-test and the degrees of randomness are indistinguishable. This fact supports the importance of the qualitative test in Figs. 3–7. We conclude that a given sequence passes the RMT-test if it passes the qualitative test, and the corresponding quantitative test by means of moment analysis gives the error less than a few percent for all the moments for $k = 1$ to 6. We next deal with the examples that fail the RMT-test.

§5. Detecting off-randomness by means of the RMT-test

Table V. Randomness of the initial parts of pseudo random numbers.

k	LCG	MT
2	.0046	-.0018
3	.0102	-.0042
4	.0198	-.0064
5	.0356	-.0083
6	.0592	-.0099

In this chapter, we test the off-randomness of two examples by using the RMT test. Our first example is the initial part of LCG sequences and the second example is the log-return sequence frequently used in financial analysis.

5.1. Testing the initial part of LCG

The initial part of LCG is generally believed to have low randomness. In order

to quantitatively measure the degree of randomness, we apply the RMT-test on the collection of initial parts of LCG, and compare them with the corresponding data of MT.

We collect the initial 500 numbers generated by iterating the LCG of Eqs. (4-1) and (4-2), starting from various seeds and connect the outputs to serve as out data sequence. As is shown in the left figure of Fig. 8, RMT-test has detected a sign of deviation from RMT formula, for the case of $N = 500$, and $L = 1,500$, since some eigenvalues are larger than the theoretical maximum. On the other hand, the corresponding case of the same data without the first 500 numbers after the seeds passes the RMT-test, having all the eigenvalues within the theoretical curve, as shown in the right figure of Fig. 8. For the sake of comparison, we have done the same experiment for another generator, MT, and have confirmed that both the initial and the rest of the sequence pass the RMT-test. The quantitative measure of the off-randomness measured by the first six moments is given in Table V. From this we learn that the errors of moment ratio for the initial 500 elements of LCG are considerably large compared to the corresponding elements of MT.

5.2. Testing the randomness of log-return sequences

It is customary to convert the price time series p_1, p_2, \dots, p_L to the log-return time series r_1, r_2, \dots, r_{L-1} by means of Eq. (5-1) in the financial analysis, in order to eliminate the unit/size dependence of different stock prices.

$$r_i = \log p_i - \log p_{i-1}. \quad (5-1)$$

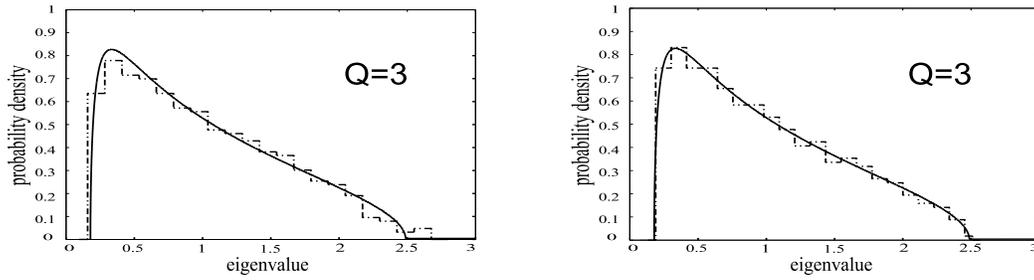


Fig. 8. The initial data generated by LCG (MT) fails (passes) the RMT-test.

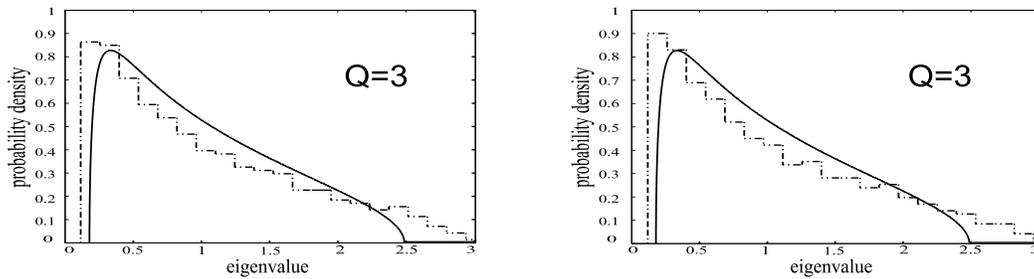


Fig. 9. RMT-test of the log-return sequences of LCG (left) and MT (right).

Table VI. Errors in RMT-test for the overlapping (left) and non-overlapping (right) log-return sequences.

k	LCG	MT
2	.1047	.1227
3	.2578	.3088
4	.4445	.5442
5	.6596	.8260
6	.9092	1.1738

k	LCG	MT
2	-.0172	.0012
3	-.0436	.0030
4	-.0744	.0050
5	-.1070	.0068
6	-.1400	.0085

However, this process involves the same p_i for r_i as well as r_{i+1} . Because of this, the time series of log-returns lose the randomness that existed in the original price time series and a certain pattern specific to the log-return time series emerges.

In this section, we measure the degree of randomness of such log-return series by using the two pseudo-random generators, LCG and MT, and identify the effect of converting financial time series to the log-return sequence. We compare the results of LCG and MT, by generating the series to make $N = 500$ and $L = 1,500$ ($Q = 3$) and execute the process of steps (1)–(4) in Chapter 3. The results are shown in Fig. 9 and the corresponding moment analysis including the step (5) in Chapter 3 is shown in Table VI (left). We also point out that this effect can be eliminated if we take the non-overlapping log-return by giving up the half of the total elements of r_i ($i = \text{even or odd}$), in exchange of the length of data L to one half of the original.

Table VII. Range of eigenvalues, $\lambda_+ - \lambda_-$ for LCG and MT are compared to the theoretical value of $4/\sqrt{Q}$ derived by RMT.

Q	RMT($4/\sqrt{Q}$)	LCG	LCG/RMT	MT	MT/RMT
2	2.82	3.43	1.22	3.43	1.22
3	2.30	2.79	1.21	2.80	1.22
4	2.00	2.40	1.20	2.41	1.21
5	1.78	2.15	1.21	2.15	1.21
6	1.63	1.97	1.21	1.96	1.20
7	1.51	1.81	1.20	1.82	1.21
8	1.41	1.70	1.21	1.70	1.21
9	1.33	1.60	1.20	1.60	1.20
10	1.26	1.50	1.19	1.49	1.18

The result of moment analysis is given in Table VI (right). The error compared to the RMT-formula is as large as 100%. This effect results in the expansion of the range of eigenvalues, $\lambda_+ - \lambda_-$ compared to the theoretical range of eigenvalues derived by RMT as $\lambda_+ - \lambda_- = 4/\sqrt{Q}$ from Eq. (2·2) and the size of expansion is approximately 20% increase of the theoretical range, as shown in Table VII.

§6. Conclusion and discussion

6.1. Cumulant analysis

We have so far used the moment analysis in Eqs. (3·9)–(3·13) for our quantitative test. Often a set of cumulants is used in place of moments. We have derived the corresponding expression of cumulants in terms of parameter Q , up to the 6th order as follows, where κ_i denotes the i -th cumulant. In this paper we do not use them for our quantitative analysis because the 6th cumulant gives an extremely large error. However, the following result may become useful for constructing a quantitative test by using the cumulants of the low degree, such as up to 3rd or 5th cumulants.

$$\kappa_1 = \mu_1 = 1, \quad (6\cdot1)$$

$$\kappa_2 = \mu_2 - \mu_1^2 = \frac{1}{Q}, \quad (6\cdot2)$$

$$\kappa_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 = \frac{1}{Q^2}, \quad (6\cdot3)$$

$$\kappa_4 = \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4 = \frac{1}{Q^3} - \frac{1}{Q^2}, \quad (6\cdot4)$$

$$\begin{aligned} \kappa_5 &= \mu_5 - 5\mu_4\mu_1 - 10\mu_3\mu_2 + 20\mu_3\mu_1^2 + 30\mu_2^2\mu_1 - 60\mu_2\mu_1^3 + 24\mu_1^5 \\ &= \frac{1}{Q^4} - \frac{5}{Q^3}, \end{aligned} \quad (6\cdot5)$$

$$\begin{aligned} \kappa_6 &= \mu_6 - 6\mu_5\mu_1 - 15\mu_4\mu_2 + 30\mu_4\mu_1^2 - 10\mu_3^2 + 120\mu_3\mu_2\mu_1 - 120\mu_3\mu_1^3 \\ &\quad + 30\mu_2^3 - 270\mu_2^2\mu_1^2 + 360\mu_2\mu_1^4 - 120\mu_1^6 = \frac{1}{Q^5} - \frac{16}{Q^4} + \frac{5}{Q^3}. \end{aligned} \quad (6\cdot6)$$

6.2. Discussion

Compared to other conventional methods of testing randomness, the RMT-test that we proposed in this paper can be applied on wide range of numerical data, independent of its data format or types. Moreover, the result is visually presented in a graph that can be grasped intuitively. It is particularly suitable to test the randomness of very long, massive data sequences. No null hypothesis, or other complicated process is required. On the other hand, the method uses a very long data sequence. In order for the RMT-formula to work, we need N strings of length L larger than N , where N is larger than 400. For example, to make $N = 500$ for $Q = 3$, we need a data string of length 750,000. Thus the application is limited to the world in which plenty of numerical data can be accumulated.

6.3. Conclusion

In this paper, we proposed a new method of testing randomness, RMT-test, as a by-product of the RMT-PCA that we used to extract trends of stock markets. In order to examine its effectiveness, we tested it on two random number generators, LCG and MT, and three physical random numbers made by Toshiba, Hitachi and Tokyo-Electron. The result shows that all of them pass the RMT-test for a wide range of parameters. Although the physical random numbers by Tokyo-Electron are relatively better than the two other physical random numbers, the degrees of randomness of the three are indistinguishable both in our qualitative test and in our quantitative test. We further tested the validity of our RMT-test on the sequences of low randomness and showed that the RMT-test can detect off-randomness successfully.

Acknowledgements

The authors thank the Yukawa Institute for Theoretical Physics at Kyoto University. Discussions during the YITP workshop YITP-W-11-04 on “Econophysics 2011” were useful to complete this work.

References

- 1) M. Mehta, *Random Matrices*, 3rd edition (Academic Press, 2004).
- 2) V. Plerou, P. Gopikrishnan, B. Rosenow, L. Amaral and H. Stanley, *Phys. Rev. E* **65** (2002), 066126.
- 3) L. Laloux, P. Cizeaux, J. Bouchaud and M. Potters, *Phys. Rev. Lett.* **83** (1998), 1467.
- 4) M. Tanaka-Yamawaki, *Intelligent Information Management* **3** (2011), 65.
- 5) V. Marcenko and L. Pastur, *Mathematics of the USSR-Sbornik* **1** (1994), 457.
- 6) A. Sengupta and P. Mitra, *Phys. Rev. E* **60** (1999), 3389.
- 7) S. Park and K. Miller, *Communication of ACM* **31** (1988), 1192.
- 8) M. Matsumoto and T. Nishimura, *ACM Trans. on Modeling and Computer Simulation* **8** (1998), 3.
- 9) Y. Tamura, Random Number Library, <http://random.ism.ac.jp/random>.

Numerical Study of Generalized Random Correlation Matrices: Autocorrelation Effects

Yuta ARAI^{1,*} and Hiroshi IYETOMI^{2,**}

¹*Graduate School of Science and Technology, Niigata University,
Niigata 950-2181, Japan*

²*Department of Physics, Niigata University, Niigata 950-2181, Japan*

We report the numerical calculations of the maximal eigenvalue for random correlation matrices which contain autocorrelations in data. Here the AR(1) model is adopted for such a study, we work out an empirical formula for autocorrelation correction of the maximal eigenvalue, which are accurate in a wide range of parameters. As an application of this formula, we propose a criterion to single out statistically meaningful correlations in the principal component analysis. The new criterion within the AR(1) model incorporates autocorrelation effects into the current method based on the random matrix theory.

§1. Introduction

In econophysics, the random matrix theory (RMT) has been successfully used for principal component analyses of various market data.^{1)–7)} The theoretical expressions of the eigenvalue distribution and its upper edge λ_+ for random correlation matrices work as a very useful null model for extracting significant information on correlation structures in the time series data.

We begin with remarking that the analytical results of the RMT are valid only in the case where the elements of random correlation matrices are totally independent. When the matrix elements have autocorrelations, the eigenvalues of random correlation matrices may have a certain distribution in the region beyond λ_+ , which is called “autocorrelation effects” here. If time series data under study contain no appreciable autocorrelations, such an effect can be neglected. In general, however, time series data can not be free of autocorrelations. Therefore, we should take account of the autocorrelation effect on the RMT formulae to improve the RMT criterion in the principal component analyses of actual data.

In this paper, we numerically discuss the autocorrelation effect on the RMT within the autoregressive model of order 1, denoted as AR(1). Especially, we focus on the maximal eigenvalue of correlation matrices constructed by AR(1) random number series. We then propose a new criterion for the principal component analysis to single out genuine correlations in multivariate time series data which are basically described by AR(1). In passing we note that we have recently studied finite size effects on the eigenvalue distribution of random correlation matrices.⁸⁾

*) E-mail: y.arai@phys.sc.niigata-u.ac.jp

**) E-mail: hiyetomi@sc.niigata-u.ac.jp

§2. Autocorrelation effects on random matrix theory

We begin with introducing a random matrix \mathbf{H} which is a $N \times T$ matrix with elements $\{h_{ij}; i = 1, \dots, N; j = 1, \dots, T\}$. Its elements h_{ij} are random variables following normal distribution $N(0, 1)$ and hence mutually independent. The correlation matrix is then defined by

$$\mathbf{C} = \frac{1}{T} \mathbf{H} \mathbf{H}^T. \tag{2.1}$$

In the limit $N, T \rightarrow \infty$ with $Q \equiv T/N$ fixed, the probability density function $\rho(\lambda)$ of eigenvalue λ of the random correlation matrix \mathbf{C} is analytically obtained as

$$\rho(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \tag{2.2}$$

$$\lambda_{\pm} = (1 \pm 1/\sqrt{Q})^2, \tag{2.3}$$

for $\lambda \in [\lambda_-, \lambda_+]$, where λ_- and λ_+ are the minimum and maximum eigenvalues of \mathbf{C} , respectively. Here we should note that the analytic result Eq. (2.2) is valid for $\{h_{ij}\}$ are mutually independent.

Using principal component analysis, we want to extract significant correlation structures in the multivariable time series data. In general, the time series data contains cross-correlation and autocorrelation at the same time:

- Cross-correlation
Cross-correlation is a measure of similarity of two different time series.
- Autocorrelation
Autocorrelation is a representation of similarity between a given time series and a lagged version of itself.

In multivariable time series data analysis, we are interested in the cross-correlation structures. Therefore we want to distinguish between the cross-correlations and the autocorrelations in principal component analysis.

To demonstrate the influence of autocorrelations, we generate time series data which consists of difference of random variables. First, we generate random number series $\{\epsilon_{i,t}\}$ which follows the normal distribution $N(0, 1)$. Then, we calculate time series $\{w_{i,t}\}$ which is defined by

$$w_{i,t} \equiv \epsilon_{i,t+1} - \epsilon_{i,t}. \tag{2.4}$$

Here we calculate the autocorrelations of τ time lag of the time series $\{w_{i,t}\}$, which are calculated as

$$\phi(\tau) = \frac{\sum_{t=1}^{T-\tau} (w_{i,t} - \bar{w}_i)(w_{i,t+\tau} - \bar{w}_i) / (T - \tau)}{\sum_{t=1}^T (w_{i,t} - \bar{w}_i)^2 / T}, \tag{2.5}$$

where \bar{w}_i denotes the time average of $w_{i,t}$. Then the autocorrelation of the time series data $\{w_{i,t}\}$ is $-1/2$ for adjacent elements ($\tau = 1$ in Eq. (2.5)). Beyond $\tau = 1$, the autocorrelations exactly vanish.

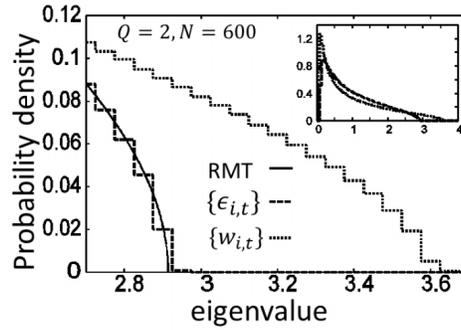


Fig. 1. Three eigenvalue distributions are compared. Note that the eigenvalue distribution of the time series $\{w_{i,t}\}$ appreciably differs from that of RMT.

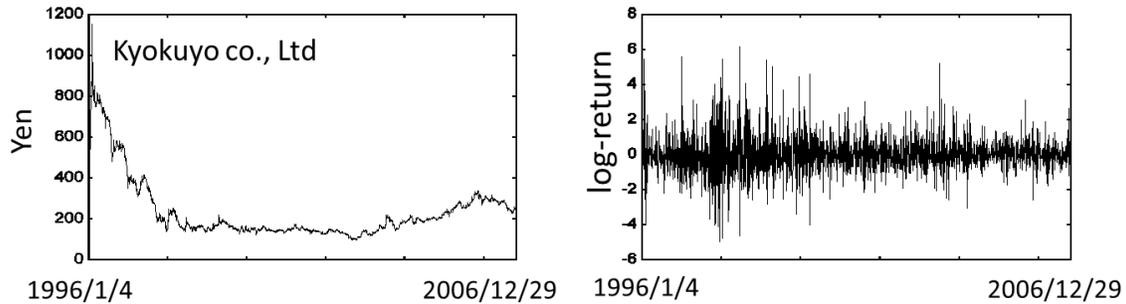


Fig. 2. The left panel shows stock price of Kyokuyo Co., Ltd. The right panel shows standardized log-return of the price change of Kyokuyo Co., Ltd.

Figure 1 shows three eigenvalue distributions. Block line, broken line and dotted line correspond to those based on RMT, time series $\{\epsilon_{i,t}\}$ and time series $\{w_{i,t}\}$, respectively. The eigenvalue distribution of the time series $\{\epsilon_{i,t}\}$ follows that of RMT. However, when the time series data contains the autocorrelations such as $\{w_{i,t}\}$, the eigenvalue distribution differs from that of RMT, which is called “autocorrelation effects” here.

§3. Example of the data analysis

In this section, we show two examples of the data analysis using actual data.

3.1. Stock data

First, we show the daily prices of $N = 557$ stocks belonging to the Tokyo Stock Exchange (TSE) for 10-year period ($T = 2,706$ daily returns).

The left panel of Fig. 2 shows one of the stock prices. To make the stock data stationary, we take log-return and standardize of those as shown in the right panel of Fig. 2. Figure 3 shows the autocorrelations of 1 day time lag in the stock data. Here the autocorrelations of τ day time lags are calculated by Eq. (2.5), where $w_{i,t}$ is standardize log-return of stock i at time t . Most of the autocorrelations fluctuate around 0. The same is time for 2 day and 3 day time lags. We thus infer that this

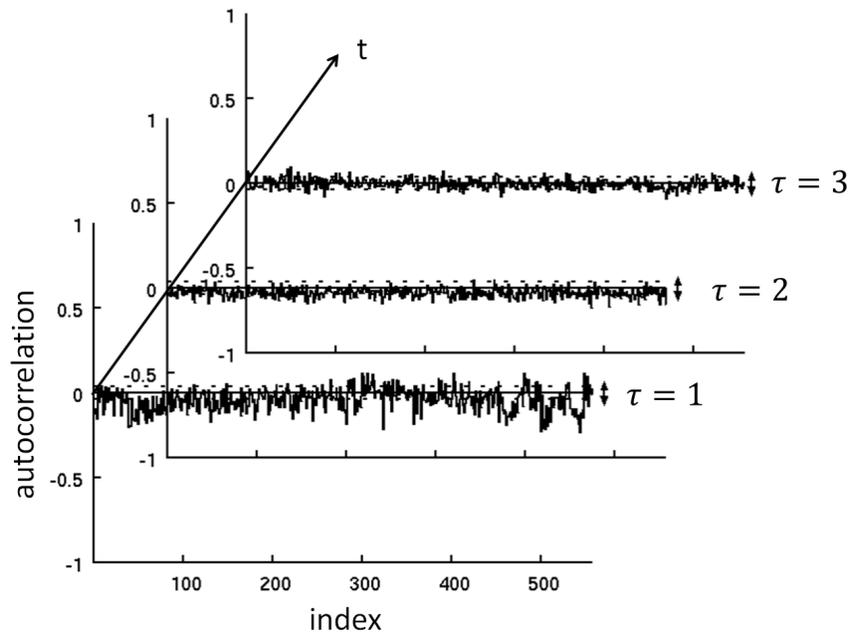


Fig. 3. Autocorrelation coefficients of the stock prices for 1, 2, and 3 day time lags. The vertical arrowed lines represent deviation of 2σ for autocorrelations of the corresponding random data.

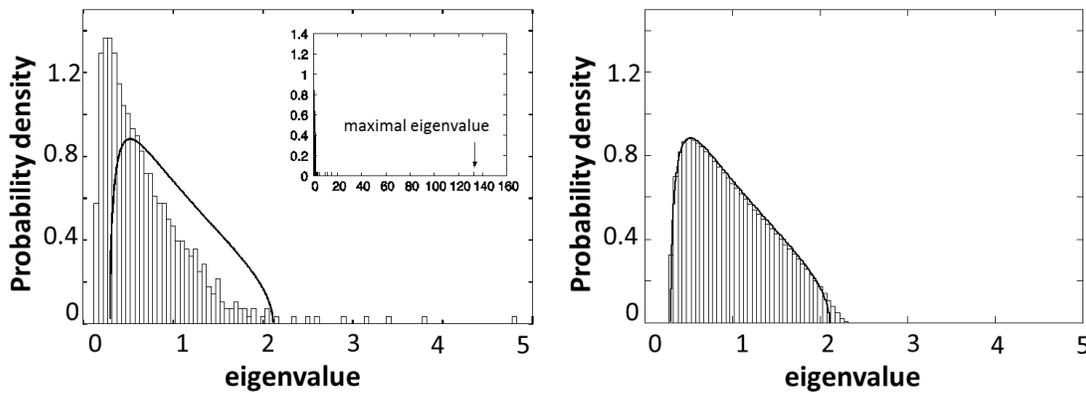


Fig. 4. The left panel shows the probability density histogram of the eigenvalues for the stock data (the maximal eigenvalue is 132.95), and the right panel, that for the rotationally shuffled data. The solid curve in each panels represents the eigenvalue distribution of the RMT.

stock data contains no appreciable autocorrelations.

In Fig. 4, the right panel shows the eigenvalue distribution of rotational random shuffled stock data. Here, the rotational random shuffling destroys only the cross-correlations, with the autocorrelations preserved.⁹⁾ The eigenvalue distribution of rotational random shuffled stock data almost follows that of RMT. Therefore the eigenvalues, which are larger than λ_+ , are regarded as the appearance of the cross-correlation structures.

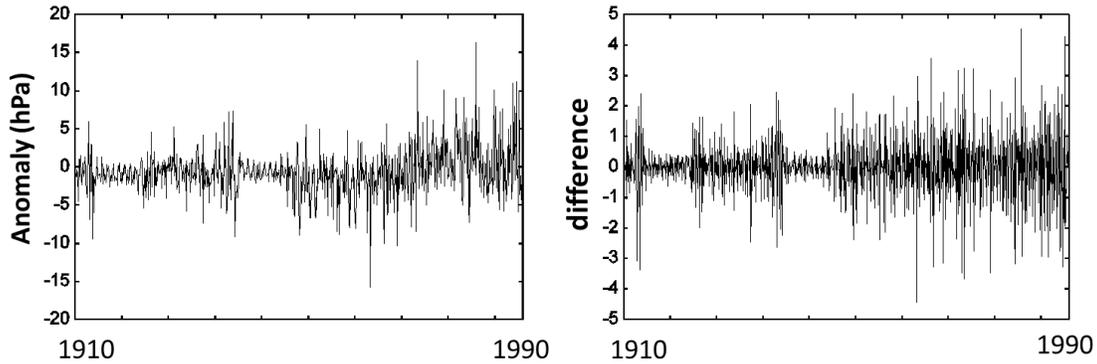


Fig. 5. The left panel shows one of the SLP data. The right panel shows standardized difference of the SLP data at some observation point.

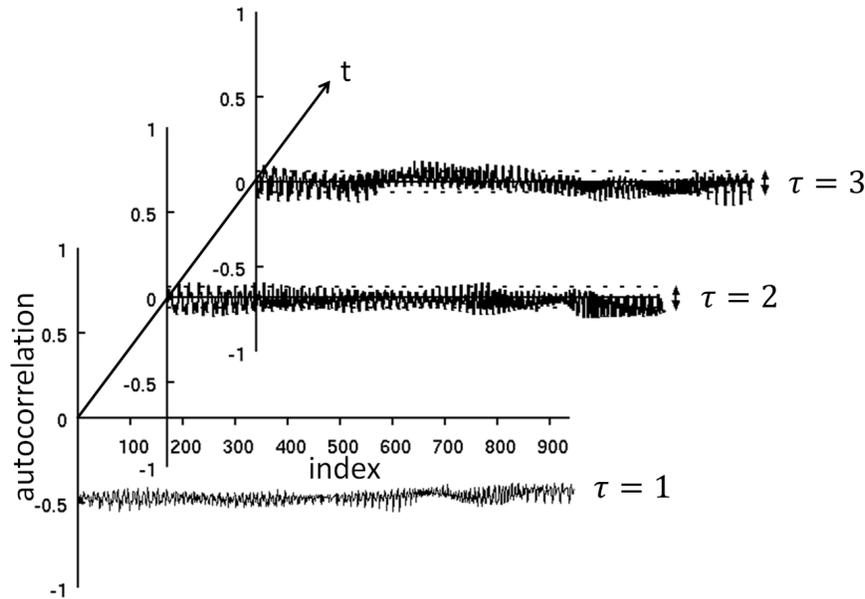


Fig. 6. Autocorrelation coefficient of 1 month, 2 month and 3 month time lags of the SLP data. The vertical arrowed lines have the same meaning as in Fig. 3.

3.2. SLP data

Next, we show the monthly Sea Level Pressure (SLP) anomaly data for 80-year period 1910–1990 ($T = 959$ monthly difference). The number of observation points is 948 ($N = 948$).

In Fig. 5, the left panel shows one of SLP data. To make the SLP data stationary, we take difference and standardize (right panel of Fig. 5).

Figure 6 shows the autocorrelations in SLP data. As with the stock data, the autocorrelations of τ month time lags are calculated by Eq. (2.5). This data indicates that SLP data contains strong autocorrelations $-1/2$ for the 1 month time lag. For 2 month and 3 month time lags, most of the autocorrelations fluctuate around 0 same as the stock data.

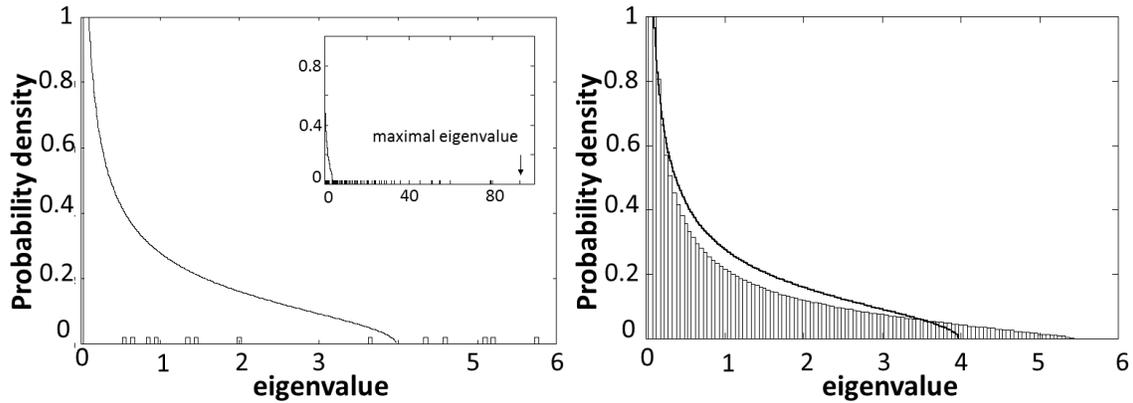


Fig. 7. The left panel shows the probability density histogram of the eigenvalues for the SLP data (the maximal eigenvalue is 93.25), and the right panel, that for the rotational random shuffled data. The solid curve in each panels represents the eigenvalue distribution of the RMT.

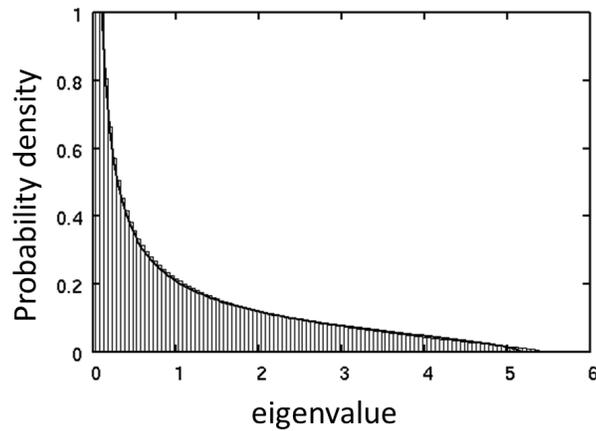


Fig. 8. This figure compares the eigenvalue distribution of $\{w_{i,t}\}$ with that of the rotationally shuffled SLP data. The solid curve corresponds to the eigenvalue distribution of the $\{w_{i,t}\}$ and the histogram is that of the rotationally shuffled SLP data.

Figure 7 shows the eigenvalue distributions of original and rotationally shuffled SLP data. The autocorrelation effects are clearly visible. These eigenvalues, which are larger than λ_+ , are appearance of autocorrelation structures. The rotational shuffling completely destroys cross-correlations preserving autocorrelations. Figure 8 instead compares the eigenvalue distribution of the SLP data so randomized with that of $\{w_{i,t}\}$. We see that $\{w_{i,t}\}$ plays a more appropriate role of the reference than $\{\epsilon_{i,t}\}$ does in the case of SLP data. Therefore, when we adopt the criterion λ_+ , we might regard the autocorrelations as cross-correlations. So we wish to have a criterion taking account of the autocorrelation effects.

§4. Multivariate autoregressive model

To generate time series data with autocorrelations, we adopt AR(1) model. AR(1) model is given by

$$x_{t+1} = \rho x_t + \epsilon_t, \quad (4.1)$$

where ϵ_t is a random variable following the normal distribution $N(0, 1)$. If absolute value of ρ is less than 1, this process is wide-sense stationary. Then we take difference of the time series $\{x_{i,t}\}$

$$w_{i,t} = x_{i,t+1} - x_{i,t} \quad (i : 0 \sim N, t : 0 \sim T + 1) \quad (4.2)$$

and standardize

$$\omega_{i,t} \equiv \frac{w_{i,t} - \langle w_i \rangle_t}{\sigma_i} \quad (i : 0 \sim N, t : 0 \sim T), \quad (4.3)$$

$$\langle w_i \rangle_t \equiv \frac{1}{T} \sum_{t=1}^T w_{i,t}, \quad \sigma_i \equiv \sqrt{\frac{1}{T} \sum_{t=1}^T (w_{i,t} - \langle w_i \rangle_t)^2}. \quad (4.4)$$

Clearly, this standardized time series data $\{\omega_{i,t}\}$ contain no cross-correlations, but have autocorrelations built in.

For $\rho = 1$ in Eq. (4.1), AR(1) model corresponds to Brownian motion. So the time series $\{\omega_{i,t}\}$ just reduces to random number series. Therefore the eigenvalue distribution follows that of RMT. On the other hands, for $\rho = 0$ in Eq. (4.1), AR(1) model corresponds to random number series. The eigenvalue distribution does not follows that of RMT as we have already shown in §3.

In this study, changing ρ ($0 \leq \rho \leq 1$), we numerically study the eigenvalue distribution with particular attention to the maximal eigenvalue.

§5. Numerical results for autocorrelation effects

In the following, λ_m denotes the statistical average of the maximal eigenvalue obtained by numerical calculation for ρ and Q . The number of samples is 10^4 , leading to statistical errors negligible in the present fitting. We define $\delta(\rho, Q)$ as

$$\delta(\rho, Q) \equiv \lambda_m(\rho, Q) - \lambda_m(\rho = 1, Q). \quad (5.1)$$

Then this $\delta(\rho, Q)$ represents autocorrelation correction to the maximal eigenvalue.

We calculate δ for $\rho = 0, 0.1, 0.2, \dots, 1$ and $Q = 1, 2, 3, 4, 5, 6, 7, 10, 15, 20, 25, 30, 35, 40, 45, 50$. To obtain an empirical formula δ_e for the numerical results δ , we assume that the ρ dependence of δ_e given by

$$\delta_e = \frac{a(1 - \rho)}{1 - b(1 - \rho)}. \quad (5.2)$$

The parameters a and b are determined by the least square fit to δ for each value of Q . Figure 9 shows δ_e so determined along with δ at $Q = 2, 10$ and 50 .

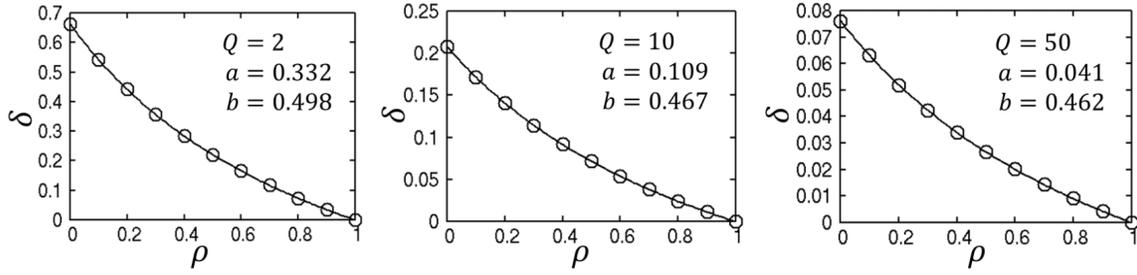


Fig. 9. The results of least squares fitting of δ_e for three values of Q .

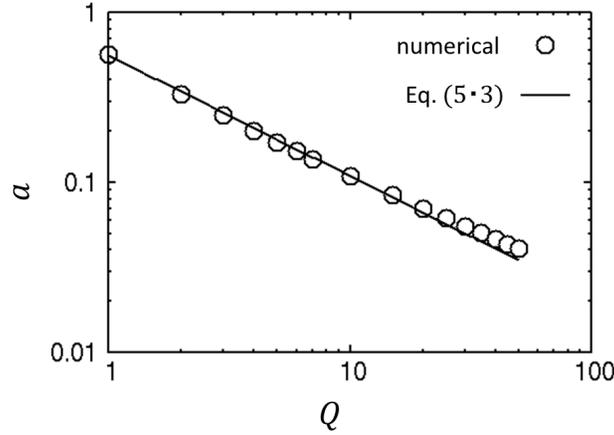


Fig. 10. The Q -dependence of the parameter a in Eq. (5.2).

We then determine the Q -dependence of a and b . In Fig. 10, we see that the fitted results for a have power-law behavior as a function of Q , which is explicitly given by

$$a(Q) = 0.56Q^{-0.71}. \tag{5.3}$$

In constant, the parameter b shows no visible dependence on Q as shown in Fig. 11. Thus we set

$$b = 0.48. \tag{5.4}$$

Finally, Eq. (5.2) with Eqs. (5.3) and (5.4) provides us with an empirical formula for the autocorrelation correction.

To evaluate accuracy of δ_e quantitatively, we calculate the absolute and relative errors defined as

$$\Delta \equiv |\delta_e(\rho, Q) - \delta(\rho, Q)|, \tag{5.5}$$

$$\delta \equiv \frac{|\delta_e(\rho, Q) - \delta(\rho, Q)|}{\lambda_m(\rho, Q)}. \tag{5.6}$$

The results are summarized in Table I.

Table I. $\bar{\Delta}$ and Δ_{\max} refer to the average and maximum of Δ . The same notations are used for δ .

$\bar{\Delta}$	4.2×10^{-3}
Δ_{\max}	$7.9 \times 10^{-2} (\rho = 0, Q = 1)$
$\bar{\delta}$	1.9×10^{-3}
δ_{\max}	$1.5 \times 10^{-2} (\rho = 0, Q = 1)$

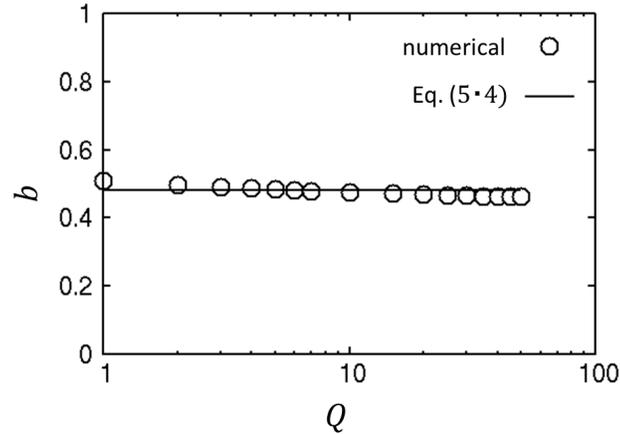


Fig. 11. The Q -dependence of the parameter b in Eq. (5.2).

The absolute errors of the empirical formula are less than 10^{-2} in average. This form is very simple, however, this empirical formula reproduce the numerical results quite well. As an application of this result, we propose a new criterion.

§6. New criterion for principal components within AR(1)

Table II. Comparison of the criterion for the principal component analysis.

Q	λ_+	$\lambda_{\text{new}}(\rho = 0)$
2	2.91	3.59
3	2.49	2.99
4	2.25	2.66
5	2.09	2.44
6	1.98	2.28
7	1.90	2.17
8	1.83	2.08
9	1.78	2.04
10	1.73	1.94

We propose the following new criterion

$$\lambda_{\text{new}} \equiv \lambda_+ + \delta_e(\rho, Q). \quad (6.1)$$

δ_e represents the autocorrelation correction.

Table II compares λ_{new} with λ_+ for various Q for $\rho = 0$. The critical value λ_{new} is appreciably larger than λ_+ over the parametric ranges as covered in this table. This new criterion may be especially useful when the time series data

contains strong autocorrelation such as SLP data.

§7. Conclusion

The principal component analysis based on the RMT is a powerful tool to extract statistically significant correlations in multivariate data, where the RMT works as a null hypothesis for the procedure. However this analytical method can not distinguish between the cross-correlations and the autocorrelations; we are not interested in the latter. Therefore we like to develop a method to separate the cross-correlations from the autocorrelations in the principal component analysis.

In this paper, we numerically investigated the maximal eigenvalue distribution for AR(1) model, where the time series data contain autocorrelations. The autocor-

relation correction δ to the maximal eigenvalue is given by Eq. (5.2) with Eqs. (5.3) and (5.4). This empirical formula reproduces the numerical results quite well. As an application of the formula, we finally proposed the new criterion λ_{new} to single out genuine correlations in the principal component analysis. This new criterion thus takes accurate account of the autocorrelation correction to the RMT prediction. The new criterion is especially useful when the time series data contains the autocorrelation $-1/2$ in adjacent elements.

Acknowledgements

The authors thank the Yukawa Institute for Theoretical Physics at Kyoto University, where this work was completed during the YITP workshop (YITP-W-11-04) entitled “Econophysics 2011 — The Hitchhikers’s Guide to the Economy”. This work was also supported, in part, through *the Program for Promoting Methodological Innovation in Humanities and Social Sciences by Cross-Disciplinary Fusing* and Grants-in-Aid for Scientific Research (B), No. 22300080 (2010–12), provided by the Japan Society for the Promotion of Science and by the Ministry of Education, Science, Sports and Culture, respectively.

References

- 1) L. Laloux, P. Cizeau, J. P. Bouchaud and M. Potters, Phys. Rev. Lett. **83** (1999), 1467.
- 2) M. S. Santhanam and P. K. Patra, Phys. Rev. E **64** (2002), 016102.
- 3) V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr and H. E. Stanley, Phys. Rev. E **65** (2002), 066126.
- 4) A. Utsugi, K. Ino and M. Oshikawa, Phys. Rev. E **70** (2004), 026110.
- 5) D. H. Kim and H. Jeong, Phys. Rev. E **72** (2005), 046133.
- 6) V. Kulkarni and N. Deo, Eur. Phys. J. B **60** (2007), 101.
- 7) R. K. Pan and S. Sinha, Phys. Rev. E **76** (2007), 046116.
- 8) Y. Arai, K. Okunishi and H. Iyetomi, Intelligent Decision Technologies, SIST **10** (2011), 557.
- 9) H. Iyetomi, Y. Nakayama, H. Aoyama, Y. Fujiwara, Y. Ikeda and W. Souma, Phys. Rev. E **83** (2011), 016103.

Issues for SME Credit Information DB Institutions and Expectations for the Econophysics^{*)}

— *Scientific Economic Policies for Avoiding Moral Hazard* —

Toshinori TANABE

CRD Association,

Tokyo Tatemono Muromachi Bldg. 8th Floor,

4-3-18 Nihonbashi Muromachi, Chuo-ku, Tokyo 103-0022, Japan

The CRD database, which has been accumulating financial data on SMEs over the ten years since its founding, and has gathered approximately 12 million records for around 2 million SMEs, approximately 3 million records for somewhere around 900,000 sole proprietors, also collected default data on these companies and sole proprietors. The CRD database's weakness is anonymity. Going forward, therefore, it appears the CRD Association is faced with questions concerning how it will enhance the attractiveness of its database whether new knowledge should be gained by using econophysics or other research approaches. We have already seen several examples of knowledge gained through econophysical analyses using the CRD database, and I would like to express my hope that we will eventually see greater application of the SME credit information database and econophysical analysis for the development of Japan's SME policies which are scientific economic policies for avoiding moral hazard, and will expect elucidating risk scenarios for the global financial, natural disaster, and other shocks expected to happen with greater frequency. Therefore, the role played by econophysics will become increasingly important, and we have high expectations for the role to be played by the field of econophysics.

§1. Introduction

I would like to begin by expressing my appreciation for this opportunity to follow in the footsteps of Shigeru Hikuma, a past executive director of the CRD Association,¹⁾ who spoke at the 2nd *Econophysics Symposium*. Also representing the CRD Association, it is my honor to present to this 5th meeting of the *Econophysics Symposium* my thoughts on issues facing SME credit information database institutions and expectations for Econophysics.

First, I would like to tell you a little about myself to give you some idea of why I have been chosen to give an address on this occasion. After training for about 25 years in financial and economic analysis and policy research at the Bank of Japan, I joined a think tank, the Economic Research Center of the Fujitsu Research Institute, where I performed research on the creation of new industries in Japan. I focused in particular on solutions for IT, environmental, energy, and agricultural issues, and on necessary infrastructure, like industrial finance and others.

Shortly after the establishment of the CRD Association, while I was still at the Fujitsu Research Institute and just as the Association was getting on its feet, I engaged in deep discussions with the Association's past executive director, Mr. Hikuma, about this new SME credit risk database institution should be managed

^{*)} This article was written based on my invited talk in the 5th Econophysics Symposium.

and organized. Having begun my relationship with the Association in this manner, I was appointed one of its directors this past June. In this position, my role is to help the CRD Association, which just celebrated the 10th anniversary of its founding in March 2011, answer certain questions regarding the coming ten years. Among these are questions about the kinds of new data the Association can gather, the kinds of new services it can provide to its member credit guarantee associations and private-sector financial institutions, and how the Association can help to solve economic policy problems facing our country. Therefore, today, I would like to begin by talking about the status of the Japanese economy in terms of the three key economic issues that are the focus of work by the American economist Paul Krugman — namely, growth, employment, and income redistribution. In my discussion, I will discuss conditions currently facing SMEs, main players which account for approximately 70% of employment in Japan, and I would be very pleased if what I have to say is of some use to you in that most important of research tasks, the selection of a research theme.

§2. Long-term growth trend of the Japanese economy

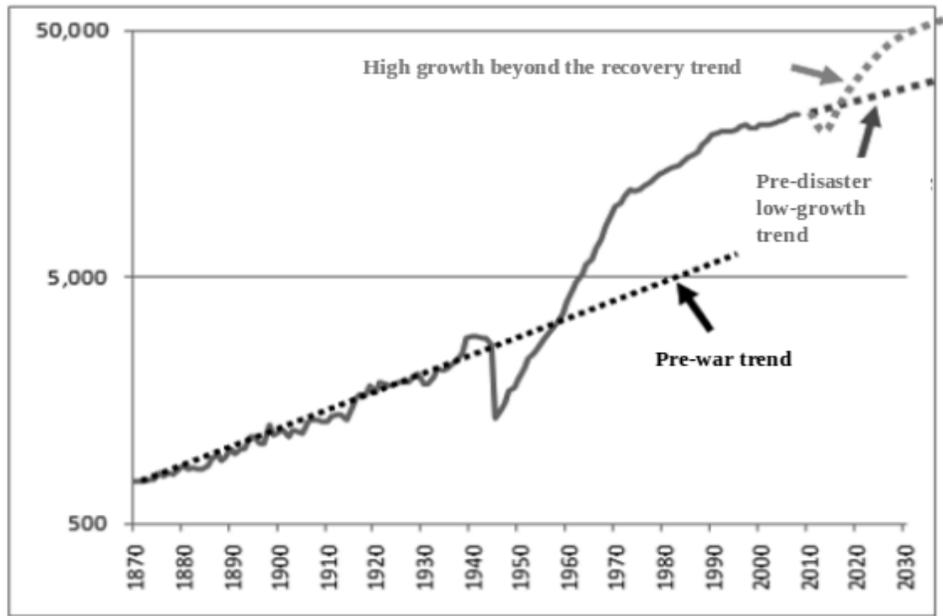
I would like to start by looking at the long-term growth trend of the Japanese economy (see Fig. 1). Following a war or major disaster, various countries have in fact seen upticks in their long-term growth rates. In Japan's case, the economy grew at rate of about 3–4% between the Meiji Restoration in 1867 and Second World War, but accelerated to rates as high as around 10% — on a par with what is being seen in China now — after the war. Over the more-than-20 years since the collapse of its bubble in the 1990s, however, the Japanese economy has seen problems with non-performing loans at financial institutions followed by deflation with no end in sight, and, more recently, government debt surpassing 200% of GDP, and intractable problems in government finance and the social security system. What is more, it appears that Japan is quickly heading down the path of “Japanization” in which the world's leading economies are following after Japan. Breaking this situation down into more manageable pieces, the Japanese economy, with a falling birthrate and aging population, is deemed to be experiencing increasingly noticeable impacts from the shrinking of its working-age population, which has been in decline for more than 10 years. Indeed, the economy has recently seen its annual growth potential decline to 1–2%.

Furthermore, considering factors like the relocation of export-oriented companies to overseas locations, which has accelerated in the wake of the Great East Japan Earthquake, the outlook inevitably calls for a prolongation of the low economic growth that has prevailed for more than 20 years since the collapse of bubble economy.

Weighed upon by a high yen, high corporate taxes, strict labor regulations, CO2 emissions regulations, and electricity supply problems, Japanese companies, particularly those of export-oriented, are finding it more and more advantageous to shift production overseas, and that is adding significantly to deflationary pressures.

Against that background, the top issues confronting the SMEs that account for

Japanese Per Capita Real GDP(USD)



Source: Angus Maddison website (<http://www.ggd.net/MADDISON/oriindex.htm>)

“A New Era of High-Growth Surpassing Recovery,” April 8, 2011 Special Column by Yasuyuki Todo, Faculty Fellow, Research Institute of Economy, Trade and Industry

Fig. 1. Extracting benefits from disasters over the long term requires a complete restructuring of Japan’s economy. Long-term economic growth rates for various countries following major wars or disasters have actually exceeded pre-war or pre-disaster growth rates. This phenomenon is believed to be the result of countries taking advantage of the structural devastation they have suffered to remake economic systems and industries in ways that are more growth-oriented.

around 70% of Japanese employment are said to be credit, bank borrowing, and the securing of staffing and successors. On the particularly important matters of credit and bank borrowing, policy measures aimed at streamlining SME financing were significantly strengthened following the Lehman shock in September 2008 and the Great East Japan Earthquake in 2011. After the Lehman shock, for example, initial lending term changes resulted in the reclassification of borrowers “which require special attention” as borrowers of “normal loans” and the application of 100% government guarantees. As for government policies instituted to address the problem of overlapping debt following the Great East Japan Earthquake, some have claimed that the scope of companies eligible for assistance is actually too wide.

In this sense, even the granting of credit to SMEs is lacking in income redistribution fairness, presents a moral hazard, and ultimately increases the likelihood of even greater fiscal inflexibility.

§3. Impasse in income redistribution

With the Japanese economy bogged down by fiscal and social security problems, the threat of Japanization looming over the U.S. and leading European economies, and rising inflation in emerging economies, income redistribution is facing considerably stronger headwinds. In Japan, the lack of a civil consensus on what kind of society Japan should aspire to be and whether government should be big or small has been a cause for the failure to set a clear direction for income redistribution, and the country's approach to economic policy is being ad hoc. Regarding income redistribution, in particular, it is important to secure trust in the fairness of the related rights and responsibilities of citizens is critical. Furthermore, it must be said that transparency in income redistribution holds the key to the question of whether a society is best served by shareholder capitalism, on the one hand, or public interest capitalism on the other, the topic to be addressed in tomorrow's *Econophysics Symposium* session.

On this point, I would like to mention that even in America, a country that highly values the protection of privacy, punishing tax evasion is seen as more important than protecting privacy.

President Nixon, for example, resigned ostensibly because of the Watergate wire-tapping. The truth, however, is that he resigned because of tax evasion he committed prior to the Watergate incident and, threatened by the disclosure of that activity, was left with no option other than to resign.

Why is tax evasion viewed so harshly? When the Pilgrims traveled from Europe to the American continent in search of a new home, they established communities in various locations in the continent, and community members were obliged to pay taxes to provide the funds necessary for building schools and other social infrastructure. Tax evasion was viewed as a betrayal of the community, the greatest affront to fairness. Nixon's evasion of taxes, therefore, effectively disqualified him from representing the country as its president. In contrast, it has been pointed out that in ancient nations like Egypt, China, and Japan, tax evasion can be seen as an exercise of the right to oppose an oppressive government. Even the right to privacy, therefore, is regarded differently depending on the country, and I believe that unique econophysics perspectives based on such nation-specific social principles will take on greater importance.

§4. Strengths and weaknesses of CRD data

Now, I would like to discuss the current status of the CRD database, which has been accumulating financial data on SMEs over the ten years since its founding. Counting a set of annual financial statements as a single record, the database has gathered approximately 12 million records for around 2 million SMEs, approximately 3 million records for somewhere around 900,000 sole proprietors such as apartment building operators, and collected default data on these companies and sole proprietors. Regarding SMEs, in particular, it is estimated that the CRD database covers 70–80% of the SMEs known to national tax authorities. The database is also dis-

tinguished by its collection of financial statements and default information for sole proprietors.

Figures like Professor Hideaki Aoyama and Professor Yoshihisa Fujiwara, both of whom are experts on financial data for SMEs in various countries across the globe, have praised the CRD database not only for being among the largest databases of its kind in the world but also for leading the world in the number of financial data items it accurately tracks.

The second largest SME financial database after the Association's is in France, and, as a point of reference, I would like to say a few words about why France's has the strengths on its collection of database.

Collection of the CRD's data is accomplished not via specific legal authority, but through voluntary provision by financial institutions and other parties. On this point, it differs vastly from its French counterpart. France, which established the world's first nation state, has included in its constitution the obligation for companies, including SMEs, to disclose financial data for the benefit of not only investors but all stakeholders, for purposes like protecting the rights of workers. Under this national ethos, the Banque de France is the institution responsible for collecting data. Compared to France's SME financial database, the CRD database is larger in terms of both the number of companies covered and the number of financial items tracked.

But, if overall scale and detail are its strengths, the CRD database's weakness is anonymity. Going forward, therefore, it appears the CRD Association is faced with questions concerning how it will enhance the attractiveness of its database — whether this will be done by beginning to gather data for named companies by itself or by linking to company-identified data owned by other database institutions, and whether new knowledge should be gained by using econophysics or other research approaches.

§5. Knowledge beginning to emerge from other databases

We are beginning to see the emergence of knowledge from the application of econophysics approaches to data from the databases of other Japanese institutions.

Professor Tsutomu Watanabe of the University of Tokyo, for example, has discussed²⁾ how, in a joint project with Hitotsubashi University and Teikoku Databank, a database was used to elucidate a transaction network involving 500,000 Japanese companies for whom data on numbers of transaction partners and key transaction partners was recorded. Randomly selecting one company from among those 500,000, following each sales transaction by counting only the first link in the transaction chain as a direct-sale customer, the randomly selected company would be connected to nearly all 500,000 companies in the space of four links. With the first link, the company would be connected to around 200 other companies and with the second the number of companies connected to leaps to more than 10,000. The third link increases the number of connected companies to 230,000, approximately half of the companies in the network, and it is understandable if the management of the original randomly selected company does not know who these indirect customers are.

The fourth link connects the original company to nearly all of the companies in the network. That any two companies could be connected to each other within the space of four transaction links goes beyond the imagination of most company managers. In the field of network science, this kind of close linking of members is referred to as the “small world” phenomenon and a famous example is relationships among acquaintances. The Watanabe group’s research results show that business society, like human society, is a small world.

The fact that companies are closely connected carries important policy implications regarding system for inducing economy-wide changes. If a unique shock — in other words, one unique to a certain company — propagates and gives rise to economic changes, policies that focus on the company where the shock begins, or on companies closely related to it, would be most effective. If the origin of an economic change is a shock unique to a particular company, it is only natural — and indeed desirable — that policy management will take on a unique hue. Furthermore, if economic fluctuations arise from individual shocks, the “hub” companies with large numbers of customers are key to shock propagation and supporting these companies to stop the chain reaction of the economic fluctuation is essential. Policies supporting hub companies have already been applied to financial institutions deemed “Too Interconnected to Fail”, a pun for “Two Gig to Fail”. The concept behind these policies has been to rescue financial institutions with large numbers of customers before the wave of damage become significant.

Policies that are highly targeted carry the risk of moral hazard. However, given that company and financial institution managers do not know the full extent of the networks they are a part of, it becomes necessary to incorporate in macro policy management the concept of stopping the wave of damage coursing through the transaction network.

This implies the urgent need to construct as an element of policy infrastructure a database that can be used to monitor the transaction networks of financial institutions and companies.

This part of my discussion indicates that the role played by econophysics will become increasingly important. As for SME financial data institutions, it is important for us to gather new types of data that better respond to the macro policy management needs I have discussed, and to research on development of services that more accurately match the needs of member financial institutions and credit guarantee associations, so that they will better respond to data provision incentives.

§6. Issues facing the CRD Association and expectations for econophysics

For the CRD Association’s database to offer even greater added value going forward, it will be important to obtain data for named companies or create linkages to such data owned by other institutions. With data including the names of companies and their customers, financial institutions will gain a picture of the full extent of the networks of which their customers are a part in the manner illustrated by the research of the Watanabe group. And that, by enabling measures like carefully

selecting sound customers, can make it possible to properly prepare for the shocks that inflict damage through transaction networks. It can also contribute to the avoidance of moral hazards. Speaking on behalf of the CRD Association, I would like to say that we have high expectations for the role to be played by the field of econophysics. For instance, by analyzing the CRD's enormous, richly detailed database of financial data, the occurrence of a shock, much like that of a tsunami, can be met with forecasts of the industries and regions that will be effected, the extent of the impacts, and the amount of time necessary for impacts to attenuate. In other words, I believe it would be possible to anticipate the extent of sales and default impacts.

The CRD Association gathers financial statements for its database by collecting them from its members. Statements for any particular year are accepted over about a two-year period and up to one million data records can be collected that period. Financial statements for 2008 (account closings occurring anywhere from January through December), for example, were gradually collected over 2009 and 2010, and eventually exceeded the one million mark. I expect that this data can now be used, for example, by dividing them into duodecimal data and determining what sales and other P/L items would have looked like on a monthly or quarterly basis over the two-year collection period, not only as a historical data of two years ago, but also to forecast the course and degree of the impacts of shocks as they propagate through sales fluctuations and other phenomena beginning in the month following the shock. The closings for SME financial statements collected by the CRD, unlike those for middle-market and large companies, which concentrate in March and September, are more dispersed throughout the year, and that makes me optimistic that some clever processing could produce monthly closing data.

I also believe it could be possible to process CRD data to issue information like the Bank of Japan's Tankan, which is based on quarterly surveys of conditions at around 10,000 companies including SMEs to monitor quarterly change in their business performance. Assuming that a research using such processing is possible, it should also be possible to measure impacts on items like SME sales on a monthly basis, beginning with the admittedly reduced data volume immediately following an event like the Lehman shock (September 2008) or the Great East Japan Earthquake (March 2011) and moving forward to identify patterns in performance changes as the volume of data grows. As for the propriety of such processing, that would appear to be a topic tailor-made for the field of econophysics. If CRD data gain in value through analyses like what I have described, so might member financial institution and credit guarantee association understanding of the need to expand CRD data with not only data for named companies, but also data on their named customers. My thinking may appear to involve a "chicken-or-egg" conundrum, but I would like to ask for your cooperation in applying econophysics to the task of gaining recognition for the indispensability of CRD data for financial institutions, among others, in elucidating risk scenarios for the global financial, natural disaster, and other shocks expected to happen with greater frequency. Viewed from a different perspective, if CRD data is enhanced and it becomes possible to use data on named companies to identify transaction relationships — in other words, draw a clear picture of the

network structure connecting companies to one another — econophysical analysis will also have taken significant leap forward, and it can be expected that the CRD Association and the field of econophysics will thrive together.

§7. Conclusion

We have already seen several examples of knowledge gained through econophysical analyses using the CRD database covered in the fiscal 2011 white paper on SMEs in Japan,³⁾ and I would like to conclude my address by expressing my hope that we will eventually see greater application of the SME credit information database and econophysical analysis for the development of Japan's SME policies and their own mutual benefit.

Acknowledgements

The author would like to thank the organizers and the audience at the *YITP Workshop on Econophysics: Econophysics 2011*.

References

- 1) S. Hikuma, *10 years of CRD: 2001–2010* (CRD Association, Tokyo, 2011), in Japanese.
- 2) T. Watanabe, *Keizai Kyoshitsu*, column in the Nikkei Shinbun Newspaper (October 7, 2011).
- 3) SME Agency, *2011 White Paper on Small and Medium Enterprises in Japan*, Sections 1 and 2, Chapter 1, Part II (July 2011), http://www.chusho.meti.go.jp/pamflet/hakusyo/h23/h23_1/2011hakusho_eng.pdf

New Systems for the Realization of a Comfortable and Prosperous Society in the 21st Century

George HARA^{*)}

*Alliance Forum Foundation,
Mitsuidaini-bekkan 7F, Nihonbashi-Hongoku-chou 4-4-20,
Chuo-ku, Tokyo 103-0021, Japan*

In the wake of the Great Northern Japan Earthquake and Tsunami, the Japanese people have impressed the world with their strong spirit of cooperation and optimism in the midst of crisis. Japan is now poised to lead the world toward a new economic paradigm and into the age of Public Interest Capitalism which can replace short-termism and shareholder centric capitalism. As the legacy economic and social systems that have dominated advanced countries reach a critical juncture, Japan now stands at the crux of a great opportunity. In the past, Japan's success was driven by the skillful exportation of "Hard" (manufactured goods) and "Soft" (entertainment/culture) products. In the coming era, the key to Japan's success lies with our ability to leverage our inherent values and disseminate new economic and social systems that trigger innovation, create new industries, and promote a more comfortable, balanced, and sustainable form of global development.

§1. Japan is positioned to signal the future course for global society

Upon returning to Japan in April after the Great Eastern Japan Earthquake, I visited the affected areas in the Tohoku region and was completely shocked by what I saw. Words cannot express the sorrow that I initially felt upon seeing the destruction and tragedy up close with my own eyes. Moreover, since my return to Japan I have been touched to see people here join together to conserve power under the energy constraints caused by the disaster. Frankly speaking though, without fear of misunderstanding, upon reflection I feel intuitively hopeful that "Japan has been given an opportunity to lead the world towards a new paradigm shift".

What other advanced country could be as successful as Japan when faced with the challenge of making the best use of inventiveness to get by with scarce resources? In the aftermath of the earthquake and tsunami disaster here, the people of the U.S. and other foreign countries have expressed their sincere respect for the efforts and spirit of the Japanese. The earthquake disaster was a horrible tragedy. However, it can become an opportunity for us to regain our strength. All of us living today must keep this firmly in our minds.

Yet many people here are still trapped in outdated perspectives and remain negative. Many thoughtleaders and specialists cautiously warn that "a power shortage of say 10% in Kanto region might lead to GDP decline of more than 4% in the summer and 2% year-round". Their logic is that "if corporations and citizens make great efforts to save electricity, power consumption will decrease effecting production and

^{*)} Chairman of the Board. Also, Group Chairman, DEFTA Partners; Senior Advisor, Japanese Ministry of Finance; and Ambassador, IIMSAM United Nations Economic and Social Council (retired).

consumption driving GDP down further”. Further, they reason that “as profits decrease, stock prices will decline and so too assets”. I feel that people who think only in this way are “brainwashed” in a sense. Indeed, such reasoning is naturally viewed from the lens of “economic common sense”. However, what’s really wrong with this if we ultimately create a more comfortable society as a result of, for example, a 10% reduction in our power usage? We must keep in mind that world population has exceeded 7 billion people and developing countries like China and India are trying to catch up with the advanced nations at a furious pace. If all these countries become an energy consumer at the same level as the U.S. the earth won’t survive. In this way, the tragic earthquake and tsunami has given Japan a rare opportunity to demonstrate to other advanced countries a new more sustainable direction.

It is time that we realize that the overuse of resources has nothing to do with human happiness. From now on, we must build a more resource-efficient, sustainable and cyclical society. This major turning point is now evident as most Japanese are beginning to become keenly aware of this. I feel that this will become the first step for Japan to become a “great nation of hope”. The 20th century was an era in which the whole world longed for a U.S.-style mass consumption society. But from now on, a different logic will start in full swing from this point on. At this major turning point, the rationale exhibited by thought leaders and specialists whose only worry is the “decline of GDP” is totally out of touch with the times. The real wisdom lies in the normal common sense reaction of the public to voluntarily turn off electricity switches and cherish things. Going forward, what is most important is for us is to make use of this wisdom and align our systems and policies with this common sense to create a more comfortable society.

§2. End of shareholder capitalism

ROE (Return On Equity) has been respected as an index highly correlated with stock price, but ultimately it is really only a mere “fad”. In a sense it is no different than in the dot-com era, when venture firms in Silicon Valley were valued at ten thousand or even a hundred thousand dollars per user. This too was just another short-lived fad to boost shareholder value in the short term. Japan experienced a similar trend in the form of the “myth of ever-rising land prices” during our era of a bubble economy.

Consider for example, if instead we measured corporate value based on a new comfort index that examined *Sustainability, Fairness or Innovation?* When society comes to realize that these factors are much more important to realizing stable mid to long-term return on investment and enhancing innovation, such a valuation system will surely replace ROE as a more effective investment indicator. The concept of this index is currently being researched to support a new theory of theoretical economics termed *Public Interest Capitalism*. Research into the theory of Public Interest Capitalism began with several years of joint research by a team from the Harvard Graduate School of Management. Beginning this July, it will be expanded through joint research by the Public Interest Capitalism Division of the Alliance Forum Foundation, the Faculty of Economics at Tokyo University, and the Faculty

of Science at Kyoto University.

In 2008, many investors became painfully aware of how unreliable ROE and ratings like AAA are after the Lehman Shock and subsequent financial crisis. These investors are eager for the arrival of new indicators of value that can serve as more effective rationale for investment decisions. If the investments increase into companies that carry out business based on the tenets of *public interest capitalism*, this will drive mid- to long-term innovation for our future. Today's dominant "shareholder-centric capitalism" which pursues short-term profits will finally ultimately morph into financial capitalism and the game will be over when society has only a few winners and a lot of losers. However, I feel that as our beliefs and understanding of how our corporations are creating value change, the stock market itself will reflect this and naturally change as well.

§3. Budget deficits created by GDP supremacy

The same holds true for the supremacy of GDP as a global indicator associated with measuring societal progress. Consider what will happen if people around the world learn about the Japanese concept of "Mottainai" (cherishing things) feel empathic towards it, and this concept becomes a mainstream idea. Surely, no society in the world can disagree with this concept once they have grasped its importance. Certainly, GDP is a dominant indicator in use around the world. However in recent years, each country has printed more and more money and this excessively liquidity has flowed into developing countries. While the ruling class gained huge wealth in these countries, a 'phantom' boom economy was created at the same time in which a vast majority don't feel that their lives have been enriched. Considering the reality of this situation, how effective is GDP as an indicator alone?

Let's look at another example of the deficiency of GDP as an indicator and driver for decision-making. Let's say that there is a road with a durability of more than 100 years but costs 10 billion yen and a road with durability of only one year but costs only 100 million yen. If we consider only GDP value, "Building of a road that costs 100 million yen every year" is better in terms of GDP. As accidents will frequency occur because of the fragility of the less durable road, extra costs will be incurred for medical care, the purchase or repair of cars, road repair, environmental expenditures, etc. In this way, the effects of building the road with one year durability will actually serve to increase GDP to a greater extent, but surely making such a decision in real life is irresponsible and ridiculous.

Some readers may counter this example by arguing that "If GDP does not increase, tax revenue will not increase and the budget deficit cannot be eliminated". To such a comment, I would rather answer that "Isn't it in fact GDP supremacy that is partially responsible for the increase in our budget deficits?" Consider Alzheimer's disease for instance. Effective treatment for these diseases has not been established yet so society requires not only treatment but also nurses and nursing homes to care for patients. Alzheimer's disease is a great misfortune to the nation, but looking at it in terms of GDP only, like the above-mentioned bad road case, the ridiculous comment goes again that "Alzheimer's disease is contributing to an increase in GDP".

And the government will have to invest a huge amount of medical expenses for that portion of social security.

Moreover, there is another side to this issue. In many cases, government spending actually increases for the portion of GDP boosted by business activities. One example of this is the Bottom of Pyramid sales approach in which specialized products are sold to low income consumers. Many companies are actually selling products at lower prices by repackaging small quantities into plastic bags. This approach has become popular among firms in western countries that are now marketing to the developing world. Some European food manufacturers have been increasing their sales in rural areas of developing countries by packaging chocolates in small plastic bags and selling them to consumers there. However, in many cases the consumers are not accustomed to recycling and end up scattering the plastic bags all over as litter. Anyone who visits developing countries will notice that there is an abundance of trash created from products brought in from developed countries. In a cyclical society, all food was returned to the soil and there was no need for garbage disposal. To counter this change, companies are actually hiring NGOs to collect garbage using a portion of the profits because they cannot profit if they collect the plastic garbage by themselves. Yet it goes beyond trash and the environment, these same chocolates end up causing cavities creating a need for more toothpaste and toothbrushes and even dental services. Companies operating in this manner will increase sales further and boost GDP, but as we have seen government expenditures will increase as garbage disposal and health care costs rise.

Let us now turn our attention to the common conception regarding electrical power. It is assumed that cheap electricity from nuclear power has been supporting the development of the national economy, but I wonder if this is genuinely true considering the tremendous size of the costs to dispose of nuclear fuel and the expenditures and compensation costs for accidents like the one at Fukushima. Aren't there major hidden costs which in fact have required a huge amount of government expenditure behind the scenes in cases where GDP growth has been supported by nuclear power?

Now consider the following. What if a magic bullet for Alzheimer's disease is developed today? Medical care facilities, medical care expenses and the burden on the family may be reduced significantly and GDP may decline. However, it would be fair to say that a great number of people would be much happier. Furthermore, national medical expenditure of social security will be significantly reduced by the effect of this magic bullet. What if companies develop bags that safely and naturally disintegrate into the earth based on the tenets of *public interest capitalism*? Mitsubishi Chemical Holdings Corp., a company with a corporate philosophy that aims to achieve 'global comfort', has done just that by developing plastic materials that disintegrate and return to the earth in just a few weeks. If such materials are correctly evaluated and widely used, the government expenditure for garbage disposal could be significantly reduced.

We can create a truly sustainable society by raising "power-saving consciousness" through adoption of a system in which the higher the household use of electricity, the higher the bill and by further developing photovoltaic, wind-power and geothermal

generation while at the same time working to develop new nuclear power technologies like thorium molten-salt reactors. Moreover, if we can promote wide use of solar power generation by households, as the demand for electricity peaks in summer daytime, power generation under the midsummer sun will greatly help to cut peak demand while reducing the need for greater capacity of power companies. To promote this we must encourage purchase of solar power generation facilities by households and small-scale businesses by utilizing eco-points or investment tax cuts to create a flat-rate buyback program. In this regard, however, a flat-rate buyback program is not necessary for a large-scale commercial solar power plant as they are not related to peak cut and should operate based on the principle of the market mechanism. Such efforts may decrease the GDP lower than the government policy to “increase the proportion of the current nuclear power generation system by 50% by 2030”, but the people would be more happy.

I don't want to be misunderstood by saying that we should deny GDP supremacy and I don't mean that we should return to the living in the Dark Age. I do not believe that we should stop nuclear power generation immediately by overestimating the natural energy that accounts for only about 1–2% of total power generation. However, what I'm trying to say is that, from now on, we must focus on contributing to a more sustainable and prosperous society by designing a social framework for the long-term based not only on economic growth but on some other value such as, *Sustainability, Fairness and Innovation*.

§4. Activities of AFDP in Africa and Asia

To build a more sustainable and prosperous society, the examples which I have given may seem like a dream. But reality can change, and when it changes, it will do so suddenly and dramatically. Let me change gears and talk now about developing countries where change too is happening very quickly.

The Alliance Forum Foundation Development Program (AFDP) has a project to deploy Spirulina (a microalgae high in protein content) to eradicate malnutrition in Africa and realize the world's leading advanced distance learning system (using XVD video communication technology) in the Republic of Zambia. In addition to these projects, AFDP is working to advocate the use of microfinance (small, uncollateralized loans to low income populations) to help create a new financial system to grow the middle class in Africa. Through these efforts, AFDP has attracted attention and drawn interest in the concept of *Public Interest Capitalism* from presidents and politicians in South African countries.

Of late, our foundation has seen more and more serious interest in the concept of *Public Interest Capitalism*, especially given the unstable political situation in North Africa. The countries from North Africa through the Middle East introduced capitalism along with the democracy, but the gap between rich and poor expanded within these countries. Since many of these countries were previously dictatorships or monarchies that were allies of western world, they had been back on safe ground for a long time without these contradictions being pointed out. Under this system though, those in power reigned as major shareholders in the companies with the

power of patronage. The strain of Shareholder Capitalism that gives priority to the shareholders' interest has now become the major cause of democratic revolutions of late. To ensure that upheavals like that in the North Africa region do not happen in our own country, we need a model of capitalism that can truly enrich the middle tier of society and avoid the pitfall of creating extreme income inequality. This is the very foundation of Public Interest Capitalism.

Last year, I requested that the governor of the Central Bank of Zambia appoint two talented young executives to attend the Financial Services for the Poor Professional Development Programme in microfinance organized by our foundation in Bangladesh. This specialist course was established in 2009 and has at present produced 86 graduates in total. These graduates are made up of professionals and graduate students who come from around the world to deeply study the practice of sustainable microfinance and small business development in developing countries. Unfortunately, the many of the financial institutions prevalent today operate under Banking and Securities Industry Acts that were originally introduced by the UK in the colonial era. These existing systems of finance only enrich the wealthy and cannot enrich the poor by their very design.

To alleviate poverty, resources are required. The poor require funds to start business and support their self-dependence. However, often the impoverished and low income population cannot obtain loans from traditional banks. To that end, we need to institutionalize a new financial system for this purpose. The two staff from the Zambia Central Bank who attended the Financial Services for the Poor Professional Development Programme learned this in Bangladesh. As a result, a banking institution that is able to do humane finance business will be born in Zambia next year under a new system. The Alliance Forum Foundation expects to support this trend in other countries in Africa, Asia and Latin America.

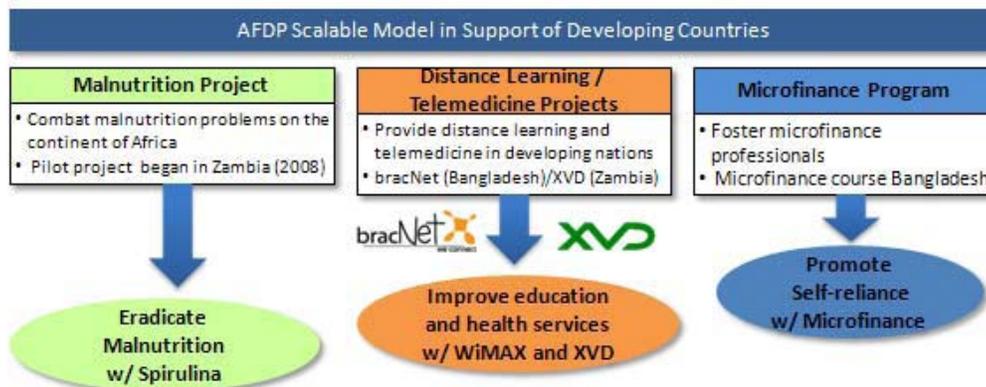


Fig. 1. The AFDP Scalable Model focuses on malnutrition, education/health and economic self-reliance.

§5. Which county will lead the world going forward?

Japan has been very successful in exporting made-in-Japan “Hard (products)” as well as “Soft” such as comics, cartoon films and cuisine culture. But from now on, the time will come for our country to export “Systems (institutions)”. At present, various “systems” from Western countries are prevailing around the world including the schooling system, legal system and parliamentary system. Most of them have been “exported” by Western powers since the age of imperialism. The same goes with the compliance system, asset-impairment accounting and market-value accounting rules, even today. However, these systems are now fraying at the edges here and there. Both the economic and social situation in advanced countries appears to have reached a critical point. A fury of excess liquidity expanded by financial capitalism shows no sign of slowing down and each country is sinking under a huge budget deficit. Even the succession of various policy instruments looks as if they are continuing “to retreat to the front” to cover up the contradiction. Moreover, it is estimated that the people in developing countries will account for 85% of the world’s population by 2050. These regions are growing stronger, headed by emerging countries, but if these countries earnestly aim for mass-consumption capitalism, the earth will not survive much longer. Belief in economic growth rate and shareholder capitalism can no longer be a viable development model. In what form will it change then?

In order to “export” a system, two conditions must be met. First, whether or not both the economy and technology has reached a critical point. I once wrote that “Economy creates culture, and technology creates politics”. Many people in advanced countries will reach a critical point when they realize that they are not satisfied with “things” even if the economy is prosperous. The fact that we are becoming free from the curse of belief in GDP suggests that a great change in the whole concept of culture will occur anytime in the future.

The reason that a large-scale country, not a small city — such as the ancient Greek cities, could realize democracy was because there was development of transportation technology such as roads and railways as well as communication technology such as printing. Further, the way democracy works has significantly changed along with the spread of radio and television. The development of internet technology will inevitably change the political processes. Technological development that enables free utilization of internet is not even half done yet and when post-computer technology is completed in 10 years to come, the current era of democracy will end and reach the critical point.

The second condition is whether or not the system is something that “people around the world yearn” like all culture, new industry or products prevalent in the world. Japanese people adopted western artifacts with a great deal of curiosity in Meiji era, perhaps because they seemed novel, attractive, interesting and convenient. There must have been more aspects of adopting them actively than grudgingly.

Going back to my point about the “economy reaching a critical point”. the situation applies to all advanced countries including Japan, Europe and the U.S. For the Japanese, who have become aware of the how unsustainable our current systems are after the great earthquake disaster this year, this period is an enormous

opportunity. It was thought often speculated that Japanese, who developed a strong affinity for mass consumption, would not be able to tolerate modest living, but we are now showing different. Such an attitude on the part of the Japanese people is having an impact on people around the world and quietly and deeply resonating with their hearts. Of course, it is not that Japan can necessarily open a new era on its own, but what is important is for us to show the earth a new set of values by taking advantage of this opportunity borne out of sacrifice and given to us by God to support historical transformation. Japan should take the leading role, while grounded in its traditions, to proudly declare new social values, create a system for a “comfortable society” and actively “export” and disseminate these to the world.

§6. Let’s make ideals a reality

If we can develop a new value-based economy, someone may argue that the “Economy would rapidly shrink”. Yet this is wrong. Think of it this way, if countries across the world (advanced and developing) can transition to an economy based on new values, a new frontier will be opened there. In much the same way as the “core technology” of the internal combustion engine created new manufacturing industries such as automobiles, vessels and aircrafts, and then created a wave of service industries derived from these industries such as the travel and transport, new values based on *Public Interest Capitalism* will trigger innovation and create a new phase of economic development.

Marketing strategies which previously fueled peoples’ passion for endlessly purchasing physical products, will be deemed uncivilized and society will shift to associating ‘cool’ instead with continue using good products for a long time. If the accounting system that values investment in long-term research and development (not the pursuit of short-term profit) becomes main stream then technological innovation will rapidly progress at an even higher rate. Further, when people value the attitude to respect sustainability, “useful technology for human well-being”. Such as the afore-mentioned plastic materials that disintegrate into the earth and the thorium molten-salt reactors, will be the first to be developed.

Consider what will happen if developing countries are quick to adopt such a movement? This is quite possible. Consider the case of advanced technologies, wrongly thought of as only adoptable in developing countries after their adoption in advanced nations. However this is simply not true. Take telephone network technology as an example. Fixed telephones have not spread throughout developing countries, but instead have been rapidly overtaken by mobile communication networks. The same thing that happens with technology can happen with values and systems. It may in fact become developing countries that end up realizing the first societies to use only safe nuclear power generation or clean, natural power production technologies. These countries may also very well realize better financial systems that enable people to start businesses with vigor and enthusiasm by utilizing micro-finance and invigorating society with creativity and innovation. What will happen when the current developing countries, which will embrace most of the world population, start to build a society based on the values of Public Interest Capitalism

and introduce appropriate technologies. Although these countries may be currently facing a lack of both energy and food, if a system that enhances the happiness of the poor is introduced widely in society, it will give rise to the possibility of developing countries overtaking more developed nations to become the “most advanced” and prosperous societies. In this way, a new form of demand and associated development will be produced there. By that time, a new definition of growth will be created to replace GDP. A society with high efficiency, less government spending and a high degree of social well-being of the people can be created.

We should disseminate the systems (institutions) that can be the foundation of such a society. It will be a mutually beneficial way for us too. In order to export such systems, the idea (concept/philosophy) and the people to practice it need to be fully integrated. So, Japanese people spread your wings around the world!

§7. Concluding remark

To lead a society that has reached the critical point toward new development, we need to develop innovative ideas and to that end, stimulation from civilizations different from that of the western world will be very useful. Now is the time for Japan to rise from the rubble and tragedy of the big earthquake and serve a major role as the “core” of global interaction for a more comfortable and sustainable world. As Japanese, the philosophy that our society has accumulated over a long period of history, the virtue of gratitude and the shared values to voluntarily conserve and not waste, can become the world’s next “leading edge” movement and spread globally as appealing values to replace those that exist today. Isn’t now the time for us to uphold these values and appeal to this sensibility! It is through human wisdom that our ideals will become a reality.

Coupled Oscillator Model of the Business Cycle with Fluctuating Goods Markets

Yuichi IKEDA,¹ Hideaki AOYAMA,² Yoshi FUJIWARA,³ Hiroshi IYETOMI,⁴
Kazuhiko OGIMOTO,¹ Wataru SOUMA⁵ and Hiroshi YOSHIKAWA⁶

¹*Institute of Industrial Science, University of Tokyo, Tokyo 153-8505, Japan*

²*Graduate School of Sciences, Kyoto University, Kyoto 606-8502, Japan*

³*Graduate School of Simulation Studies, University of Hyogo,
Kobe 650-0047, Japan*

⁴*Department of Physics, Niigata University, Niigata 950-2181, Japan*

⁵*College of Science and Technology, Nihon University, Funahashi 274-8501, Japan*

⁶*Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan*

The sectoral synchronization observed for the Japanese business cycle in the Indices of Industrial Production data is an example of synchronization. The stability of this synchronization under a shock, e.g., fluctuation of supply or demand, is a matter of interest in physics and economics. We consider an economic system made up of industry sectors and goods markets in order to analyze the sectoral synchronization observed for the Japanese business cycle. A coupled oscillator model that exhibits synchronization is developed based on the Kuramoto model with inertia by adding goods markets, and analytic solutions of the stationary state and the coupling strength are obtained. We simulate the effects on synchronization of a sectoral shock for systems with different price elasticities and the coupling strengths. Synchronization is reproduced as an equilibrium solution in a nearest neighbor graph. Analysis of the order parameters shows that the synchronization is stable for a finite elasticity, whereas the synchronization is broken and the oscillators behave like a giant oscillator with a certain frequency additional to the common frequency for zero elasticity.

§1. Introduction

The business cycle is an example of synchronization, and this has been studied in nonlinear physics. In particular, the stability of synchronization under a shock, e.g., a fluctuation in supply, is a matter of interest. Sectoral synchronization was observed for the Japanese business cycle in the Indices of Industrial Production data from 1988 to 2007.^{1),2)} By synchronization, we mean there is a constant phase difference between industry sectors. For instance, if random noise components are removed in an appropriate manner, the business cycle is written as

$$x_i = \sin(\omega t + \theta_i), \quad (1.1)$$

$$\omega = \frac{2\pi}{T}, \quad (1.2)$$

where x_i , θ_i , ω , and T are the normalized growth rate of production for sector i , the phase of the business cycle for sector i , the common angular frequency, and the common period of the business cycle, respectively.

We consider an economic system made up of industry sectors and goods markets. Various types of shock may occur in the economic system of interest. If production

in a specific sector suddenly decreases significantly, the imbalance of demand and supply may destroy the synchronization. However, if the price of the goods in the market is quickly adjusted, a sudden change of production in the sector will be absorbed by consumers as a decrease in demand in coupled sectors.

In this paper, we describe a synchronized coupled oscillator model with fluctuating goods markets. The effects on synchronization with respect to the magnitude of shock, price flexibility, and network topology are studied using the coupled oscillator model. The paper is organized as follows. In §2, existing theories are briefly reviewed. In §3, we explain the formulation of our coupled oscillator model, and we describe our analysis of sectoral synchronization in §4. Finally, §5 presents our conclusions.

§2. Existing theories

In this section, existing theories for the synchronization of economic systems, power systems, and physical and biological systems are reviewed. It is noted that the first two are large-scale manmade systems that exhibit similar dynamical behavior in terms of synchronization.

2.1. *Economic systems*

The business cycle is observed in most of industrialized economies. Economists have studied this phenomenon by means of mathematical models, including various kinds of linear non-linear, and coupled oscillator models.

Interdependence, or coupling, between industries in the business cycle has been studied for more than half a century. A study of the linkages between markets and industries using nonlinear difference equations suggests a dynamical coupling among industries.³⁾ A nonlinear oscillator model of the business cycle was then developed using a nonlinear accelerator as the generation mechanism.⁴⁾ In this paper, we stress the necessity of nonlinearity because linear models are unable to reproduce sustained cyclical behavior, and tend to either die out or diverge to infinity.

However, it is noted that a simple linear economic model, based on ordinary economic principles, optimization behavior, and rational expectations, can produce cyclical behavior much like that found in business cycles.⁵⁾ An important question aside from synchronization in the business cycle is whether sectoral or aggregate shocks are responsible for the observed cycle. This question has been empirically examined, and it was clarified that business cycle fluctuations are caused by small sectoral shocks, rather than by large common shocks.⁶⁾

As the third category of model, coupled oscillators were developed in order to study noisy oscillating processes like national economies.^{7),8)} Simulations and empirical analyses showed that synchronization between the business cycles of different countries is consistent with such mode-locking behavior. Along this line of approach, a nonlinear mode-locking mechanism was further studied that described a synchronized business cycle between different industrial sectors.⁹⁾

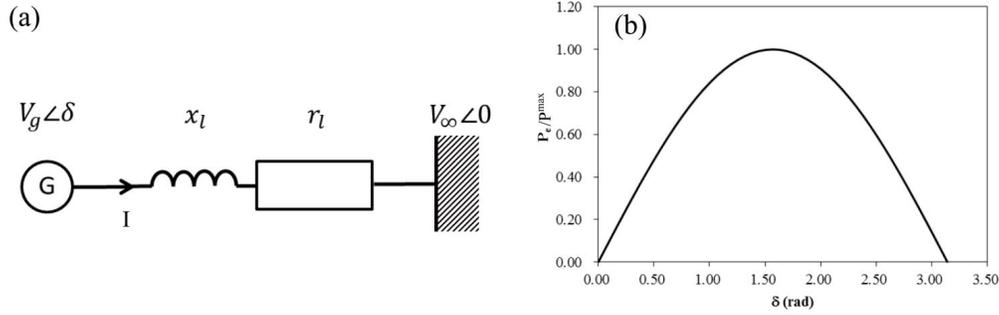


Fig. 1. Single machine connected to infinite bus and synchronizing force.

2.2. Power systems

Power systems made up of many synchronous machines and transmission lines exhibits synchronization. The stability and control of this synchronization of the machines has been studied in power system engineering.¹⁰⁾

The dynamical property of a single synchronous machine is often studied in the single machine connected to the infinite bus shown in Fig. 1(a), where V_g , $\angle\delta$, I , x_l , r_l , V_∞ , and $\angle 0$ are the generator voltage, generator phase, current, transmission line reactance, transmission line resistance, infinite bus voltage, and infinite bus phase, respectively. Here the infinite bus is characterized by constant voltage and constant frequency. The active power P_e and reactive power Q_e are calculated by

$$P_e + iQ_e = V_g \exp(i\delta) I^*, \tag{2.1}$$

where $*$ indicates the complex conjugate, and the current I is given by

$$I = \frac{V_g \exp(i\delta) - V_\infty}{r_l + ix_l}. \tag{2.2}$$

By substituting Eq. (2.2) into Eq. (2.1), the active power P_e is obtained as the real part of the equation

$$P_e = \frac{V_g V_\infty}{x_l} \sin \delta = P^{\max} \sin \delta, \tag{2.3}$$

by considering relation $r_l \ll x_l$. P_e is plotted as a function of the generator phase δ in Fig. 1(b). It should be noted that P_e is the electric power transmitted through the transmission line and works as a synchronizing force in power systems.

2.3. Physical and biological systems

Many collective synchronization phenomena are known in physical and biological systems.¹²⁾ Physical examples include clocks hanging on a wall, an array of lasers, microwave oscillators, and Josephson junctions. Biological examples include synchronously flashing fireflies, networks of pacemaker cells in the heart, and metabolic synchrony in yeast cell suspensions.

Kuramoto proposed a coupled oscillator model to explain this rich variety of synchronization phenomena.^{11)–13)} In his model, the dynamics of the oscillators are

governed by

$$\dot{\theta}_i = \omega_i + \sum_{j=1}^N k_{ji} \sin(\theta_j - \theta_i), \quad (2.4)$$

where θ_i , ω_i , and k_{ji} are the oscillator phase, the natural frequency, and the coupling strength, respectively. The second term of the RHS of Eq. (2.4) is identical to Eq. (2.3). If the coupling strength k_{ij} exceeds a certain threshold, the system exhibits synchronization.

§3. Coupled oscillator model

In this section, we first explain the formulation of our coupled oscillator model, and then give an illustrative example to understand the basic behavior of the model.

3.1. Formulation

Our model is developed based on the Kuramoto model with inertia¹⁴⁾ by adding goods markets. We consider the system to consist of oscillator i and oscillator j . The angle θ_i of oscillator i is written as

$$\theta_i = \omega t + \tilde{\theta}_i, \quad (3.1)$$

where ω and $\tilde{\theta}_i$ are the frequency and phase, respectively.

The energy dissipated as heat from oscillator i at a rate proportional to the square of the angular velocity is

$$P_d = K_D (\dot{\theta}_i)^2, \quad (3.2)$$

and the kinetic energy accumulated in oscillator i at a rate proportional to the square of the angular velocity is

$$P_a = \frac{1}{2} I \frac{d}{dt} (\dot{\theta}_i)^2, \quad (3.3)$$

where K_D and I are a dissipation constant and a moment of inertia, respectively. From Eq. (3.1), the angular difference $\Delta\theta_{ji}$ is written as the phase difference $\Delta\theta_{ji} = \theta_j - \theta_i = \tilde{\theta}_j - \tilde{\theta}_i$ using the phase $\tilde{\theta}_i$. The power transmitted from one oscillator to another is given by

$$P_t = -k_{ji} \sin \Delta\theta_{ji}, \quad (3.4)$$

using the phase difference $\Delta\theta_{ji}$. In Eq. (3.4), the negative sign is used to indicate that power is lost from oscillator i .

By substituting relations (3.2), (3.3), and (3.4) into the power balance equation for oscillator i ($P_s = P_d + P_a + P_t$), we obtain an equation corresponding to Eq. (2.4)

$$I\omega\ddot{\theta}_i = P_s - K_D\omega^2 - 2K_D\omega\dot{\tilde{\theta}}_i + k_{ji} \sin \Delta\theta_{ji}, \quad (3.5)$$

using the approximate relation $\dot{\tilde{\theta}}_i \ll \omega$. From the above discussion, without loss of generality, we obtain an equation describing the dynamics of an N -oscillator system,

$$\ddot{\theta}_i = P_i - \alpha\dot{\theta}_i + \sum_{j=1}^N k_{ji} \sin \Delta\theta_{ji}, \quad (3.6)$$

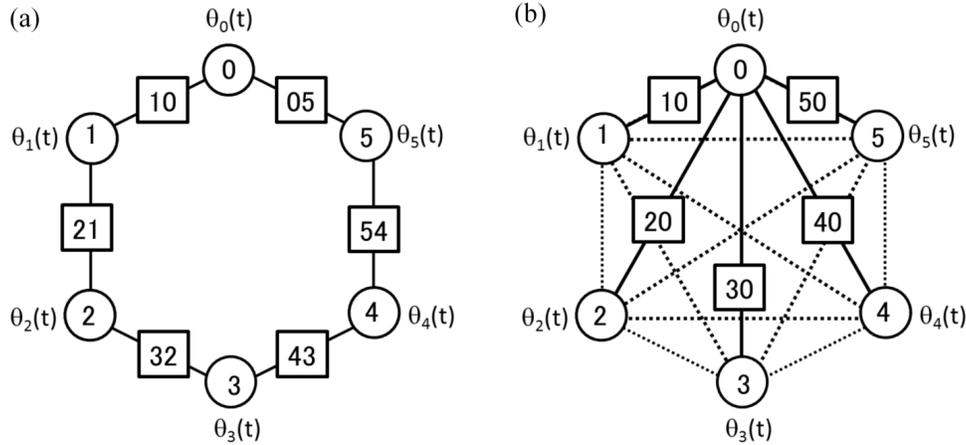


Fig. 2. Network topology.

where $I\omega$, $P_s - K_D\omega^2$, and $2K_D\omega$ are replaced by 1, P_i , and α , respectively. P_i is regarded as the net input to oscillator i , which is equal to the difference between the input to oscillator i from outside the N -oscillator system and the output from oscillator i to outside the N -oscillator system. Hereafter $\tilde{\theta}_i$ is written as θ_i for simplicity.

It is noted that the synchronizing force is interpreted as being the goods markets. The meaning of the goods markets is clarified by adding the sectoral fluctuations of demand or supply, δ_{ji} , as

$$\ddot{\theta}_i = P_i - \alpha\dot{\theta}_i + \sum_{j=1}^N \{k_{ji} \sin \Delta\theta_{ji} + \delta_{ji}\}. \quad (3.7)$$

The introduction of these sectoral fluctuations is intended to be consistent with the importance of the small sectoral shocks found in Ref. 6). Figure 2 depicts a system consisting of six oscillators. Figure 2(a) is the Nearest Neighbor (NN) graph, and Fig. 2(b) is the Complete (C) graph. Oscillators and goods markets are indicated by circles and rectangles, respectively. In Fig. 2(a), the number of the markets is six, because they are only open to the neighboring oscillators. In Fig. 2(b), however, the number of markets is five for each oscillator and the total number of markets is equal to $(6 \times 5)/2 = 15$.

Demand d_{ij} for, and supply s_{ji} of, good i can be written as

$$d_{ij} = d_0 + \delta_{ij}, \quad (3.8)$$

$$s_{ji} = s_0 + \delta_{ji}, \quad (3.9)$$

and are determined through the goods market ij . d_0 and s_0 are the equilibrated demand and equilibrated supply of good i in goods market ji , and δ_{ij} and δ_{ji} are the sectoral fluctuations of demand and supply, respectively. In Eqs. (3.8) and (3.9) the equilibrated demand d_0 and equilibrated supply s_0 are given by

$$d_0 = s_0 = k_{ji} |\sin(\theta_j^* - \theta_i^*)|, \quad (3.10)$$

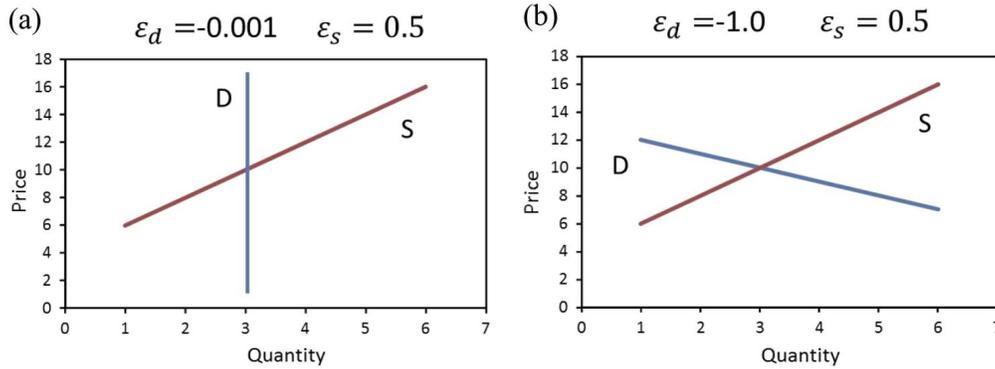


Fig. 3. Demand and supply.

where the equilibrated phase θ_i^* ($i = 1, \dots, N$) are solution of Eq. (3.7) with $\ddot{\theta}_i = \dot{\theta}_i = 0$. The relations between demand d or supply s and the price p of good i are written using the price elasticity of demand ϵ_d or the price elasticity of supply ϵ_s ,

$$\frac{d}{d_0} = \left(\frac{p}{p_0}\right)^{\epsilon_d}, \tag{3.11}$$

$$\frac{s}{s_0} = \left(\frac{p}{p_0}\right)^{\epsilon_s}. \tag{3.12}$$

The price as a function of demand and supply is depicted in Fig. 3. Figure 3(a) is plotted for small value of ϵ_d . In this case, the demand is shown as a vertical line. Oscillator j responds to the sectoral fluctuations of supply δ_{ji} according to price p . Depending on the market flexibility, oscillator j changes its demand by δ_{ij} , given by

$$\delta_{ij} = \begin{cases} -\delta_{ji}, & (\epsilon_d < 0) \\ 0, & (\epsilon_d = 0) \end{cases} \tag{3.13}$$

by responding to the fluctuation of supply δ_{ji} .

3.2. Illustrative example

We consider an oscillator system in order to understand the basic behavior of the coupled oscillator model. For the NN graph, the model equations are written as

$$\begin{aligned} \ddot{\theta}_1 &= P_1 - \alpha\dot{\theta}_1 + \{k_{21} \sin(\theta_2 - \theta_1) + \delta_{21}\}, \\ \ddot{\theta}_2 &= P_2 - \alpha\dot{\theta}_2 + \{k_{21} \sin(\theta_1 - \theta_2) + \delta_{12}\} + \{k_{32} \sin(\theta_3 - \theta_2) + \delta_{32}\}, \\ &\dots \\ \ddot{\theta}_{N-1} &= P_{N-1} - \alpha\dot{\theta}_{N-1} + \{k_{N-1N-2} \sin(\theta_{N-2} - \theta_{N-1}) + \delta_{N-2N-1}\}, \\ &\quad + \{k_{NN-1} \sin(\theta_N - \theta_{N-1}) + \delta_{NN-1}\}, \\ \ddot{\theta}_N &= P_N - \alpha\dot{\theta}_N + \{k_{NN-1} \sin(\theta_{N-1} - \theta_N) + \delta_{NN-1}\}. \end{aligned} \tag{3.14}$$

Analytic solutions of the stationary state will be obtained by solving the simultaneous equations of (3.14) with $\dot{\theta}_i = \ddot{\theta}_i = 0$ ($i = 1, \dots, N$). We consider the case, where

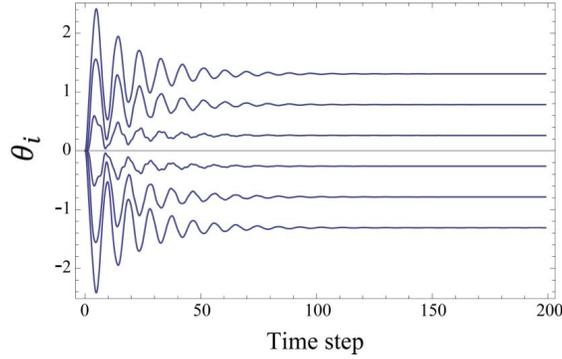


Fig. 4. Stationary solution.

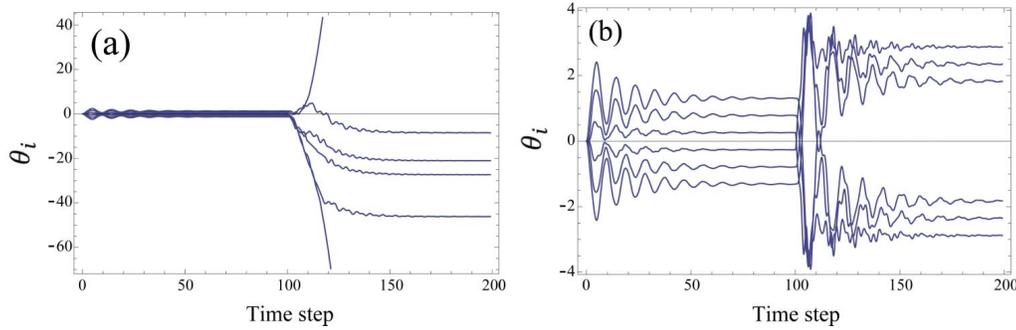


Fig. 5. Fluctuation and elasticity.

$P_1 = P = 1$, $P_i = 0$ ($i = 2, \dots, N - 1$), $P_N = -P = -1$, and $\theta_N = 0$. The analytic solutions obtained for the stationary state θ_i^* and the synchronizing coupling strengths k_{ji} are

$$\theta_i^* = \theta_{i+1}^* + \arcsin \frac{P}{k_{ji}}, \quad (3.15)$$

$$k_{ji} = \frac{P}{\sin(\theta_i^* - \theta_{i+1}^*)}. \quad (3.16)$$

We simulate the behavior of the system by solving Eq. (3.14) numerically with the initial condition $\dot{\theta}_i(0) = \ddot{\theta}_i(0) = 0$ ($i = 1, \dots, N$). Synchronization was reproduced as an equilibrium solution in a simple NN graph as shown in Fig. 4. The response to a sectoral fluctuation at $t = 100$ is also simulated in the NN graph. In the case of zero elasticity ($\epsilon_d = 0$), the synchronization was broken, as shown in Fig. 5(a). In contrast, in the case of finite elasticity ($\epsilon_d < 0$), stability was restored after a shift of phase, as shown in Fig. 5(b).

§4. Analysis of sectoral synchronization

In this section, we analyze the sectoral synchronization observed in the Japanese business cycle using the coupled oscillator model described in the previous section. The effects on synchronization resulting from a sectoral shock are studied for systems with different price elasticities and coupling strengths.

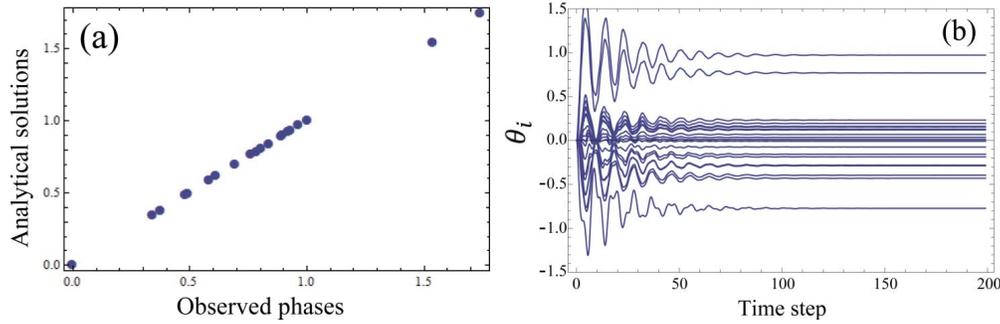


Fig. 6. Parameter calibration and synchronization.

4.1. Synchronization

For the business cycle with a common period of $T = 60$ months, the phases of production, shipment, and inventory are given in the article.²⁾ The sectors are rearranged in decreasing order of the observed phase of production. If we assume that the network is the NN graph, we can calibrate the coupling strengths k_{ij} using Eq. (3.16) with the observed phases θ_i^* . Analytical solutions of the stationary phases are then obtained using Eq. (3.15).

The analytical solutions of stationary phases are compared with the observed phases in Fig. 6(a). The agreement between the analytical solutions and the observed phases is quite good. We simulate the behavior of the system by solving Eq. (3.14) numerically with the initial condition $\dot{\theta}_i(0) = \ddot{\theta}_i(0) = 0$ ($i = 1, \dots, N$). Synchronization was reproduced as an equilibrium solution in a simple NN graph, as shown in Fig. 6(b).

4.2. Fluctuating goods market

The complex order parameter

$$q(t) = \frac{1}{N} \sum_{j=1}^N e^{i\theta_j(t)} = r(\cos(\phi) + i\sin(\phi)) \quad (4.1)$$

is defined as a macroscopic quantity that corresponds to the centroid of the phases of oscillators.¹²⁾ The radius r measures the coherence and ϕ is the average phase. If $\text{Re}(q(t)) \approx 1$ and $\text{Im}(q(t)) \approx 0$, the oscillators remain in the synchronization region where the phase differences are fairly small. In contrast, if $\text{Re}(q(t))$ and $\text{Im}(q(t))$ oscillate between 1 and -1 , the oscillators behave like a giant oscillator with a frequency additional to the common frequency ω .

First, we simulate the response to the sectoral fluctuation at $t = 100$ in the NN graph. The result for a finite elasticity ($\epsilon_d < 0$) is shown in Fig. 7(a). It is seen that $\text{Re}(q(t)) \approx 1$ and $\text{Im}(q(t)) \approx 0$, even after the sectoral shock was applied in the middle of the network at $t = 100$. This means that the synchronization is stable, i.e., the oscillators remain in the region where phase differences are fairly small. For the case of zero elasticity ($\epsilon_d = 0$), $\text{Re}(q(t))$ and $\text{Im}(q(t))$ oscillate rapidly between 1 and -1 as shown in Fig. 7(b). This means that synchronization is broken and

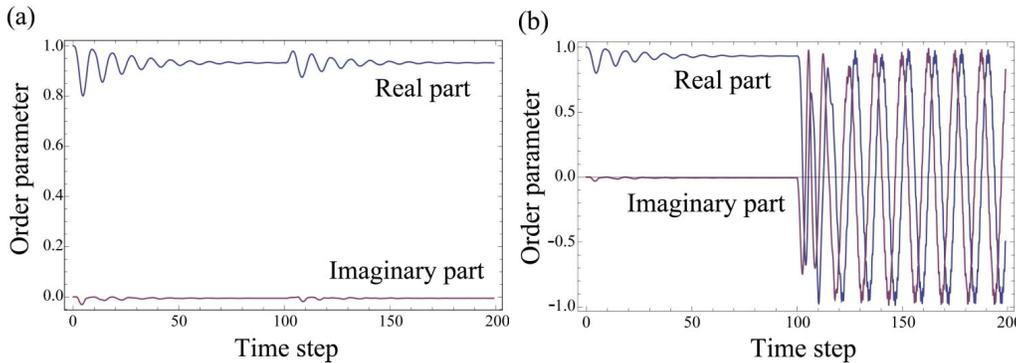


Fig. 7. Price elasticity and stability of synchronization.

the oscillators behave like a giant oscillator with a high frequency additional to the common frequency ω .

Next, we simulate the oscillator system with weaker coupling strengths k_{ij} . We expect that the system to behave as if the oscillators were uncoupled for k_{ij} below than a certain threshold k_{ij}^c . For the Kuramoto oscillator, the exact formula of the critical coupling strength k_{ij}^c has been derived and verified with the results of numerical simulations. The results of our simulation with weaker coupling strengths of $0.3k_{ij}$ are shown in Fig. 8. The time evolution of the phases θ_i shown in Fig. 8(a) depicts a few oscillators separating from the main part of the coupled oscillator system, which still exhibits synchronization. The oscillator system acts like a giant oscillator with a low frequency additional to the common frequency ω , as seen in Fig. 8(b). The phases θ_i at $t = 200$ in Fig. 8(c) show that the system disintegrated into four parts and that the main part moves with small phase differences.

This result is quite different from our expectation, but this is also reasonable, because the system under consideration has $P_1 = 1, P_i = 0$ ($i = 2, \dots, N - 1$), $P_N = -1$ ($N = 21$). Hence, one end of the system is pulled in the positive direction and the other end is pulled in the opposite direction. Therefore, if the coupling strengths are weak enough, the oscillators in both ends of the system are separated from the main part. The oscillators in the main part lose tension and shrink to the small phase difference. However, it is noted that the system in the C graph with $P_i \neq 0$ ($i = 1, \dots, N$) might behave similarly to the Kuramoto oscillator.

§5. Conclusions

The sectoral synchronization observed for the Japanese business cycle in the Indices of Industrial Production data is an example of synchronization. The stability of this synchronization under a shock, e.g., fluctuation of supply or demand, is a matter of interest in physics and economics.

First, existing theories for synchronization for economic systems, power systems, and physical and biological systems were reviewed.

We then considered an economic system consisting of industry sectors and goods markets. A coupled oscillator model exhibiting synchronization was developed based

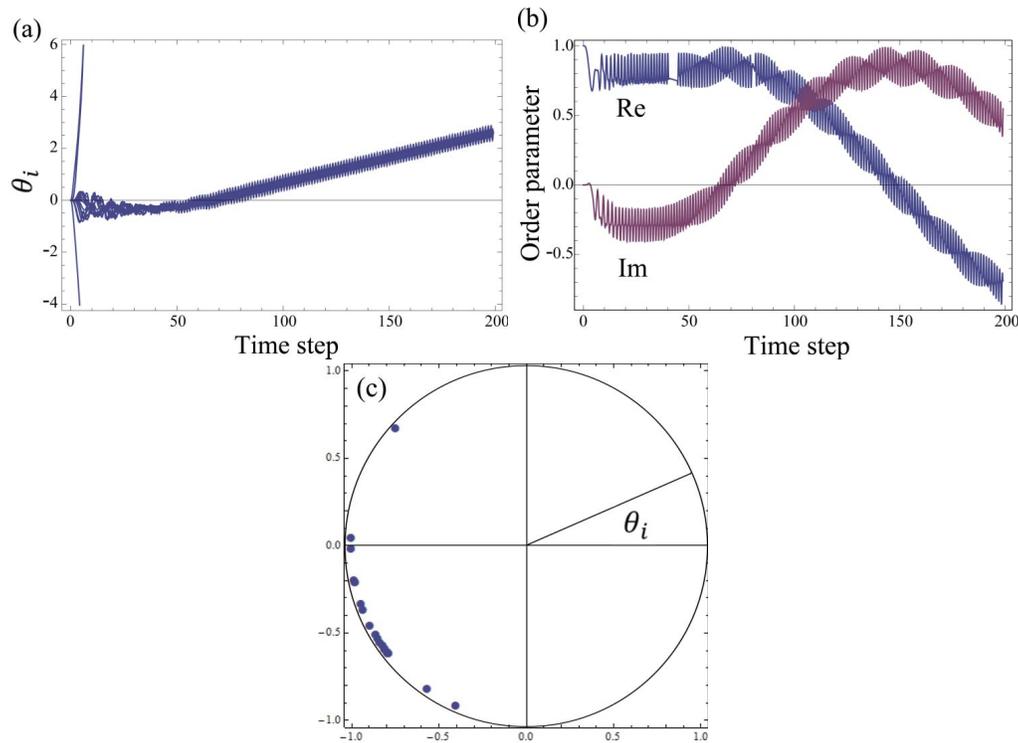


Fig. 8. Case with the coupling strength multiplied by 0.3.

on the Kuramoto model with inertia by adding goods markets, and analytic solutions of the stationary state and the coupling strength were obtained.

Finally, we analyzed the sectoral synchronization observed in the Japanese business cycle using the coupled oscillator model. Effects on synchronization resulting from a sectoral shock were simulated for systems with different price elasticities and coupling strengths. The agreement between the analytical solutions and the observed phases was quite good. Synchronization was reproduced as an equilibrium solution in a simple NN graph. Analysis of the order parameters showed that the synchronization was stable for a finite elasticity ($\epsilon_d < 0$), whereas the synchronization was broken and the oscillators behaved like a giant oscillator with a frequency additional to the common frequency ω for zero elasticity ($\epsilon_d = 0$).

In the future work, we intend to study

1. Simulation of the oscillator system for the complete graph with $P_i \neq 0$ ($i = 1, \dots, N$).
2. Theory of phase transition for synchronization, including derivation of the exact formula for the critical coupling strength k_{ij}^c .
3. Development of the economic implications for sectoral fluctuations and the propagation of risk through the economic network.

Acknowledgements

The authors thank the Yukawa Institute for Theoretical Physics at Kyoto University. Discussions during the YITP workshop YITP-W-11-04 on “Econophysics 2011 —The Hitchhiker’s Guide to the Economy—” were useful for completing this work.

References

- 1) H. Iyetomi, Y. Nakayama, H. Yoshikawa, H. Aoyama, Y. Fujiwara, Y. Ikeda and W. Souma, *J. of the Japanese and International Economies* **25** (2011), 246.
- 2) H. Iyetomi, Y. Nakayama, H. Aoyama, Y. Fujiwara, Y. Ikeda and W. Souma, *Phys. Rev. E* **83** (2011), 016103.
- 3) R. M. Goodwin, *Econometrica* **15** (1947), 181.
- 4) R. M. Goodwin, *Econometrica* **19** (1951), 1.
- 5) J. B. Long and C. I. Plosser, *J. of Political Economy* **91** (1983), 39.
- 6) J. B. Long and C. I. Plosser, *American Economic Review* **77** (1987), 333.
- 7) H. M. Anderson and J. B. Ramsey, *Economic research reports PR # 99-01* (New York University, 1999).
- 8) D. D. Selover, R. V. Jensen and J. Kroll, *Studies in Nonlinear Dynamics & Econometrics* **7** (2003), 1.
- 9) B. Sussmuth, *Business Cycles in the Contemporary World* (Springer, Berlin, Heidelberg, 2003).
- 10) P. Kunder, *Power System Stability and Control* (McGraw-Hill, New York, 1993), p. 211.
- 11) Y. Kuramoto, in *International Symposium on Mathematical Problems in Theoretical Physics*, Lecture Notes in Phys. Vol. 39, ed. H. Araki (Springer, New York, 1975), p. 420.
- 12) S. H. Strogatz, *Physica D* **143** (2000), 1.
- 13) J. A. Acebron, L. L. Bonilla, C. J. P. Vicente, F. Ritort and R. Spigler, *Rev. Mod. Phys.* **77** (2005), 137.
- 14) G. Filatrella, A. H. Nielsen and N. F. Pedersen, *Eur. Phys. J. B* **61** (2008), 485.

Power Laws in Firm Productivity

Takayuki MIZUNO,^{1,2} Atushi ISHIKAWA,³ Shouji FUJIMOTO^{2,3} and
Tsutomu WATANABE^{2,4}

¹*Division of Information Engineering, Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba 305-8573, Japan*

²*Canon Institute for Global Studies, 1-5-1 Marunouchi, Tokyo 100-6511, Japan*

³*Department of Informatics and Business, Faculty of Business Administration and Information Science, Kanazawa Gakuin University, Kanazawa 920-1392, Japan*

⁴*Graduate School of Economics, University of Tokyo, Tokyo 113-0033, Japan*

We estimate firm productivity for about 3.2 million firms from 30 countries. We find that the distribution of firm productivity in each country, which is measured by total factor productivity (TFP), has a power law upper tail. However, the power law exponent of a TFP distribution in a country tends to be greater than that of a sales distribution in that country, indicating that the upper tail of a TFP distribution is less heavy compared to that of a sales distribution. We also find that the power law exponent of a TFP distribution tends to be greater than the power law exponents associated with the number of workers or tangible fixed assets. Given the idea that the sales of a firm is determined by the amount of various inputs employed by the firm (i.e., “production function” in the terminology of economics), these results suggest that the heavy tail of a sales distribution in a country comes not from the tail of a TFP distribution, but from the tail of the distribution of the number of workers or tangible fixed assets.

§1. Introduction

Ever since Pareto reported 120 years ago that the distribution of personal wealth follows a power law, both physicists and economists have investigated the distributions of a wide variety of size variables.¹⁾ Some economists found in the 1950s and 1960s that firm size distributions have fatter tails than log-normal distributions,^{2)–4)} while physicists showed in the 1990s and later that the distributions of firm size variables, including firm sales, the number of workers, and tangible fixed assets, are characterized by power law tails.^{5)–9)} The mechanism behind firm growth was also investigated so as to learn more about where the power law tails of firm size distributions come from.^{10)–13)}

More recently, researchers have shifted their focus to the relationship between firm sales, the number of workers, and tangible fixed assets.^{14)–16)} It was shown that this relationship is not linear but nonlinear,^{5), 14)} and that there exists a scaling law for the relationship.¹⁴⁾ On the other hand, some economists studied this relationship based on the idea of “production function”,^{17)–20)} in which the output of a firm is determined by the amount of inputs employed by the firm, including the number of workers, tangible fixed assets, and productivity. Among several alternative functional forms for production function proposed by economists, the most widely used one is a Cobb-Douglas production function,²¹⁾ which is given by $Y = AK^\alpha L^\beta$, where Y is output, L is the number of workers, K is tangible fixed assets, and A is productivity

(more precisely, total factor productivity, or TFP). The exponents α and β are constant parameters, which are usually estimated empirically.

Note that Y , L , and K are reported by firms in their financial statements, so that they are all observable. However, A is not observable, so that we need to estimate it. Recently, it was found using firm level data from major countries, including the U.S. and Japan, that the relationship between Y , L , and K is well described by a Cobb-Douglas production function at least for mega firms (i.e. firms that belong to the tail part of firm size distributions).^{14),17)} This suggests a method to estimate A from the data; namely, we first estimate the exponents α and β of a Cobb-Douglas production function using the data of Y , L , and K , and then compute A by $A = Y/(K^\alpha L^\beta)$. In the present paper, we estimate TFP in this way, and then look at the tail part of its distribution. Specifically, we are interested in whether the distribution of TFP is characterized by power law tails, as in the case of the distributions for Y , L , and K , and whether the tail of a TFP distribution is thicker or thinner than the tails associated with Y , L , and K . Note that total factor productivity estimated in this way is closely related to, but different from labor productivity, which was extensively examined by previous studies.^{15),16)} Labor productivity is defined by Y/L , therefore it coincides with total factor productivity only when α and β are equal to zero and unity, respectively. In other words, previous studies were based on the assumption that this parameter restriction is not a bad approximation. We do not impose such a restriction a priori but estimate α and β from the data. This is an important departure from previous studies.

The rest of the paper is organized as follows. In §§2 to 4, we estimate TFP for about 3.2 million firms from 30 countries, with the sample period of 2004 to 2008. The data we will use is from “Orbis”, a database compiled by Bureau van Dijk Electronic Publishing. In §2, we give a brief review of the method proposed by Ref. 14) for estimating a Cobb-Douglas production function. In §3, we estimate two key parameters, α and β , of a Cobb-Douglas function and compare them across countries, as well as across industries. In §4, we show that the distribution of TFP is characterized by a power law tail, and that its power law exponent is greater than those for Y , L , and K for most countries, as well as for most industries. In §5, we use value added instead of sales as a measure of output in estimating TFP. This is closer to the traditional treatment adopted by economists. We repeat the same exercise as in §§3 and 4 to confirm that the basic results remain unchanged. Section 6 concludes the paper.

§2. Estimation of Cobb-Douglas production functions

It is shown by previous studies that various firm size variables, including firm sales (S), the number of workers (L), and tangible fixed assets (K), follow power law distributions in many countries. That is,

$$P_{>}(S) \propto S^{-\mu_S} \quad \text{for } S > S_0, \quad (2.1)$$

$$P_{>}(L) \propto L^{-\mu_L} \quad \text{for } L > L_0, \quad (2.2)$$

$$P_{>}(K) \propto K^{-\mu_K} \quad \text{for } K > K_0, \quad (2.3)$$

where S_0 , L_0 , and K_0 represent the size thresholds, and μ_S , μ_L , and μ_K are the power law exponents associated with S , L , and K . Reference 14) found two statistical laws regarding the relationships between S , L , and K . The first law is that the joint probability distributions of S and L , S and K , L and K are characterized by nonlinear functions of the form

$$P_J(L, S) = P_J \left(\left(\frac{S}{A_{LS}} \right)^{\frac{1}{\nu_{LS}}}, A_{LS} L^{\nu_{LS}} \right) \quad \text{for } L > L_0, \quad (2.4)$$

$$P_J(S, K) = P_J \left(\left(\frac{K}{A_{SK}} \right)^{\frac{1}{\nu_{SK}}}, A_{SK} S^{\nu_{SK}} \right) \quad \text{for } S > S_0, \quad (2.5)$$

$$P_J(L, K) = P_J \left(\left(\frac{K}{A_{LK}} \right)^{\frac{1}{\nu_{LK}}}, A_{LK} L^{\nu_{LK}} \right) \quad \text{for } L > L_0, \quad (2.6)$$

where the coefficients A_{LS} , A_{SK} , A_{LK} and ν_{LS} , ν_{SK} , ν_{LK} are fixed numbers. The second law is that S , L , and K satisfy Gibrat's law as

$$P \left(\frac{S}{A_{LS} L^{\nu_{LS}}} \mid L \right) = P \left(\frac{S}{A_{LS} L^{\nu_{LS}}} \right) \quad \text{for } L > L_0, \quad (2.7)$$

$$P \left(\frac{K}{A_{SK} S^{\nu_{SK}}} \mid S \right) = P \left(\frac{K}{A_{SK} S^{\nu_{SK}}} \right) \quad \text{for } S > S_0, \quad (2.8)$$

$$P \left(\frac{K}{A_{LK} L^{\nu_{LK}}} \mid L \right) = P \left(\frac{K}{A_{LK} L^{\nu_{LK}}} \right) \quad \text{for } L > L_0. \quad (2.9)$$

From these two laws, Ref. 14) derives the following relational expression (see Appendix A for details).

$$S = AK^\alpha L^\beta \quad \text{for } L > L_0 \quad \text{and} \quad K > K_0, \quad (2.10)$$

where the coefficients α , β are fixed numbers, and A is a stochastic variable which is independent of K and L , so that A satisfies

$$P(A \mid K, L) = P(A). \quad (2.11)$$

In economics, Eq. (2.10) is known as a Cobb-Douglas production function, while the variable A is referred to as total factor productivity, or TFP.

In estimating the coefficients α and β in Eq. (2.10), we need pay attention to the issue of multicollinearity; namely, K and L are highly correlated so that a simple application of ordinary least squares does not work. To overcome this issue, Ref. 14) proposes to introduce new variables, Z_1 and Z_2 , which are defined as

$$\log Z_1 = \frac{\log L}{\sigma_{\log L}} + \frac{\log K}{\sigma_{\log K}}, \quad (2.12)$$

$$\log Z_2 = \frac{\log L}{\sigma_{\log L}} - \frac{\log K}{\sigma_{\log K}}, \quad (2.13)$$

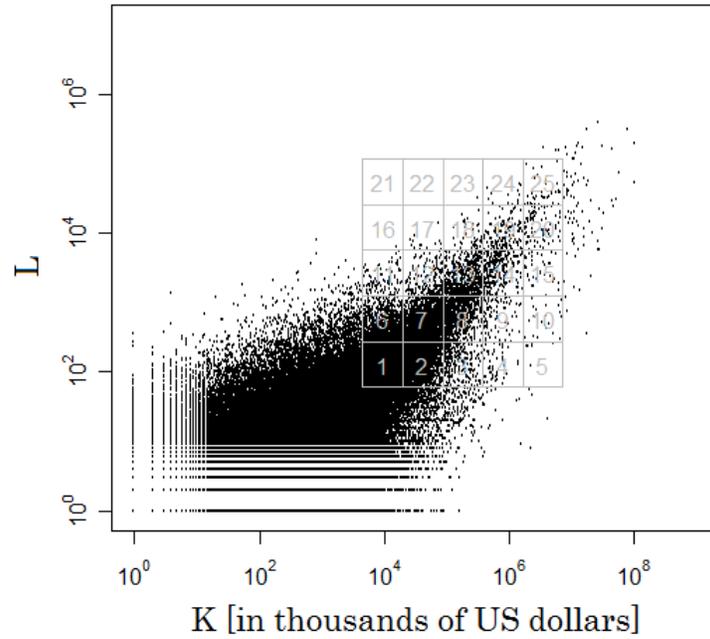


Fig. 1. Scatter plots of tangible fixed assets and the number of workers for Japanese firms in 2008. We evenly divide the entire (K, L) plane (i.e. $4.5 \times 10^3 \leq K \leq 7.2 \times 10^6$ and $60 \leq L \leq 1.2 \times 10^5$) into 5×5 sub-areas.

where $\sigma_{\log L}$ and $\sigma_{\log K}$ are the standard deviations of $\log K$ and $\log L$. Then Eq. (2.10) is transformed into

$$S = AZ_1^{\theta_1} Z_2^{\theta_2} \quad \text{for } L > L_0 \text{ and } K > K_0, \tag{2.14}$$

where

$$\alpha = \frac{\theta_1 - \theta_2}{\sigma_{\log K}}; \quad \beta = \frac{\theta_1 + \theta_2}{\sigma_{\log L}}. \tag{2.15}$$

Note that, by construction, $\log Z_1$ and $\log Z_2$ are not correlated to each other in the regions $L > L_0$ and $K > K_0$. Therefore, we can safely apply an ordinary least square to Eq. (2.14) so as to obtain estimates for α and β .

We run this regression using Japanese data in 2008 to find that α and β are equal to 0.18 and 0.85, respectively. Based on these estimates, we conduct the following exercise. In Fig. 1, we plot K and L , and divide the entire (K, L) space (i.e. $4.5 \times 10^3 \leq K \leq 7.2 \times 10^6$ and $60 \leq L \leq 1.2 \times 10^5$) evenly into 5×5 sub-areas. Then, we produce a probability density function of the estimated TFP for each of the sub-areas. The results are presented in Fig. 2. It is confirmed that the probability density functions of A are almost identical irrespective of the values of K and L , indicating that the estimated A is indeed independent of K and L .

§3. Estimates of α and β

Figure 3 presents the estimates of α and β for 30 countries, whose list is available in Table II in Appendix B. Figure 3 shows that β is greater than α for all countries

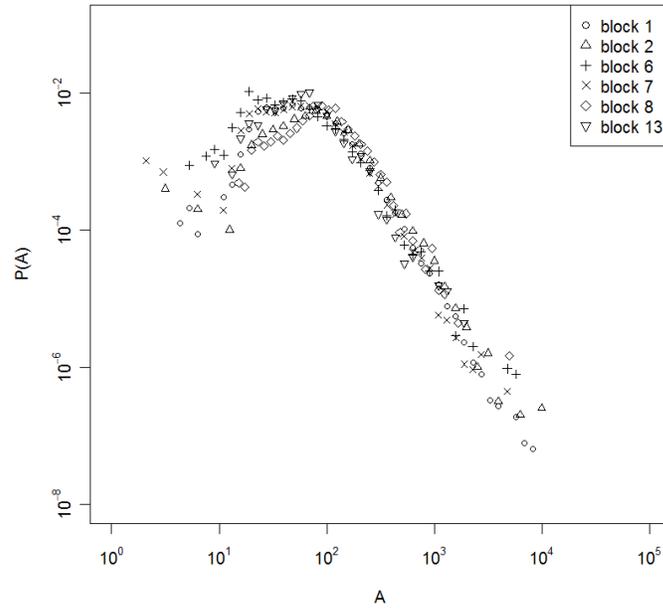


Fig. 2. PDF of total factor productivity. The block number refers to the corresponding sub-area in Fig. 1.

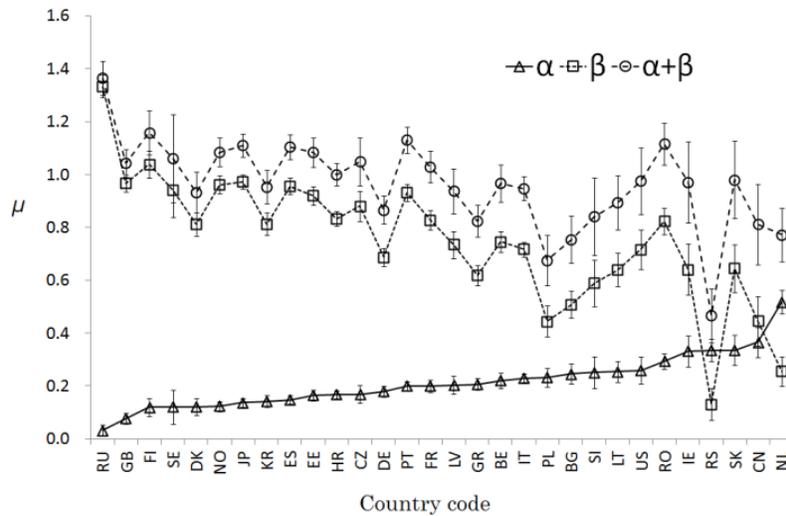


Fig. 3. The estimates of α , β , and $\alpha + \beta$ for different countries. The horizontal axis represents the country code, which is given in Table II in Appendix B.

except Serbia and the Netherlands, suggesting that firm sales depend more closely on the number of workers than on the amount of tangible fixed assets. Figure 4 presents the estimates of α and β for different industries. The horizontal axis of Fig. 4 represents the SIC (Standard Industrial Classification) code, which is given in Table III in Appendix B.²²⁾ We see that the estimate of α tends to be somewhere

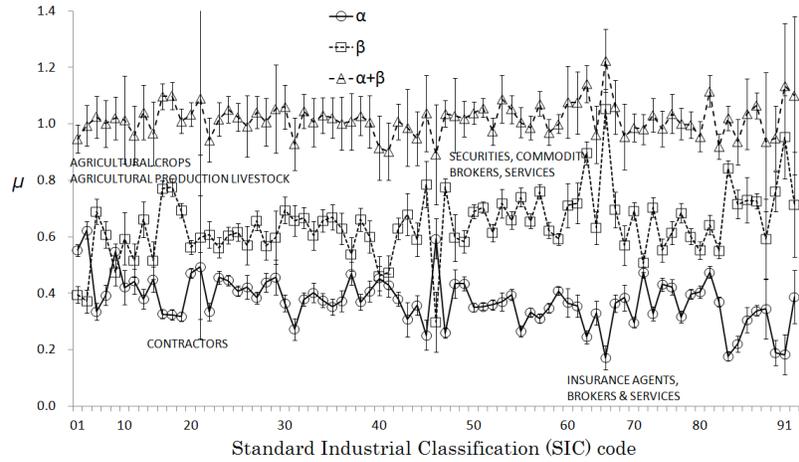


Fig. 4. The estimates of α , β , and $\alpha + \beta$ for different industries. The horizontal axis represents the SIC code, which is given in Table III in Appendix B.

around 0.4, and β tends to be around 0.6, with the exception of “insurance agents, brokers, and services” (SIC code no.64) where β is very close to unity.

Turning to the sum of α and β , we see in Fig. 3 that the sum of α and β is close to unity for most countries, although there are several countries, like Serbia, in which the sum of α and β is significantly smaller than unity. Somewhat interestingly, there is a tendency that the sum of α and β is smaller for countries with larger α . Also, we see in Fig. 4 that the sum of α and β is very close to unity for almost all industries. In economics, $\alpha + \beta = 1$ implies a constant return to scale; namely, if one increases each of K and L by, say, ten percent, then S also increases by ten percent. The above empirical results suggest that this property holds in most countries, as well as in most industries.

§4. Total factor productivity

Given the estimates of α and β in hand, we now proceed to the estimation of total factor productivity. Using Eq. (2.10), we calculate total factor productivity as $A = S/(K^\alpha L^\beta)$. Figure 5 shows the cumulative distribution functions of total factor productivity obtained in this way for Japanese firms in the years of 2004 to 2008. We see that the tail part of the distribution for each year satisfies:

$$P_>(A) \propto A^{-\mu_A} \quad \text{for } A > A_0 \tag{4.1}$$

indicating that total factor productivity follows a power law distribution with an exponent of μ_A . Note that the power law exponent is about 1.65, and it remains unchanged over time.

Figure 6 shows the power law exponents of total factor productivity for 27 countries, together with the power law exponents of sales (μ_S), the number of workers (μ_L), and tangible fixed assets (μ_K). These power law exponents are estimated using a new method proposed by recent studies.^{9),23)} That is, through a set of statistical

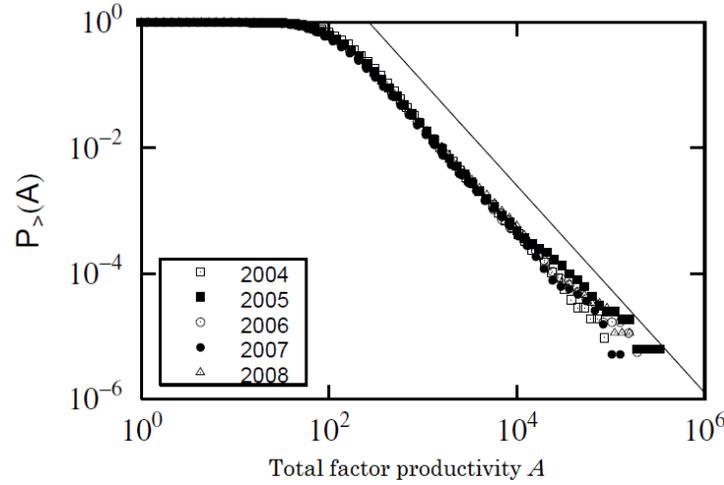


Fig. 5. CDF of total factor productivity for Japanese firms in 2004 to 2008. The straight reference line represents a power law with an exponent of 1.65.

Table I. Inequality in terms of productivity, sales, the number of workers, and tangible fixed assets.

	Productivity	Sales	Number of workers	Tangible fixed assets
20 percent of the total	1.3%	0.019%	0.016%	0.0067%
50 percent of the total	23%	0.21%	0.25%	0.066%
80 percent of the total	43%	2.9%	5.3%	0.90%

tests, we identify a range of a variable in which the variable follows a power law distribution. A power law exponent is estimated using observations only in that range. We see that

$$\mu_K < \mu_S \leq \mu_L < \mu_A \quad (4.2)$$

holds for almost all countries. This indicates that inequality across firms in terms of A tends to be smaller than inequality in terms of S , K , or L . To show this result in a different way, we count the number of firms that account for 20, 50, 80 percent of the total of A . Similarly, we count the number of firms that account for 20, 50, 80 percent of the total of S , L , and K . The result for Japanese firms in 2008 is presented in Table I, showing that we need top 23 percent firms to account for 50 percent of the total of A , while we need only 0.21 percent firms to account for 50 percent of the total of S . Even more surprisingly, we need only 0.066 percent firms to account for 50 percent of the total of K .

The result that μ_A is greater than μ_S implies the following. Note that, by construction, A , $\log Z_1$ and $\log Z_2$ in Eq. (2.14) are not correlated to each other. Therefore, the four power law exponents, μ_S , μ_A , μ_{Z_1} , and μ_{Z_2} , must satisfy the following:^{(24), (25)}

$$\mu_S = \min \left\{ \mu_A, \frac{\mu_{Z_1}}{\theta_1}, \frac{\mu_{Z_2}}{\theta_2} \right\}. \quad (4.3)$$

The empirical result that μ_A tends to be greater than μ_S implies that the tail part

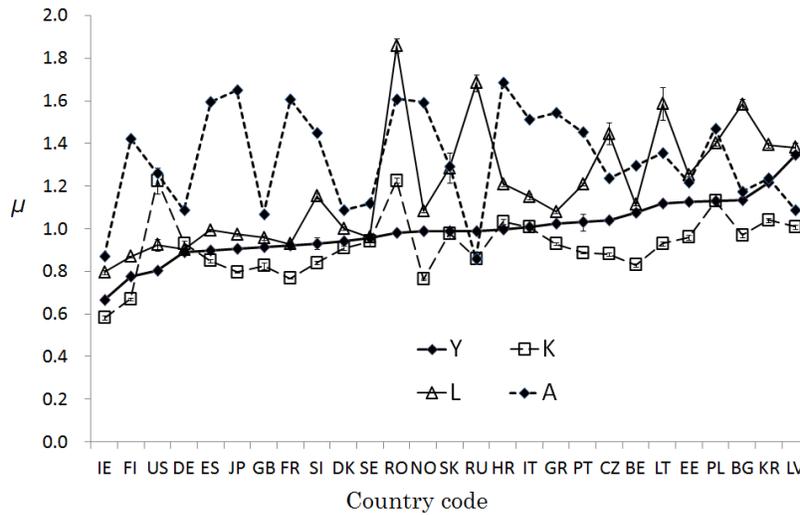


Fig. 6. Power law exponents in 2008 for different countries. The horizontal axis represents the country code, which is given in Table II in Appendix B.

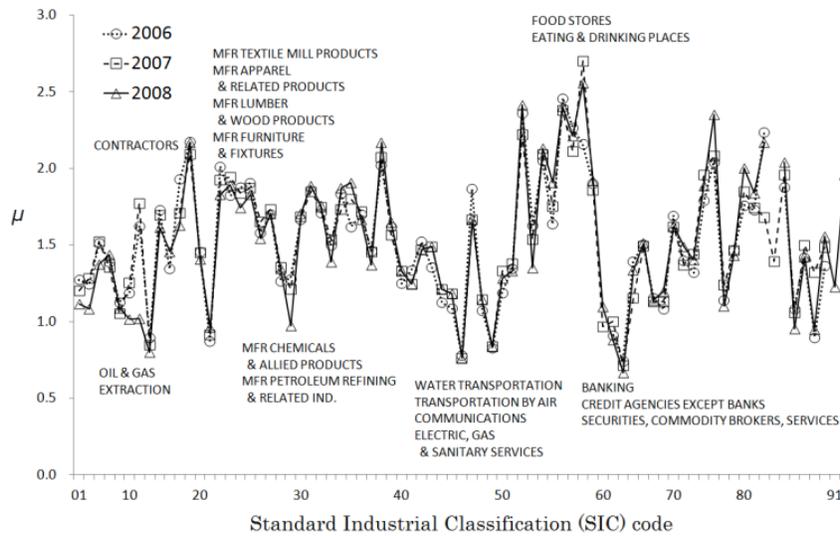


Fig. 7. Power law exponents for different industries. The horizontal axis represents the SIC code, which is given in Table III in Appendix B.

of the distribution of sales does *not* stem from the tail part of the distribution of total factor productivity.

Figure 7 shows the power law exponents of total factor productivity for different industries. The three lines represent 2006, 2007, and 2008. We see that μ_A is small for banks, credit agencies, securities, commodity brokers, and services, suggesting that total factor productivity in these industries is significantly unequal across firms. It may be the case that, in these industries, the tail of sales distribution is generated by the tail of productivity distribution. On the other hand, μ_A is relatively large for food stores, restaurants, and automobile repair shops, suggesting that total factor productivity in these industries is not so unequal across firms. It is highly likely that

the tail of sales distribution in these industries comes not from the tail of productivity distribution, but from the tail of the number of workers distribution or from the tail of tangible fixed assets distribution.

§5. Production function for value added

So far we have used firm sales as a measure of output. Firm sales data is easy to obtain since it is disclosed by almost all firms in their financial statements, irrespective of which country and which industry a firm belongs to. This is particularly important when one seeks to conduct cross country analysis as we did in the previous sections. However, some economists prefer to use value added as a measure of output, since it may not be appropriate to ignore the role of raw materials and intermediate products in the process of production, at least for some firms belonging to a certain industries. In this section, we check how the results in the previous sections will change when we use value added instead of sales as a measure of output. Unfortunately, value added for each firm is available only for Japanese manufacturing firms. We will use the dataset provided by Teikoku Databank Ltd. for about 21,000 Japanese manufacturing firms in 2008.

Value added and sales are related as:

$$V = S - F, \quad (5.1)$$

where V represents value added and F is intermediate inputs, consisting of the expenditures on raw materials, purchased parts, transportations, outsourcing, and so on. We start with looking at the correlations between the value-added to sales ratio, which is defined by $1 - F/S$, and L or K . The left panel of Fig. 8 shows how the distribution of the value-added to sales ratio depends on the value of L . Specifically, we divide the range of L , which is given by $10 \leq L < 10^{3.5}$, evenly into five. For each of the five sub-regions of L , we present a probability density function of the value-added to sales ratio conditional on that L belongs to that sub-range. We see that the five PDFs are almost identical, indicating that the value-added to sales ratio is independent of L . Similarly, the right panel of Fig. 8 shows the conditional distributions of the value-added to sales ratio for different values of K . The conditional PDFs are almost identical irrespective of the value of K , suggesting again that the value-added to sales ratio is independent of K .

Given that the value-added to sales ratio is correlated neither with L nor K , we can rewrite Eq. (2.10) as

$$V = A' K^\alpha L^\beta, \quad (5.2)$$

where A' is defined as $A' = (1 - F/S)A$. Using Eq. (5.2), we estimate total factor productivity as we did in the previous sections. Figure 9 shows the cumulative distribution function of A' , together with that of A . We confirm that the tail part of the distribution of A' is characterized by a power law, and that the power law exponent associated with A' is almost identical with the exponent associated with A .

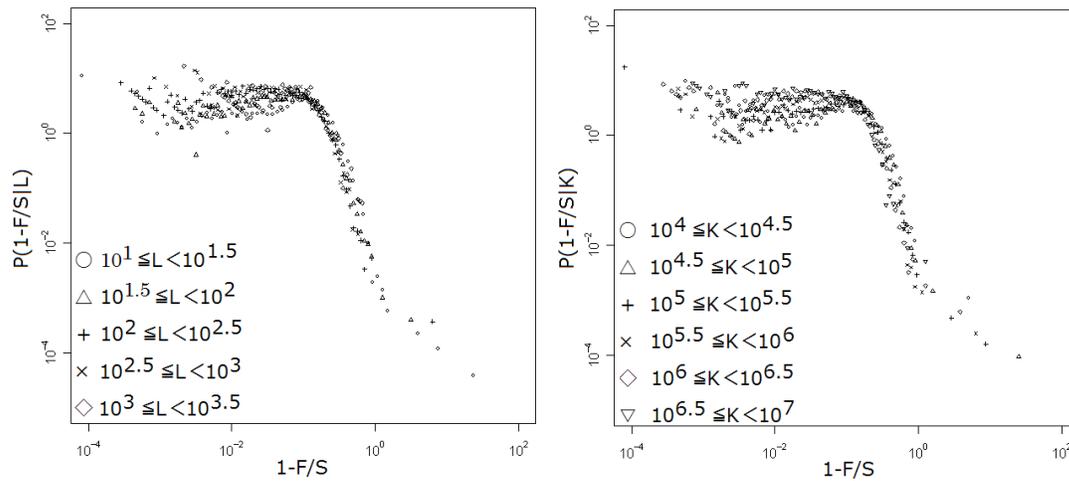


Fig. 8. PDF of the value-added to sales ratio for different values of L (left panel) and for different values of K (right panel). The range of L , $10 \leq L < 10^{3.5}$, is divided evenly into five, while the range of K , $10^4 \leq K < 10^7$, divided evenly into six. The data for Japanese manufacturing firms in 2008 is used.

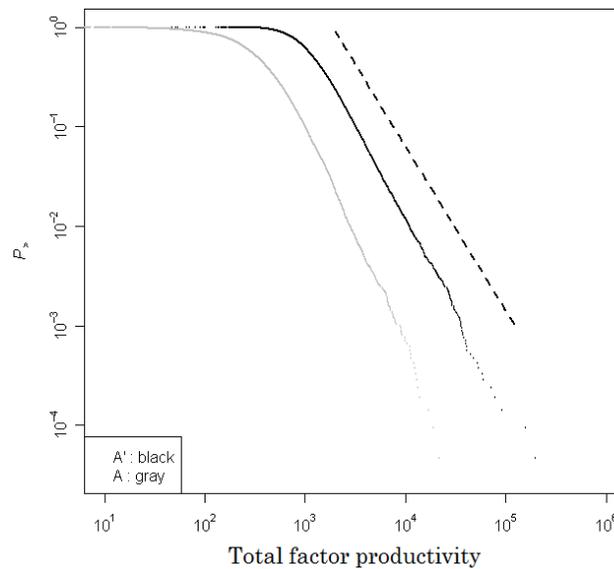


Fig. 9. CDFs of A and A' for Japanese manufacturing firms in 2008. The reference line is a power law with an exponent of 1.65.

§6. Conclusion

In this paper we have estimated total factor productivity for mega firms from 30 countries. We found that a productivity distribution is characterized by a power law tail, as in the case of other firm size variables, including firm sales, the number of workers, and tangible fixed assets. However, the power law exponent associated with a productivity distribution tends to be greater than those for the other firm size variables. We have confirmed this tendency for different countries, as well as

for different industries. This result suggests that the tail part of a sales distribution does not stem from the tail of a productivity distribution, but from the tail of the distributions of the number of workers or tangible fixed assets.

Acknowledgements

The authors thank the Yukawa Institute for Theoretical Physics at Kyoto University, where this work was completed during the YITP-W-11-04 on “Econophysics 2011”. We would also like to thank Yayoi Hatano for her careful explanation of the data. This work was supported in part by a Grant-in-Aid for Scientific Research (C) (No. 20510147) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

Appendix A

— Reversal Quasi-Symmetry and Gibrat’s Law —

Let us assume that, in the two dimensional space defined by $\log u$ and $\log v$, the joint probability $P_J(u, v)$ is invariant under the exchange of variables with respect to the line defined by $\log v = \theta \log u + \log a$ ($v = au^\theta$). That is,

$$P_J(u, v) = P_J\left(\left(\frac{v}{a}\right)^{1/\theta}, au^\theta\right), \quad (\text{A}\cdot 1)$$

where a and θ are parameters. Assume further that the conditional probability $P\left(\frac{v}{au^\theta} \mid u\right)$ does not depend on u ; namely,

$$P\left(\frac{v}{au^\theta} \mid u\right) = P\left(\frac{v}{au^\theta}\right). \quad (\text{A}\cdot 2)$$

These two equations represent “reversal quasi-symmetry” and “Gibrat’s law”, respectively. One can show that, given these two equations, each of the variables u and v follows a power law distribution.

$$P_{>}(u) \propto u^{-\mu_u}, \quad (\text{A}\cdot 3)$$

$$P_{>}(v) \propto v^{-\mu_v}. \quad (\text{A}\cdot 4)$$

This is a simple application of the results obtained in Ref. 13). Reference 14) extended their results to the three dimensional space. Specifically, it considered reversal quasi-symmetry and Gibrat’s law in the three dimensional space defined by $\log u_1$, $\log u_2$, and $\log v$. On the one hand, Eqs. (2.4)–(2.6) in the 2D projective planes imply that the invariance of the joint probability with respect to the space $\log v = \theta_1 \log u_1 + \theta_2 \log u_2 + \log a$ ($v = au_1^{\theta_1}u_2^{\theta_2}$) holds. That is,

$$P_J(u_1, u_2, v) = P_J\left(\left(\frac{v}{au_2^{\theta_2}}\right)^{1/\theta_1}, \left(\frac{v}{au_1^{\theta_1}}\right)^{1/\theta_2}, au_1^{\theta_1}u_2^{\theta_2}\right). \quad (\text{A}\cdot 5)$$

On the other hand, Eqs. (2.7)–(2.9) imply that the conditional probability is independent of the initial values:

$$P\left(\frac{v}{au_1^{\theta_1}u_2^{\theta_2}} \mid u_1, u_2\right) = P\left(\frac{v}{au_1^{\theta_1}u_2^{\theta_2}}\right). \quad (\text{A}\cdot 6)$$

These two equations represent reversal quasi-symmetry and Gibrat's law in the three dimensional space. Thus one can show that each of the three variables, u_1 , u_2 , and v , follows a power law distribution:

$$P_{>}(u_1) \propto u^{-\mu_{u_1}}, \quad (\text{A}\cdot 7)$$

$$P_{>}(u_2) \propto u^{-\mu_{u_2}}, \quad (\text{A}\cdot 8)$$

$$P_{>}(v) \propto v^{-\mu_v}. \quad (\text{A}\cdot 9)$$

Appendix B

— Country Code and Standard Industrial Classification (SIC) Code —

Table II. Country code.

BE	Belgium	BG	Bulgaria	CN	China
CZ	Czech	DE	Germany	DK	Denmark
EE	Estonia	ES	Spain	FI	Finland
FR	France	GB	The U.K.	GR	Greece
HR	Croatia	IE	Ireland	IT	Italy
JP	Japan	KR	Korea	LT	Lithuania
LV	Latvia	NL	The Netherlands	NO	Norway
PL	Poland	PT	Portugal	RO	Romania
RS	Serbia	RU	Russia	SE	Sweden
SI	Slovenia	SK	Slovak	US	United States

Table III. Standard Industrial Classification (SIC) code.

01-09	Forestry and Fishery
10-19	Mining, Building, Construction
20-29	Food, Textile, Wood, Chemicals
30-39	Rubber, Leather, Metal, Machinery
40-49	Railroad, Transportation, Postal, Telecom, Electricity, Gas, Water Supply
50-59	Wholesale, Retailing, Eatery
60-69	Bank, Finance, Security, Insurance, Real estate, Investor
70-79	Services
80-89	Medical, Legal, Educational, Social Services
90-99	International Affairs and Non-Operating Establishments

References

- 1) V. Pareto, *Cours d'Economie Politique* (Macmillan, London, 1897).
- 2) H. A. Simon and C. P. Bonini, *The American Economic Review* **48** (1958), 67.
- 3) R. E. Quandt, *The American Economic Review* **56** (1966), 416.
- 4) Y. Ijiri and H. A. Simon, *J. of Political Economy* **82** (1974), 315.
- 5) K. Okuyama, M. Takayasu and H. Takayasu, *Physica A* **269** (1999), 125.
- 6) M. H. R. Stanley, S. V. Buldyrev, S. Havlin, R. Mantegna, M. A. Salinger and H. E. Stanley, *Economics Letters* **49** (1995), 453.
- 7) T. Mizuno, M. Katori, H. Takayasu and M. Takayasu, *Empirical Science of Financial Fluctuations — The Advent of Econophysics* (Springer-Verlag, Tokyo, 2002), p. 321.
- 8) R. L. Axtell, *Science* **293** (2001), 1818.
- 9) S. Fujimoto, A. Ishikawa, T. Mizuno and T. Watanabe, submitted to *Economics*.
- 10) R. N. Mantegna and H. E. Stanley, *Nature* **376** (1995), 46.
- 11) T. Mizuno, M. Takayasu and H. Takayasu, *Physica A* **332** (2004), 403.
- 12) A. Ishikawa, S. Fujimoto and T. Mizuno, *Physica A* **390** (2011), 4273.
- 13) Y. Fujiwara, C. D. Guilmi, H. Aoyama, M. Gallegati and W. Souma, *Physica A* **335** (2004), 197.
- 14) A. Ishikawa, S. Fujimoto, T. Watanabe and T. Mizuno, submitted to *Eur. Phys. J. B*.
- 15) Y. Ikeda and W. Souma, *Prog. Theor. Phys. Suppl. No. 179* (2009), 93.
- 16) H. Aoyama, H. Yoshikawa, H. Iyetomi and Y. Fujiwara, *Prog. Theor. Phys. Suppl. No. 179* (2009), 80.
- 17) T. Watanabe, T. Mizuno, A. Ishikawa and S. Fujimoto, *The Economic Review* **62** (2011), 193.
- 18) H. S. Houthakker, *The Review of Economic Studies* **23** (1955), 27.
- 19) S. Rosen, *Economica* **45** (1978), 235.
- 20) C. I. Jones, *Quarterly J. of Economics* **120** (2005), 517.
- 21) C. W. Cobb and P. H. Douglass, *The American Economic Review* **18** (1928), 139.
- 22) U.S. Department of Labor, <http://www.dol.gov/>
- 23) Y. Malevergne, V. Pisarenko and D. Sornette, *Phys. Rev. E* **83** (2011), 036111.
- 24) A. H. Jessen and T. Mikosch, *Publ. de l'Inst. Math.* **94** (2006), 171.
- 25) X. Gabaix, *Annual Review of Economics* **1** (2009), 255.

Labor Productivity Distribution with Negative Temperature

Hiroshi IYETOMI

Department of Physics, Niigata University, Niigata 950-2181, Japan

Exhaustive financial data of firms in Japan enables us to shed light on how the labor productivity, defined here as value added produced by one worker in a year, is diverse across firms and workers. Statistical equilibrium theory reinforced with the concept of negative temperature turns out to be useful to explain the empirical facts on a major part of the distribution of workers over labor productivity states, where particle and single-particle energy are replaced by worker and labor productivity, respectively. The zero-temperature state in the negative temperature regime corresponds to the optimized state for the current mainstream economics, where all workers are allocated to a state of the highest productivity. Significant difference in temperature is observed between the manufacturing and nonmanufacturing sectors. The negative temperature in the nonmanufacturing sector is three times lower than that in the manufacturing sector, indicating that the former may suffer from a much wider demand gap. In contrast, the two sectors are almost in equilibrium with respect to exchange of workers.

§1. Introduction

The neoclassical theory, adopted by economists in the mainstream, postulates that the economy is in a full-employment state where productivity of workers is optimized, that is, all workers are allocated to an economic sector of the highest productivity. According to our ordinary experiences, this seems to be too ideal. On the other hand, the Keynesian theory emphasizes that workers are always subject to variation of demand. In such a demand-driven economy workers are distributed across sectors of different productivity.

Along the line laid by Keynes, a statistical physics theory to explain diversity of labor productivity has been developed by Yoshikawa¹⁾ and Aoki and Yoshikawa;²⁾ productivity distribution is described in terms of the Boltzmann form. The idea is an economic analogue of statistical distribution theory of particles over single-particle energy states in thermal equilibrium; worker and labor productivity correspond to particle and single-particle energy, respectively and total demand of an economic system substitutes for total energy of a physical system. Of course this is a macroscopic way to understand the behavior of workers. Certainly each of them always makes his or her own decision in daily life. But there may exist a fundamental statistical law that individuals are not able to get rid of. Recently the statistical equilibrium theory has been revisited with a dynamic probabilistic formulation by Scalas and Garibaldi.^{3),4)}

In a series of recent papers,^{5)–10)} the author and his collaborators have demonstrated that productivity of workers is in fact widely scattered in Japan and the US. References 5), 6) and 9) observed that workers belonging to the Japanese listed firms are distributed in a power-law form in the high productivity end and Refs. 7) and 8) confirmed that the same is true for the US. Also the power-law tail of the productivity

distribution was observed in European countries such as Italy and France by different authors.^{11),12)} An idea of superstatistics was employed to explain such diverging behavior of the productivity⁹⁾ and large-scale financial data covering small-to-medium sized firms as well as large firms was used for extended study.¹⁰⁾

We remark that all of the empirical results, except for Fig. 13 in Ref. 9), are for the productivity distributions as regards workers, firms or industrial sectors. However, those distributions are not the one predicted by the statistical equilibrium theory due to Aoki and Yoshikawa.^{1),2)} It is distribution of workers across states with different productivity as shown in the figure in Ref. 9). That is, information on average number of workers in a state with given productivity provides us with a most naive comparison between the theoretical and the empirical results. To this end we take full advantage of exhaustive financial data of firms in Japan, the same data as used in Ref. 10). This enables us to address the third problem on productivity dispersion left in Ref. 9): contradiction between the principle of statistical physics and that of neoclassical economics. We will see that the concept of negative temperature gives a solution to the problem.

In the next section empirical facts on the labor productivity are established. A brief review of the statistical theory for labor productivity is given in §3. Equilibrium conditions between different subsystems as regards exchange of demands and distinguishable workers are derived in §4. The Boltzmann form is then adapted to the empirical results in §5; the economic temperature takes a negative value. The section 6 finally summarizes the principal results obtained here together with discussion on directions for future research.

§2. Empirical results in Japan

Production activities of firms may be characterized by value added Y produced by a firm with given input factors such as labor L and capital K ; Y is often regarded as a function of K and L . A full-scale nationwide database on financial states of Japanese firms has been constructed by combining data on small- and medium-sized firms collected by the Credit Risk Database (CRD) Association and data on listed large firms by Nikkei Electronic Database System (NEEDS). Sum of fixed assets and number of workers were substituted for K and L , respectively. Since added value Y is not a primary financial quantity, it was calculated through accumulation of various contributions such as labor costs and profits. The database thus includes the financial information on about half a million of firms in the period of 2000 through 2006.

Figures 1 and 2 show K - L - Y scatter plots for firms belonging to the manufacturing and nonmanufacturing sectors in 2004. We see that firms are distributed in a teardrop shape in the financial space, irrespective of selection of the sectors. Such plots give us a clear perspective on growth of firms. The main body of the teardrop is an incubating domain for small-to-medium-sized firms. Its fat shape reflects diversity of their characteristics, a prerequisite for vital economy. Then some of them begin to grow, being focused sharply in a certain direction. This indicates that there exist necessary economic conditions shared by firms which are successful in their

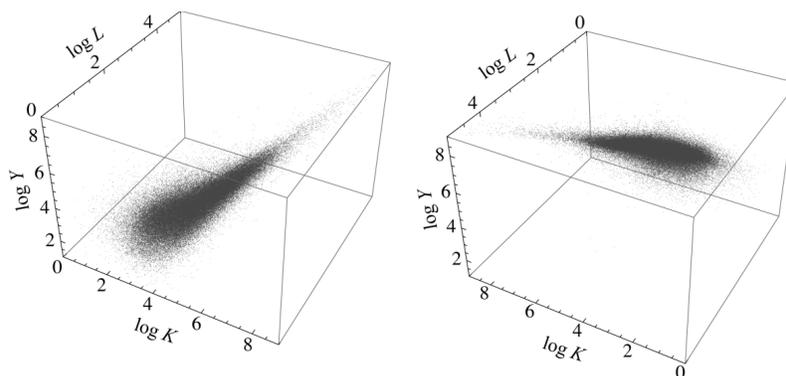


Fig. 1. K - L - Y scatter plot for the manufacturing sector with 109,856 firms in 2004.

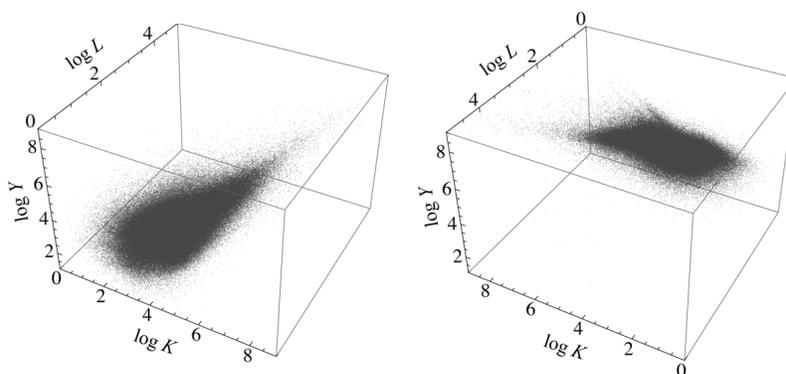


Fig. 2. K - L - Y scatter plot for the nonmanufacturing sector with 371,922 firms in 2004.

business.

We define the labor productivity c_i of firm i , a key quantity in this work, as

$$c_i = \frac{Y_i}{L_i}, \tag{2.1}$$

where Y_i and L_i are the value added and the number of workers of firm i . Then we can calculate c_i using the data as shown in Figs. 1 and 2. The results for the distribution $n(c)$ of workers over productivity states are compiled in Fig. 3, along with cumulative share as regards firms and workers counted from the upper end of productivity. We observe a sharp cusp around $c = c^*$ dividing $n(c)$ into two regimes in either of the two sectors:

$$c^* = \begin{cases} 30 \text{ Myen} & \text{for the manufacturing sector,} \\ 60 \text{ Myen} & \text{for the nonmanufacturing sector.} \end{cases} \tag{2.2}$$

The distribution $n(c)$ drastically changes its functional behavior across c^* . In the lower productivity side of $c < c^*$, $n(c)$ increases in an almost monotonic way as c is raised. In the higher productivity side of $c > c^*$, on the contrary, $n(c)$ is roughly a decreasing function; the jaggy behavior arises from inadequate statistics.

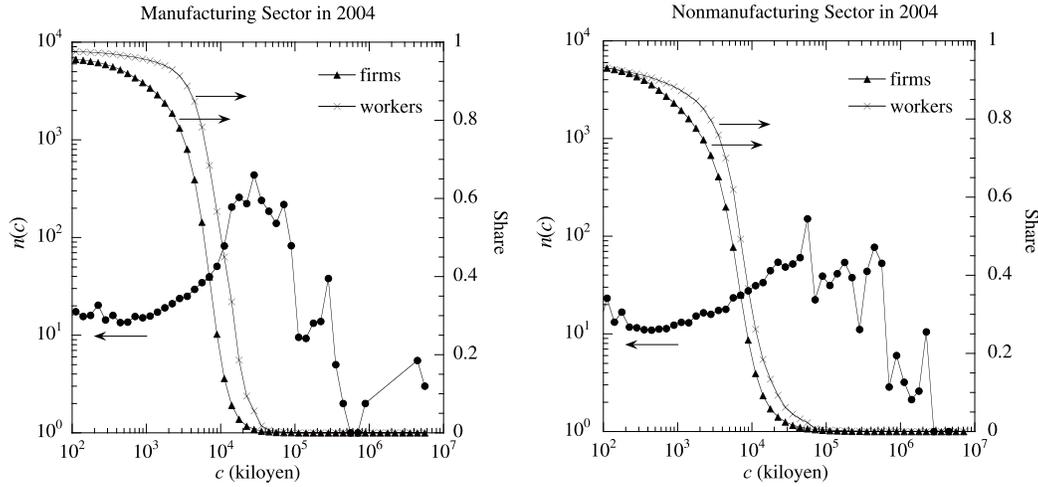


Fig. 3. Empirical results for distributions of the labor productivity of Japanese firms in 2004. Note that both axes are logarithmic scales.

We note that about 99.5% of firms and 98.2% of workers belong to the lower productivity side in the manufacturing sector and 99.5% of firms and 99.2% of workers, in the nonmanufacturing sector. Here we thus concentrate our attention on the major part of $n(c)$ in both the sectors.

§3. Statistical physics theory

The model adopted to describe diversity of labor productivity is nothing but a noninteracting Potts model in statistical physics. Workers in an economic system correspond to spins in the Potts model and states of different productivity levels available for workers, to energy levels of spins. We remark that workers, being macroscopic objects, are as distinguishable as spins which are fixed in space.

The total number W of configurations for worker's allocation in which the i -th state ($i = 1, \dots, K$) with productivity c_i is occupied by n_i workers is given by

$$W(\{n_i\}) = \frac{N!}{n_1! n_2! \cdots n_K!} . \quad (3.1)$$

Here N stands for the total number of workers:

$$N = \sum_{i=1}^K n_i . \quad (3.2)$$

Also the total (effective) demand D is calculated according to

$$D = \sum_{i=1}^K c_i n_i . \quad (3.3)$$

We suppose that the total number N of workers and the total demand D are both conserved during a relaxation process for allocation of workers.

Entropy of the system is then defined as

$$S(\{n\}) := \log W(\{n\}) \simeq N \log N - \sum_{i=1}^K n_i \log n_i, \quad (3.4)$$

where we used Stirling's approximation assuming $n_i \gg 1$. The distribution of workers in equilibrium is obtained by maximizing the entropy (3.4) under the two constraints (3.2) and (3.3). The variational condition for the entropy is written as

$$\delta S - \alpha \delta N - \beta \delta D = 0, \quad (3.5)$$

where Lagrange multipliers α and β are introduced and those are related to temperature T and chemical potential μ through

$$\beta = \frac{1}{T}, \quad (3.6)$$

$$\alpha = -\frac{\mu}{T}. \quad (3.7)$$

Finally, solution of Eq. (3.5) yields a Boltzmann distribution of workers over productivity:

$$n(c) = N\lambda \exp\left(-\frac{c}{T}\right), \quad (3.8)$$

where λ , corresponding to the fugacity in thermodynamics, is defined as

$$\lambda = e^{\mu/T}. \quad (3.9)$$

The economic temperature T measures to what extent the total demand of an economic system diminishes away from its maximum. Tracking year-to-year variation of the temperature thus provides us with in-depth information on business fluctuations. An alternative meanings of T along with that of $N\lambda$ will be discussed in the next section in relation to equilibrium conditions between subsystems.

§4. Equilibrium conditions

Economic systems are generally inhomogeneous, since they may be decomposed into various components such as industrial sectors, regional sectors and business groups. As has been shown in §2, we separated the Japanese workers to the manufacturing and nonmanufacturing sectors. Here we thus derive necessary conditions for such subsystems of workers to be in equilibrium to each other. We are required to modify the standard derivation¹³⁾ of those conditions in physics, because we have to take into account distinguishability of workers.

Let us suppose that an economic system of workers consists of two subsystems A and B with $N_A(N_B)$ workers and demand of $D_A(D_B)$. The total number W_{A+B} of microscopic states of the whole system is calculated as

$$W_{A+B}(N_A, N_B; D_A, D_B) = \frac{N!}{N_A!N_B!} W_A(N_A, D_A) W_B(N_B, D_B), \quad (4.1)$$

with $N = N_A + N_B$ and $D = D_A + D_B$. The prefactor in the right-hand side of Eq. (4.1), counting ways to distribute workers between the two subsystems, stems from the distinguishability of workers. Then we obtain the entropy of the whole system as

$$S_{A+B} \simeq S_A + S_B + N \log N - N_A \log N_A - N_B \log N_B . \quad (4.2)$$

This formula destroys the additivity of entropy. Conversely, we learn that identical particles are indistinguishable in nature, so that we do not need the prefactor in the right-hand side of Eq. (4.1). But here it should be.

The entropy maximum principle determines the most probable distribution of workers and demands between the two subsystems. Since the total number of workers and the total demand are conserved, the variational equation is given as

$$\begin{aligned} \delta S_{A+B} &= \delta S_A + \delta S_B - \delta N_A \log N_A - \delta N_B \log N_B \\ &= (\alpha_A - \alpha_B - \log N_A + \log N_B) \delta N_A + (\beta_A - \beta_B) \delta D_A \\ &= 0 , \end{aligned} \quad (4.3)$$

where δN_A and δD_A denote the infinitesimal variations of N_A and D_A which are independent of each other. We thus obtain the following equilibrium conditions:

$$T_A = T_B , \quad (4.4)$$

$$N_A \lambda_A = N_B \lambda_B . \quad (4.5)$$

The first condition (4.4), which shares the same form as in physics, is necessary for the system to be in equilibrium against exchange of demands. The second condition (4.5) guarantees no macroscopic flow of workers between the two subsystems. This is unfamiliar to physicists because the number of workers appears in either hand of it; the distinguishability of workers has changed its normal form. Demand flows from subsystem A to subsystem B if $T_A > T_B$ and in the reversed direction if $T_A < T_B$. Also workers flow from A to B if $N_A \lambda_A > N_B \lambda_B$ and vice versa in the opposite case.

§5. Negative temperature

We are in a position to apply the Boltzmann form (3.8) with T and $N\lambda$ as adjustable parameters to the empirical results for the productivity dispersion obtained in §2. However we first recall that the average number of workers increases as the productivity is raised. This empirical fact necessitates us to adopt the concept of negative temperature.

Temperature is normally positive for physical systems, which means that the population of particles in a single-particle energy state reduces as the energy is increased. If a physical system have an upper bound in its single-particle energy levels, population inversion of particles characterized by negative temperature can be artificially realized by pumping energy into the system. In the economic system, on the other hand, states of negative temperature are naturally formed with c^* , given by Eq. (2.2), as the maximum productivity. The zero-temperature state in the

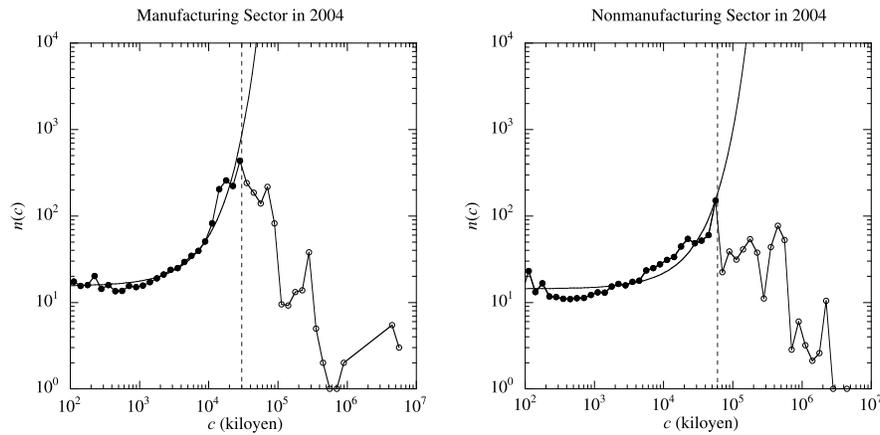


Fig. 4. Fitting of the labor productivity distributions in terms of the Boltzmann formula. Note that both axes are logarithmic scales.

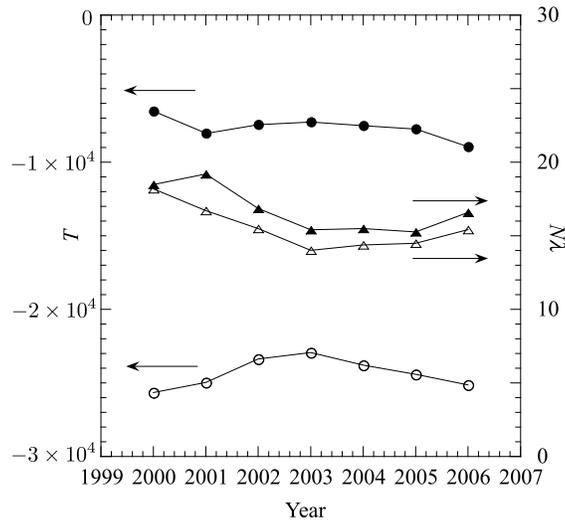


Fig. 5. Fitted results for the economic temperature T (left-hand ordinate) and the fugacity multiplied by number of workers $N\lambda$ (right-hand ordinate). The circles and triangles denote the results for T and $N\lambda$, respectively. The solid and open symbols distinguish between those for the manufacturing and nonmanufacturing sectors.

negative temperature regime corresponds to allocation of all workers to a state of the highest productivity. This is the optimized state on which the current mainstream economics (neoclassical theory) is constructed.

The results for $n(c)$ in Fig. 3 were fitted to the Boltzmann form (3-8), in which the parameters T and $N\lambda$ were determined using the least-square method. The data in $100 \text{ kiloyen} < c < c^*$ were used for the fitting procedure as shown in Fig. 4. Also the same fitting procedure was applied to the data in other years ranging from 2000 through 2006. Figure 5 summarizes the results for T and $N\lambda$ thus obtained in the two sectors.

We observe the temperature of the nonmanufacturing sector is three times as low as that of the manufacturing sector. The two sectors are thus far away from

equilibrium in demand exchange. This fact indicates the nonmanufacturing sector may suffer from lack of total demand much more than the manufacturing sector. Although the temperature for the nonmanufacturing sector gradually increases from 2000 up to 2003, it turns over its trend in 2003. In contrast, the two sectors are almost in equilibrium with respect to exchange of workers; the difference in $N\lambda$ is confined within 15% over the whole period. These results are quite understandable. Conversion of demands between sectors of different industries is generally difficult, because each demand tends to be closely tied up with a specific sector. Relaxation in demand thus takes a long time. On the other hand, workers can change their jobs across industrial sectors. Such adaptability of human beings leads to smooth exchange of workers between different sectors.

§6. Conclusion

We have demonstrated promising applicability of the statistical physics methodology to outstanding basic issues in the macroeconomy such as diversity of labor productivity. Most of workers are distributed over states of varied productivity in accordance with the Boltzmann distribution function characterized by *negative* temperature; the average number of workers in a firm thus increases with raised productivity. It should be noted that the use of the nationwide financial data on firms was crucial for us to find such a fundamental statistical law for the labor productivity distribution. The two key parameters T and $N\lambda$, measured through fitting of the full-scale data to the Boltzmann form, give a new insight into the macroeconomy. The economic temperature T evaluates how much the total demand of an economic system is close to its maximum. Also it quantifies imbalance in demand between subsystems. The economic fugacity multiplied by the number of workers, $N\lambda$, tells us to what extent subsystems are equilibrated with respect to allocation of workers.

Our measurement elucidates that the nonmanufacturing sector has a much larger gap between the maximum and actual demands as compared with the manufacturing sector over the period under this study, 2000–2006. In contrast, the two sectors are almost in equilibrium with respect to exchange of workers over the whole period. We are now proceeding in the following directions:

- to elucidate the degree of equilibration in exchange of demands and workers among industrial sectors of more detailed classification,
- to understand the behavior of the worker distribution on the supreme productivity side, where the average firm size shrinks with increased productivity, as well as that on the main-body side in a unified way.

The results will be reported elsewhere.

Acknowledgements

The author thanks the Yukawa Institute for Theoretical Physics at Kyoto University, where this work was completed during the YITP workshop (YITP-W-11-04) entitled “Econophysics 2011 — The Hitchhikers’s Guide to the Economy”. And he would like to thank Hideaki Aoyama, Yoshi Fujiwara, Yuichi Ikeda, Wataru Souma

and Hiroshi Yoshikawa for continual discussion on the present and related subjects. The exhaustive financial data used here was kindly made accessible to him by the Credit Risk Database Association. This work was also supported in part by *the Program for Promoting Methodological Innovation in Humanities and Social Sciences by Cross-Disciplinary Fusing* of the Japan Society for the Promotion of Science and by the Ministry of Education, Science, Sports and Culture, Grants-in-Aid for Scientific Research (B), No. 22300080 (2010–12).

References

- 1) H. Yoshikawa, Japanese Econ. Rev. **54** (2003), 1.
- 2) M. Aoki and H. Yoshikawa, *Reconstructing Macroeconomics: A Perspective from Statistical Physics and Combinatorial Stochastic Processes* (Cambridge University Press, New York, 2007).
- 3) E. Scalas and U. Garibaldi, Economics: The Open-Access, Open-Assessment E-Journal **3** (2009), 2009-15,
<http://dx.doi.org/10.5018/economics-ejournal.ja.2009-15>
- 4) U. Garibaldi and E. Scalas, *Finitary Probabilistic Methods in Econophysics* (Cambridge University Press, New York, 2010).
- 5) H. Aoyama, H. Yoshikawa, H. Iyetomi and Y. Fujiwara, Prog. Theor. Phys. Suppl. No. 179 (2009), 80.
- 6) H. Aoyama, Y. Fujiwara, Y. Ikeda, H. Iyetomi and W. Souma, Economics: The Open-Access, Open-Assessment E-Journal **3** (2009), 2009-22,
<http://www.economics-ejournal.org/economics/journalarticles/2009-22>
- 7) W. Souma, Y. Ikeda, H. Iyetomi and Y. Fujiwara, Economics: The Open-Access, Open-Assessment E-Journal **3** (2009), 2009-14,
<http://www.economics-ejournal.org/economics/journalarticles/2009-14>
- 8) Y. Ikeda and W. Souma, Prog. Theor. Phys. Suppl. No. 179 (2009), 93.
- 9) H. Aoyama, H. Yoshikawa, H. Iyetomi and Y. Fujiwara, J. Econ. Interact. Coord. **5** (2010), 27.
- 10) Y. Ikeda, W. Souma, H. Aoyama, Y. Fujiwara and H. Iyetomi, Eur. Phys. J. B **76** (2010), 491.
- 11) T. Di Matteo, T. Aste and M. Gallegati, Eur. Phys. J. B **47** (2005), 459.
- 12) C. Di Guilmi, F. Clementi, T. Di Matteo and M. Gallegati, J. Econ. Interact. Coord. **3** (2008), 43.
- 13) L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 3rd Ed., Part 1 (Butterworth-Heinemann, Oxford, 1980).

Subcommunities and Their Mutual Relationships in a Transaction Network

Takashi IINO^{*)} and Hiroshi IYETOMI

*Department of Physics, Niigata University,
Niigata 950-2181, Japan*

We investigate a Japanese transaction network consisting of about 800 thousand firms (nodes) and four million business relations (links) with focus on its modular structure. Communities detected by maximizing modularity often are dominated by firms with common features or behaviors in the network, such as characterized by regions or industry sectors. However, it is well known that the modularity optimization approach has a resolution limit problem, that is, it fails in identifying fine communities buried in large communities. To unfold such hidden structures, we apply the community detection to each of subnetworks formed by isolating those communities from the whole body. Subcommunities thus identified are composed of firms with finer regions, more specified sectors or business affiliations. Also we introduce a new idea of reduced modularity matrix to measure the strength of relations between (sub)communities.

§1. Introduction

Networks formed by firms through their mutual transactions are a manifestation of economic activities. This is a new way to study economic phenomena emphasizing importance of interaction between economic agents. A number of researches on complex networks have been carried out from a physical point of view.^{1),2)} The endeavors encompass development of statistical mechanics methods for quantifying network structure, construction of theoretical models for network formation and visualization of networks based on a physical model.

In general, real complex networks have nonuniform distribution of links. For example, the World Wide Web is not a random network; websites may tend to make a link to a site in the same category. In friendship networks, it is easier for us to establish friendship with a person who is a friend of one of our friends than unfamiliar persons in the world. Such nonuniform networks are often characterized by communities in which nodes are divided into densely connected groups, and on the other hand those groups are sparsely connected. Communities often contain nodes that have common features or behaviors in networks. The community analysis provides us with a coarse-graining view on structure of large complex networks.

To detect communities, Newman proposed a measure called modularity to evaluate the performance of division of a network into modules.³⁾⁻⁵⁾ The idea of modularity is based on quantifying statistically unforeseen arrangements of links. Finding the division with the highest modularity value determines an optimized community structure of a network.

Here we investigate community structure of a Japanese transaction network

^{*)} E-mail: iino@ad.sc.niigata-u.ac.jp

consisting of about 800 thousand firms and four million business relations.^{*)} In the transaction network, nodes and links correspond to firms and transaction relationships, respectively. The scale of the network virtually covers all of production activities in Japan, so that its analysis is expected to throw a new insight into the nationwide real economy. Community analysis together with visualization on the whole network was carried out in Refs. 6) and 7). The network was separated into tens of major communities and a lot of minor communities. The modularity optimization thus worked well to demonstrate that the network is quite nonuniform. Also a similar study on a partial network of manufacturers based on the same data was reported in Ref. 8).

However, it is well known that the approach using the modularity has a resolution limit problem.⁹⁾ The modularity optimization often fails to identify small but important communities buried in large communities. This indicates that the dominant communities in the transaction network may have more detailed nonuniform structures. Social organizations tend to generate hierarchical structures. For instance, Japan has several regions and each of those regions is composed of several prefectures. It may thus happen that the communities themselves are characterized by lower-level social organizations. Other causes for formation of such communities include development of supply chains in industry sectors and establishment of business affiliations.

To address the above issue in this paper, we elucidate community structures within the dominant communities of the transaction network that have been already extracted. Hereafter, the word of “community” represents a community in the whole network, and the word of “subcommunity” stands for a community in the community. Also we pay attention to relative positions of those subcommunities. We thereby evaluate quantitatively business relationships between the subcommunities by formulating a measure in terms of reduced components of the modularity.

This paper is organized as follows. In §2 the methodology to resolve a network up to the subcommunity level is explained. It is an extensive use of modularity with introduction of a new idea of reduced modularity matrix which enables us to measure the strength of relationship between (sub)communities. Section 3 is devoted to discussion of major communities in the whole network. Then the methodology developed here is applied to elucidation of internal structures of the dominant communities in §4. Concluding remarks are given in §5. A preliminary account of this paper has been reported in Ref. 10).

§2. Methodology

2.1. Recursive community detection

Let us begin with explaining a basic methodology to resolve the transaction network up to the subcommunity level. We first single out communities in the transaction network by maximizing its modularity Q . Then we define a subnetwork

^{*)} The data on firms used in this paper were collected by TOKYO SHOKO RESEARCH, LTD. in 2005 through financial statements, corporate documents and hearing-based survey.

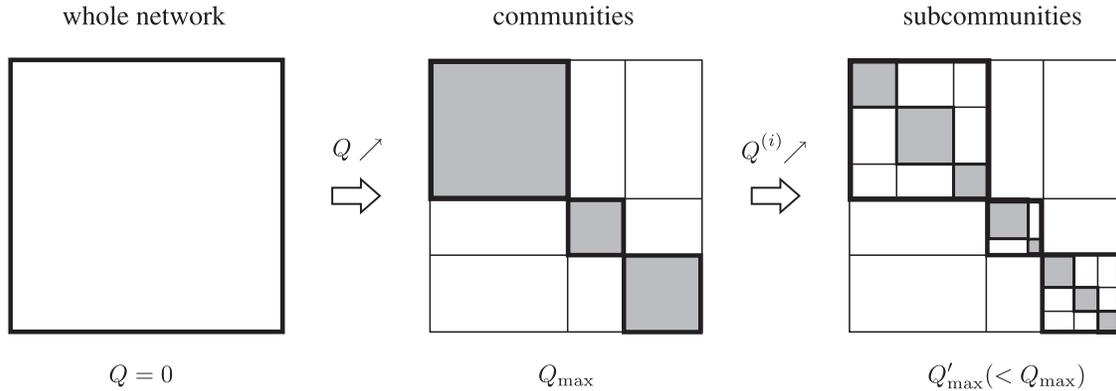


Fig. 1. Schematic view of the process to detect communities and subcommunities of a network in an adjacency matrix representation. The modularity Q of the whole network is first maximized to determine an optimized community structure with Q_{\max} . Then maximization of the modularity $Q^{(i)}$ for the subnetwork of community i further divides the group into components. It should be noted that the resulting modularity Q'_{\max} for the whole network resolved to the subcommunity level is lower than Q_{\max} .

so as to be composed only of firms belonging to a given community i by cutting off links connected to different communities, and decompose the subnetwork so obtained into groups by maximizing its modularity $Q^{(i)}$. We repeat the decomposition process for each of the subnetworks constructed from major communities. Although such decomposition of the network leads to reduction of the modularity, it has possibilities to unveil more detailed structures (subcommunities) involved in the communities. The procedure for the recursive community detection is depicted in Fig. 1.

It is actually impossible to search all possibilities for the optimized division of a large network such as the transaction network under study, because the modularity maximization is an NP-hard problem.¹¹⁾ For practical purposes approximate methods are employed, including greedy agglomeration,^{5),12),13)} simulated annealing,^{14)–16)} extremal optimization¹⁷⁾ and spectral methods.³⁾ We adopt a bisection method⁶⁾ which has simulated annealing built in. It spends significantly more computational time than the greedy algorithm. Although there is a trade-off between computational time and performance for optimization methods, the bisection method can detect communities in the submillion transaction network in realistic time with a much better optimized Q value. Details of the method are described in Ref. 6).

2.2. Modularity matrix

How strongly (sub)communities are connected to each other is a key concern in this paper, which enables us to delve into nonuniform structure of the transaction network. The modularity matrix³⁾ is a suitable quantity to this end. The element B_{lm} of the modularity matrix between nodes l and m for a network with the total number M of links is defined as

$$B_{lm} = A_{lm} - \frac{k_l k_m}{2M}, \quad (2.1)$$

where A_{lm} is the corresponding element of the adjacency matrix of the network and $k_l = \sum_m A_{lm}$ is degree of node l , the total number of links associated with the node. The modularity Q for a given partition of the network is expressible in terms of the modularity matrix as

$$Q = \frac{1}{2M} \sum_{i=1}^L \sum_{l \in V_i} \sum_{m \in V_i} B_{lm} , \quad (2.2)$$

where we assume that all nodes in the network are partitioned to L groups denoted by V_i ($i = 1, 2, \dots, L$). Maximization of Q decomposes the network into groups in an optimal way.

2.3. Reduced modularity matrix

We project the modularity matrix onto (sub)community space by taking partial summation within the groups as

$$q_{ij} = \frac{1}{2M} \sum_{l \in V_i} \sum_{m \in V_j} B_{lm} = e_{ij} - a_i a_j , \quad (2.3)$$

where e_{ij} and a_i are defined by

$$\begin{aligned} e_{ij} &= \frac{1}{2M} \sum_{l \in V_i} \sum_{m \in V_j} A_{lm} , \\ a_i &= \frac{1}{2M} \sum_{l \in V_i} k_l . \end{aligned} \quad (2.4)$$

The quantity e_{ij} measures the fraction of links connecting two groups V_i and V_j , and a_i represents the fraction of links associated with V_i . If links are randomized in the network under a constraint that degree of each node is fixed, the expected value of e_{ij} for such a random network is given by $a_i a_j$. We refer to the matrix defined by $\mathbf{q} = \{q_{ij}\}$ as reduced modularity matrix. We note that trace of the reduced modularity matrix amounts to the modularity Q :

$$Q = \text{Tr} \mathbf{q} = \sum_{i=1}^L q_{ii} . \quad (2.5)$$

As shown below, the off-diagonal element q_{ij} ($i \neq j$) of the reduced modularity matrix is related to increment of the modularity when two groups are combined into one group with the remainders untouched. We compare modularity Q of a partition $C = \{V_1, V_2, \dots, V_L\}$ with modularity Q' of the resulting partition C' in which two groups V_i and V_j in C are merged into $V_h = V_i \cup V_j$. The link density within the group V_h is calculated as

$$e_{hh} = \frac{1}{2M} \sum_{l \in V_i \cup V_j} \sum_{m \in V_i \cup V_j} A_{lm}$$

$$\begin{aligned}
&= \frac{1}{2M} \left(\sum_{l \in V_i} \sum_{m \in V_i} A_{lm} + \sum_{l \in V_j} \sum_{m \in V_j} A_{lm} + \sum_{l \in V_i} \sum_{m \in V_j} A_{lm} + \sum_{l \in V_j} \sum_{m \in V_i} A_{lm} \right) \\
&= e_{ii} + e_{jj} + e_{ij} + e_{ji}.
\end{aligned} \tag{2.6}$$

And the fraction of links connected to V_h is given by

$$a_h = \frac{1}{2M} \sum_{l \in V_i \cup V_j} k_l = \frac{1}{2M} \left(\sum_{l \in V_i} k_l + \sum_{l \in V_j} k_l \right) = a_i + a_j. \tag{2.7}$$

Increment $\Delta Q = Q' - Q$ of modularity is thus calculated as

$$\begin{aligned}
\Delta Q &= e_{hh} - a_h^2 - (e_{ii} - a_i^2 + e_{jj} - a_j^2) \\
&= e_{ij} + e_{ji} - a_i a_j - a_j a_i \\
&= q_{ij} + q_{ji} \\
&= 2q_{ij},
\end{aligned} \tag{2.8}$$

where the last line is valid only for undirected networks, with $q_{ij} = q_{ji}$.

The reduced modularity matrix for the communities that are obtained by maximizing the modularity should have *negative* off-diagonal elements. Otherwise one could obtain a larger value of the modularity by merging the two communities associated with the negative element into one. On the contrary, the reduced modularity matrix for the subcommunities has *positive* off-diagonal elements. This statement of course depends on how to define the reduced modularity matrix for the subcommunities. In this paper, it is always calculated for the whole network with all of links including those across the communities. Partition of nodes by the subcommunities thus have a room to optimize the modularity by unifying two of them, that is, $\Delta Q > 0$ in Eq. (2.7); see Fig. 1.

2.4. Community and subcommunity distance

A large value of q_{ij} indicates that two (sub)communities i and j are strongly related to each other, irrespective of its sign. For communities with $q_{ij} < 0$, the situation means $\Delta Q \simeq 0$. Thus merging of the two communities does not harm the optimized decomposition at all. This is interpreted as a sign of that those are in fact tightly connected. For subcommunities with $q_{ij} > 0$, the situation tells that merging of the two subcommunities gives rise to a large gain of the modularity, which can be regarded as a manifestation of strong coupling of them.

Here we define a distance d_{ij} between (sub)communities i and j in the reduced modularity matrix as

$$d_{ij} = \max(q_{ij}) - q_{ij}, \tag{2.9}$$

which takes account of the fact that distance should behave inversely to their closeness and take a positive value. More closely related are two groups, shorter is the distance between them. This distance is extensively used to evaluate the strength of relationship between (sub)communities throughout this paper.

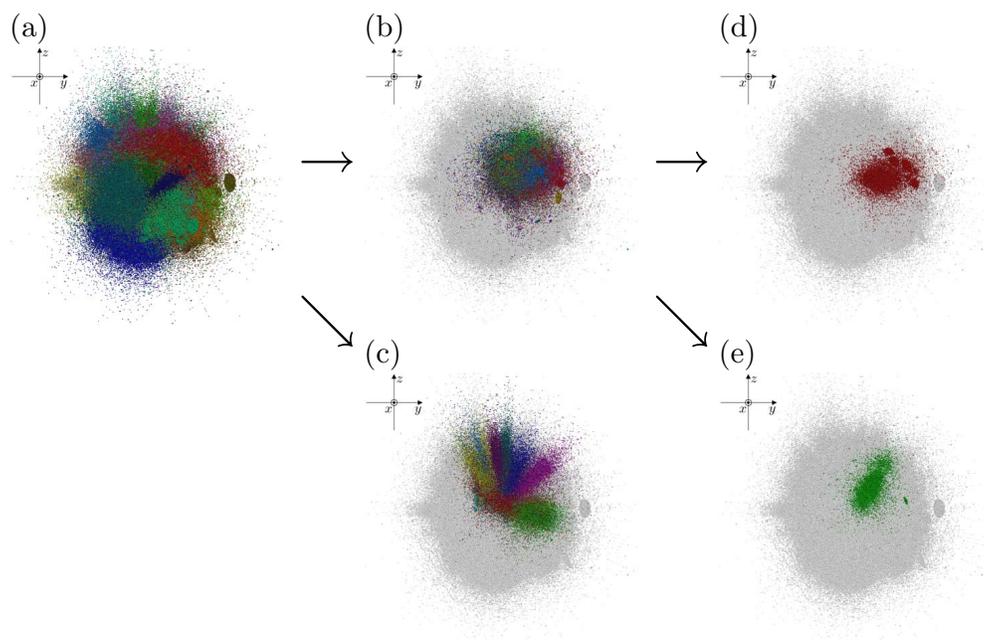


Fig. 2. The Japanese transaction network drawn in three-dimensional space by a spring-electrical model.⁷⁾ Dots in these images represent nodes (firms) whose communities are distinguished by colors. The image (a) shows the whole network. The images (b) and (c) illuminate only the first and second largest communities, respectively. And those communities are further decomposed into subcommunities as displayed with different colors. The first and second largest subcommunities in the largest community are selected in (d) and (e), respectively.

§3. Communities

We purified the transaction data by removing firms having bankruptcy flag set on or lacking important information such as classification of industry sector. To simplify our community analysis, we ignore direction of transactions between firms so that the data is treated as an undirected network. Although the resulting network is not a connected one, the largest connected component with 773,670 nodes and 3,192,582 links encompasses more than 99 percent of the total nodes and links; the remaining connected components have at most eight nodes. Therefore, we pay attention to only the largest connected component here. The whole part of the transaction network is shown in Fig. 2(a), where a spring-electrical model⁷⁾ optimally arranges nodes for visualization in three-dimensional space.

We extracted 118 communities in the transaction network with $Q_{\max} = 0.654$.*) The community structure thus obtained is well noticeable on the images in Fig. 2(a) as has been already demonstrated in Ref. 6). The community size distribution is shown in Fig. 3, where the size of communities is measured by how many firms they

*) This number of communities significantly differs from that given in Ref. 6), where 1,352 communities were detected with $Q_{\max} = 0.653$. The previous result was obtained for the network containing all of the connected components, while the present result is based on the largest connected component. This is the origin of the difference in the number of communities. However, the results for the community detections do not change so much.

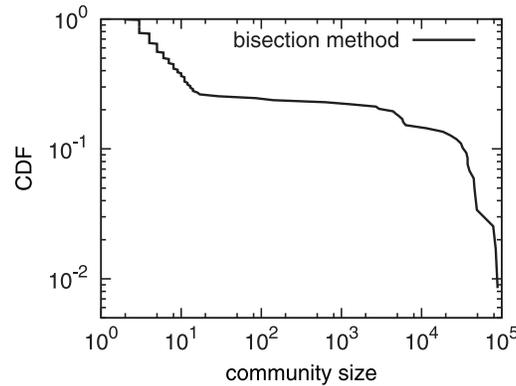


Fig. 3. Cumulative distribution function (CDF) for size of the communities in the transaction network obtained by a bisection method;⁶⁾ the size of a community is number of firms in it.

have. There are 17 large communities that have over 10,000 nodes and about 90 tiny communities that have less than 20 nodes. The top 10 communities cover about 71% nodes in the whole network. Those colossal communities^{*)} are characterized by attributes of firms belonging to them in Table I, where prefectures and industry sectors of firms are looked up community by community. Industry sectors conformed to Japan Standard Industrial Classification are shown in Table II, which lists only sectors appearing in this paper. To simplify notations, the names of industry sectors are abbreviated by three capitals.

We see that each of the dominant communities is well featured in regions and industry sectors of the constituent firms. For instance, the largest community is composed of manufactures and wholesalers of machinery to a large extent. The second, fourth, fifth, sixth and ninth largest communities can be distinguishable as construction groups. Especially among them, the fourth and ninth communities have strong local characteristics. The fraction of firms located in Kyushu area in the fourth community and that of firms in Tohoku area in the ninth community both exceed 90%. The third and eighth largest communities mainly consist of firms dealing with food and apparel, respectively. The seventh largest community mainly contains firms of transport. The 10th largest community is characterized by firms of printing and manufacturers of paper products.

We then elucidate the strength of connections between the dominant communities as given in Table I. The reduced modularity matrix for the communities, calculated through Eq. (2.3), plays a key role. The numerical result is depicted in Fig. 4. To have a more comprehensive view on relationship between the communities, we constructed such a dendrogram as shown in Fig. 5. This is a result of the cluster analysis with the distance defined by Eq. (2.9). Each end of branches in the dendrogram is labeled rank in size of the associated community. Figure 5 enables us to find that the Japanese economy has a trilateral structure. One can separate the

^{*)} The largest community corresponds to the third largest community in Table 3 of Ref. 6). The second, third, fourth and fifth largest communities correspond to the second, first, fifth and fourth largest communities detected previously.

Table I. Attributes of firms in the 10 largest communities detected for the transaction network. Top three major prefectures and industry sectors (see Table II for the meanings of abbreviations) are listed in each community. Decimals shown in the parentheses represent occupation ratios of firms with specified attributes.

rank	size	prefecture	industry
1	88,840	Tokyo (0.189)	M-GM (0.144)
		Aichi (0.120)	W-ME (0.124)
		Osaka (0.110)	M-FM (0.105)
2	84,280	Niigata (0.117)	C-GE (0.400)
		Tokyo (0.094)	C-SP (0.228)
		Aichi (0.086)	C-EI (0.075)
3	78,529	Tokyo (0.110)	W-FB (0.262)
		Hokkaido (0.088)	M-FO (0.172)
		Aichi (0.055)	R-FB (0.137)
4	48,903	Fukuoka (0.298)	C-GE (0.359)
		Kagoshima (0.137)	C-SP (0.146)
		Kumamoto (0.134)	C-EI (0.124)
5	47,085	Aichi (0.135)	C-GE (0.362)
		Kanagawa (0.121)	C-SP (0.205)
		Tokyo (0.113)	W-BM (0.094)
6	45,622	Tokyo (0.152)	C-EI (0.473)
		Kanagawa (0.112)	C-GE (0.101)
		Osaka (0.098)	W-ME (0.072)
7	44,736	Tokyo (0.167)	T-RF (0.163)
		Aichi (0.094)	R-MB (0.123)
		Osaka (0.083)	R-FH (0.120)
8	39,591	Tokyo (0.180)	R-DA (0.251)
		Osaka (0.118)	W-TA (0.201)
		Aichi (0.080)	M-AP (0.132)
9	37,524	Fukushima (0.197)	C-GE (0.357)
		Aomori (0.167)	C-SP (0.159)
		Miyagi (0.154)	C-EI (0.089)
10	37,488	Tokyo (0.320)	M-PR (0.175)
		Osaka (0.096)	W-MI (0.127)
		Aichi (0.068)	M-PP (0.077)

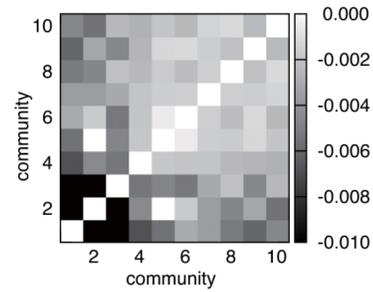


Fig. 4. Reduced modularity matrix for the 10 largest communities in the whole network.

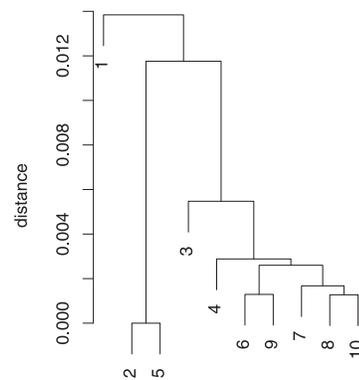


Fig. 5. Dendrogram derived from the reduced modularity matrix as given in Fig. 4.

whole system into three groups by choosing any cutoff distance in the wide range of 0.006 through 0.011. The largest community itself forms one group. This is the manufacturing sector leading the Japanese economy, as will be ascertained by the subcommunity analysis in §4. The second and fifth communities, so tightly connected, form an alternative construction-base group. And the rests are assembled into the third group, in which the communities of food industry (third), transport (seventh), apparel (eighth) and printing (10th) participate together with the fourth, sixth and ninth communities of construction. Also we find the construction sector, one of the leading economic players in Japan, is separable into two parts.

Table II. Abbreviation of industry sectors in conformity with Japan Standard Industrial Classification defined by Statistics Bureau in Ministry of Internal Affairs and Communications. Although wholesale and retail trade in the divisions are merged into one category in the standard classification, we distinguish those two sectors here.

Abbr.	Major groups	Divisions	
C-EI	equipment installation work		
C-GE	construction work, general, including public and private construction work	construction	
C-SP	construction work by specialist contractor, except equipment installation work		
M-AP	manufacture of apparel and other finished products made from fabrics and similar materials		
M-BT	manufacture of beverages, tobacco and feed	manufacturing	
M-EM	manufacture of electrical machinery, equipment and supplies		
M-EP	electronic parts and devices		
M-FM	manufacture of fabricated metal products		
M-FO	manufacture of food		
M-GM	manufacture of general machinery		
M-PL	manufacture of plastic products, except otherwise classified		
M-PP	manufacture of pulp, paper and paper products		
M-PR	printing and allied industries		
M-RU	manufacture of rubber products		
M-TR	manufacture of transportation equipment		
R-DA	retail trade (dry goods, apparel and apparel accessories)		retail trade
R-FB	retail trade (food and beverages)		
R-FH	retail trade (furniture, household utensil and household appliance)		
R-MB	retail trade (motor vehicles and bicycles)	transport	
T-RF	road freight transport		
W-BM	wholesale trade (building materials, minerals and metals, etc.)	wholesale trade	
W-FB	wholesale trade (food and beverages)		
W-ME	wholesale trade (machinery and equipment)		
W-MI	miscellaneous wholesale trade		
W-TA	wholesale trade (textile and apparel)		

§4. Subcommunities

As remarked in the Introduction, the transaction network is quite nonuniform. We are now ready to reiterate the community analysis on each of the dominant communities detected in the previous section to elucidate such nonuniform structure in the network.

Table III. Results of the subcommunity analysis for the top 10 communities shown in Table I.

rank	number of subcommunities	optimized modularity	character of subcommunities
1	48	0.525	industry-sectoral
2	57	0.728	regional
3	88	0.598	regional
4	67	0.729	prefectural
5	76	0.722	marginally regional
6	116	0.685	marginally regional
7	69	0.772	industry-sectoral
8	49	0.638	marginally regional
9	48	0.740	prefectural
10	85	0.629	industry-sectoral

The results of the subcommunity analysis applied to the top 10 communities are summarized in Table III. The maximized values of $Q^{(i)}$ for the subnetworks are either similar to or even well surpassing $Q_{\max} = 0.654$, the maximized modularity for the whole network. This indicates the major communities have appreciable modular structures at the resolution level of subcommunities. Especially, half of them with $Q_{\max}^{(i)} > 0.7$, which are construction-base or transport-base groups, are expected to have very clear subcommunity structures in the light of modularity.

To elucidate characteristic features of the subcommunities thus detected, we examined attributes of firms in them as has been done for firms in the communities. The subcommunities are likewise characterized by geographical regions or industry sectors. The first, seventh and 10th largest communities are decomposed into groups with more specific industry classifications. The remainders have regional characteristics, more or less. The fourth and ninth largest communities are broken up even into pieces characterized by prefectures.^{*)} Figures 2(b) through 2(d) illustrate how the largest and second largest communities are decomposed into subcommunities. The detailed structures are well appreciable in the network visualized by the spring-electrical model.

We have already unveiled that the Japanese economy is organized from three components in the previous section. The largest, second largest, and third largest communities belong to those groups separately. Here we concentrate our efforts on exploring internal structures of the three principal communities.

We first show attributes of firms in the top 10 subcommunities in the largest community. Table IV summarizes the results. We see that the largest community is separated into subcommunities each of which has its own industry-sectoral characteristics. The largest subcommunity contains representative electric-appliance makers as hub nodes, including Matsushita, Toshiba and Hitachi. The second largest subcommunity is an automobile manufacturing industry cluster led by Toyota and its group firms such as Denso and Aishin. The third largest subcommunity, a machinery-oriented group, has Amada, Fanuc and NSK. The fourth largest subcommunity is

^{*)} Japan has eight regions which are constituted by several neighboring prefectures except that Hokkaido itself forms a region; the total number of prefectures is 47.

Table IV. Attributes of firms in the 10 largest subcommunities detected for the subnetwork of the largest community. The notations are the same as those in Table I.

rank	size	prefecture	industry
1	17,020	Tokyo (0.329)	W-ME (0.157)
		Kanagawa (0.131)	M-EP (0.107)
		Osaka (0.101)	M-EM (0.095)
2	11,152	Aichi (0.722)	M-GM (0.202)
		Gifu (0.073)	M-FM (0.133)
		Mie (0.048)	W-ME (0.095)
3	9,872	Osaka (0.199)	M-GM (0.290)
		Tokyo (0.135)	W-ME (0.200)
		Shizuoka (0.081)	M-FM (0.118)
4	9,561	Shizuoka (0.237)	M-GM (0.154)
		Tokyo (0.157)	M-FM (0.121)
		Kanagawa (0.103)	M-TR (0.112)
5	7,337	Tokyo (0.249)	W-ME (0.186)
		Osaka (0.135)	M-EM (0.144)
		Kyoto (0.110)	M-GM (0.132)
6	6,781	Nagano (0.293)	M-GM (0.196)
		Tokyo (0.172)	M-FM (0.160)
		Kanagawa (0.067)	M-EM (0.071)
7	6,356	Tokyo (0.264)	M-FM (0.102)
		Osaka (0.133)	M-GM (0.096)
		Ibaraki (0.105)	W-ME (0.094)
8	3,773	Tokyo (0.251)	M-PL (0.250)
		Osaka (0.172)	M-GM (0.096)
		Gunma (0.101)	M-FM (0.082)
9	3,397	Niigata (0.367)	M-FM (0.214)
		Tokyo (0.102)	W-ME (0.125)
		Osaka (0.089)	W-MI (0.120)
10	2,957	Tokyo (0.214)	M-RU (0.143)
		Osaka (0.186)	W-ME (0.102)
		Hyogo (0.085)	M-GM (0.096)

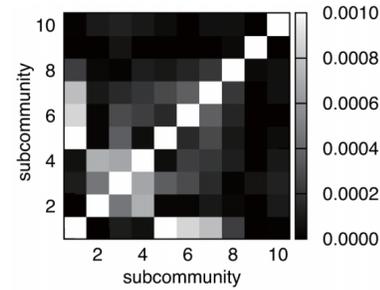


Fig. 6. Reduced modularity matrix for the major subcommunities in the largest community.

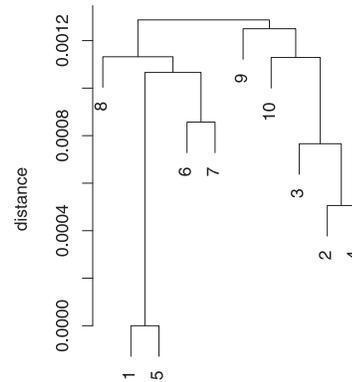


Fig. 7. Dendrogram corresponding to the reduced modularity matrix in Fig. 6.

another automobile manufacturing industry group in which Honda, Yamaha and Nissan are core members. The fifth largest subcommunity is characterized by the representative hub nodes such as Fujifilm, Asahi Glass and Shimazu. The reduced modularity matrix for the subcommunities in the largest community is depicted in Fig. 6. The dendrogram derived from the matrix is shown in Fig. 7. It is widely known that the automotive and electrical industries form two main streams in the manufacturing industry in Japan. This fact is reflected quite well on the dendrogram.

The results for the second largest community are shown in Table V. Its subcommunities are characterized to a large extent by geographical regions. This fact is understandable recalling that the community is a cluster of firms in the construction industry, in which local connections are very important. The reduced modularity matrix and the associated dendrogram are given in Figs. 8 and 9, respectively. The

Table V. Same as Table IV, but for the subnetwork of the second largest community.

rank	size	prefecture	industry
1	14,135	Tokyo (0.321)	C-SP (0.349)
		Kanagawa (0.187)	C-GE (0.192)
		Osaka (0.107)	W-BM (0.067)
2	9,941	Niigata (0.966)	C-GE (0.402)
		Tokyo (0.010)	C-SP (0.207)
		Shizuoka (0.002)	C-EI (0.119)
3	7,978	Aichi (0.765)	C-GE (0.384)
		Gifu (0.160)	C-SP (0.290)
		Mie (0.012)	W-BM (0.056)
4	7,215	Osaka (0.311)	C-GE (0.525)
		Hyogo (0.284)	C-SP (0.135)
		Kyoto (0.160)	W-BM (0.067)
5	6,654	Chiba (0.236)	C-GE (0.490)
		Tokyo (0.204)	C-SP (0.162)
		Kanagawa (0.090)	W-BM (0.065)
6	6,633	Shizuoka (0.946)	C-GE (0.364)
		Kanagawa (0.011)	C-SP (0.288)
		Aichi (0.011)	C-EI (0.129)
7	5,522	Tochigi (0.615)	C-GE (0.439)
		Gunma (0.162)	C-SP (0.195)
		Ibaraki (0.133)	C-EI (0.062)
8	4,485	Ishikawa (0.947)	C-GE (0.420)
		Toyama (0.009)	C-SP (0.200)
		Osaka (0.007)	C-EI (0.122)
9	4,436	Fukui (0.955)	C-GE (0.456)
		Ishikawa (0.008)	C-SP (0.150)
		Shiga (0.007)	C-EI (0.131)
10	3,976	Kanagawa (0.779)	C-GE (0.479)
		Tokyo (0.126)	C-SP (0.181)
		Chiba (0.020)	W-BM (0.071)

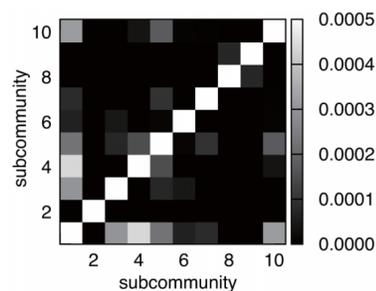


Fig. 8. Reduced modularity matrix for the major subcommunities in the second largest community.

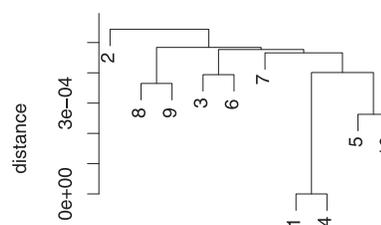


Fig. 9. Dendrogram corresponding to the reduced modularity matrix in Fig. 8.

flat internal structure of the community, observed through its dendrogram, highly contrasts with the hierarchical structure of the largest community as revealed in Fig. 7. This indicates the subcommunities occupy rather equal positions. Such relationship between the subcommunities conforms to the radial shape of the community in Fig. 2(c).

Table VI demonstrates the third largest community is separated into subcommunities characterized by well-defined geographical regions; the original community itself has no such regional characteristics. This community is a food industry cluster as has been identified. The reduced modularity matrix for the subcommunities is shown in Fig. 10 along with the associated dendrogram in Fig. 11. We see there are strong relations between subcommunities located closely to each other. In the dendrogram, the subcommunities placed on the right-hand side are mainly eastern groups such as located in Hokkaido, Tohoku and north Kanto regions. In contrast, the subcommunities on the left-hand side are characterized by western prefectures

Table VI. Same as Table IV, but for the subnetwork of the third largest community.

rank	size	prefecture	industry
1	10,907	Hokkaido (0.169)	W-FB (0.387)
		Tokyo (0.103)	M-FO (0.222)
		Shizuoka (0.069)	R-FB (0.055)
2	9,082	Tokyo (0.203)	M-FO (0.267)
		Aichi (0.128)	W-FB (0.250)
		Osaka (0.088)	R-FB (0.110)
3	7,574	Tokyo (0.178)	W-FB (0.257)
		Osaka (0.070)	R-FB (0.231)
		Aichi (0.067)	M-BT (0.128)
4	6,809	Fukuoka (0.235)	W-FB (0.230)
		Kagoshima (0.172)	M-FO (0.216)
		Miyazaki (0.141)	R-FB (0.120)
5	6,663	Hiroshima (0.140)	W-FB (0.263)
		Kagawa (0.126)	M-FO (0.204)
		Okayama (0.116)	R-FB (0.127)
6	5,294	Fukushima (0.204)	W-FB (0.198)
		Iwate (0.163)	M-FO (0.195)
		Aomori (0.163)	R-FB (0.155)
7	5,024	Tokyo (0.166)	W-FB (0.365)
		Kanagawa (0.084)	R-FB (0.181)
		Osaka (0.069)	M-FO (0.136)
8	4,947	Hokkaido (0.886)	R-FB (0.276)
		Tokyo (0.031)	W-FB (0.176)
		Osaka (0.007)	M-FO (0.133)
9	3,579	Tokyo (0.225)	W-FB (0.299)
		Tochigi (0.179)	M-FO (0.183)
		Gunma (0.131)	R-FB (0.084)
10	3,113	Aichi (0.220)	W-FB (0.275)
		Ishikawa (0.184)	M-FO (0.209)
		Toyama (0.165)	R-FB (0.119)

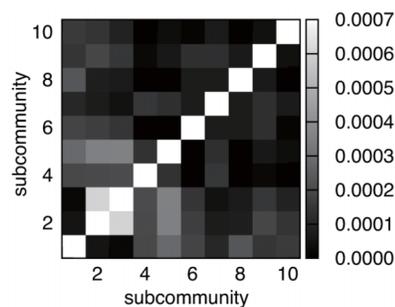


Fig. 10. Reduced modularity matrix for the major subcommunities in the third largest community.

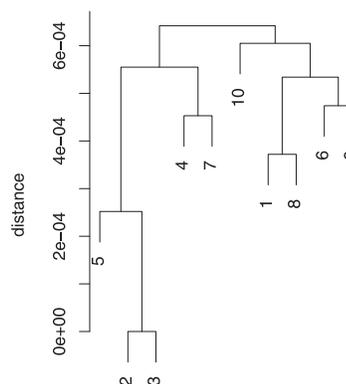


Fig. 11. Dendrogram corresponding to the reduced modularity matrix in Fig. 10.

in Chugoku, Shikoku and Kyusyu regions. And Tokyo bridges the two sides. Foods have expiration dates and some of them are prevailing locally, so that the firms in the food community may tend to establish business relations with their neighbors.

We also carried out the same analysis on the other communities in Table I. The geographical distance and industrial closeness between subcommunities are important factors to understand the strength of relationship between them. Of course the two factors provide us with no perfect solution for our understanding.

§5. Concluding remarks

In this paper we have resolved the transaction network formed by submillion Japanese firms into components up to the subcommunity level taking advantage of the recursive community detection technique. A wide variety of components were thus identified and the internal structures of the three major communities were eluci-

dated in view of regions, industry sectors and business affiliations. Furthermore, we succeeded in measuring the strength of relationship between the (sub)communities by our new tool based on the reduced modularity matrix. It is no exaggeration to say that the current economic activities arise from connections between firms; a typical example is formation of supply chains. We thus assure that the present study open an alternative and useful way to overview the industrial structure in Japan.

Acknowledgements

The authors thank the Yukawa Institute for Theoretical Physics at Kyoto University, where this work was completed during the YITP workshop (YITP-W-11-04) entitled “Econophysics 2011 — The Hitchhikers’s Guide to the Economy”. This work was also supported, in part, through *the Program for Promoting Methodological Innovation in Humanities and Social Sciences by Cross-Disciplinary Fusing* and Grants-in-Aid for Scientific Research (B), No. 22300080 (2010-12), provided by the Japan Society for the Promotion of Science and by the Ministry of Education, Science, Sports and Culture, respectively. We thank Hideaki Aoyama, Yoshi Fujiwara, Yuichi Ikeda and Wataru Souma for useful discussions on the subjects treated here. We also thank the Research Institute of Economy, Trade and Industry (RIETI) for having provided us with the data on transaction relationship between firms in Japan.

References

- 1) R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74** (2002), 47.
- 2) M. E. J. Newman, *SIAM Rev.* **45** (2003), 167.
- 3) M. E. J. Newman, *Proc. Natl. Acad. Sci.* **103** (2006), 8577.
- 4) M. E. J. Newman and M. Girvan, *Phys. Rev. E* **69** (2004), 026113.
- 5) M. E. J. Newman, *Phys. Rev. E* **69** (2004), 066133.
- 6) T. Iino, K. Kamehama, H. Iyetomi, Y. Ikeda, T. Ohnishi, H. Takayasu and M. Takayasu, *J. Phys.: Conf. Ser.* **221** (2010), 012012.
- 7) K. Kamehama, T. Iino, H. Iyetomi, Y. Ikeda, T. Ohnishi, H. Takayasu and M. Takayasu, *J. Phys.: Conf. Ser.* **221** (2010), 012013.
- 8) Y. Fujiwara and H. Aoyama, *Eur. Phys. J. B* **77** (2010), 565.
- 9) S. Fortunato and M. Barthélemy, *Proc. Natl. Acad. Sci.* **104** (2007), 36.
- 10) T. Iino and H. Iyetomi, in *Intelligent Decision Technologies, SIST 10* (Springer, Berlin, 2011), p. 537.
- 11) U. Brandes, D. Delling, M. Gaertler, Görke, M. Hoefer, Z. Nikoloski and D. Wagner, *IEEE Transactions on Knowledge and Data Engineering* **20** (2008), 172.
- 12) A. Clauset, M. E. J. Newman and C. Moore, *Phys. Rev. E* **70** (2004), 066111.
- 13) V. D. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre, *J. Stat. Mech.* **10** (2008), P10008.
- 14) R. Guimerà and L. A. N. Amaral, *Nature* **433** (2005), 895.
- 15) A. Medus, G. Acuña and C. O. Dorso, *Physica A* **358** (2005), 593.
- 16) J. Reichardt and S. Bornholdt, *Phys. Rev. E* **74** (2006), 016110.
- 17) J. Duch and A. Arenas, *Phys. Rev. E* **72** (2005), 027104.

Omori Law after Large-Scale Destruction of Production Network

Yoshi FUJIWARA

Graduate School of Simulation Studies, University of Hyogo, Kobe 650-0047, Japan

After the two giant earthquakes, Kobe 1995 and East Japan 2011, a large-scale destruction took place on supplier-customer network. These primary and exogenous shocks were propagated on the production network, which caused a secondary effect resulting in chained bankruptcies. By employing data of bankruptcies occurring in a neighbor to the primarily damaged firms or regional economy, we show that the number of neighboring failures obeys a Omori law, a power-law relaxation. This finding implies that the recovery from such a huge shock on production network is much more sluggish than one can naively expect.

§1. Introduction

Several complex systems have abnormally long-time relaxation after extreme events. An example is what we call *Omori law* today, which is a power-law decay of aftershock activity with time after a large earthquake. This law was discovered in the pioneering work of seismology by Fusakichi Omori during the 1890's.¹⁾ His discovery was based on the Great Nobi (Mino-Owari) Earthquake in 1891, the largest inland earthquake in the historical records of Japan, two other earthquakes and their aftershocks. After the centenary of his discovery, an accumulated data of aftershocks due to the Nobi event until today reveals the validity of Omori law, surprisingly long relaxation time for more than 10^4 days after the earthquake (see Fig. 1 in the review paper²⁾).

Social and economic systems as well as natural systems have been known to exhibit long-time relaxation processes similar to the Omori law. Examples include financial market crashes,³⁾ book sales ranking dynamics accompanying shocks,⁴⁾ and so forth. See Refs. 3) and 4) and the references therein for example. These empirical findings are addressing the same question: how is the dynamics of such a system affected and reacting to extreme events or shocks? This question is important to the understanding the system, predicting recovery processes after shocks and, in some cases, precursory signals.

The Great East Japan Disaster on March 11, 2011 was a huge shock to the domestic and overseas economy. The supplier-customer or production network was severely damaged at a nation-wide scale. Firstly, a mass destruction of firms and a set of intervention in many industrial sectors and geographical locations took place in such a devastating way that a number of firms instantly ceased their production activity. Secondly, financially fragile firms went into bankruptcies and defaults. If such failed firms are irreplaceable nodes in supplier-customer links, many places in the production network were severely damaged. Thirdly, such a damage was so serious that other firms in upstream or downstream side of the network eventually went into bankruptcies or financially ill-conditions as a kind of *Tsunami* effect.

In this paper, we shall show a preliminary result that Omori law holds for the

“aftershocks” following after a large-scale destruction of production network. Aftershocks are defined and measured by bankruptcies, which were directly caused by other failed firms due to such a large-scale shock. The identification of causes for bankruptcies was done carefully by credit research agencies in their datasets. By using such datasets, failed *and* linked nodes can be identified so that one can count the number of failed nodes in the *neighbor* of the nodes damaged by the initial shock, which is our definition of aftershock. We employ two episodes of nation-wide shocks, the Kobe Earthquake in 1995 and the Great East Japan Earthquake in 2011 for empirical tests.

Section 2 describes aftershocks defined by bankruptcies and a basic idea behind the method of identifying neighboring failures. In §3, we shall examine the validity of Omori law by using two episodes of the Kobe and East Japan. The aftershocks are quantitatively characterized by the cumulative number of chained bankruptcies, which are shown to satisfy a modified Omori formula. This finding would provide potentially useful estimation on how long and large those aftershocks of defaults can last. We shall discuss these points as well as things to be investigated further in §4. Section 5 summarizes the paper.

§2. Data of neighboring bankruptcies as aftershocks

According to the 2011 White Paper on Small and Medium Enterprises in Japan,⁷⁾ a large number of firms were directly affected by the East Japan Disaster due to the earthquake, tsunami, nuclear and other disasters, especially along the north-east coast. This primary and exogenous shock resulted in business failures, mostly temporary but often permanent, in a considerable fraction in the economy.

Because more than 0.7 million firms were present in the region among 2 million firms in the entire country, the suppliers and customers, who depend their business on the damaged firms and the regional economy, were affected afterward. For example, suppliers had a delay, or often a loss, in the receipt of accounts receivable, causing an abrupt drop in sales which may have deteriorated their financial conditions subsequently.

To measure such secondary effects, we shall focus on bankruptcies. A bankruptcy or business failure is a critical financial state of a firm; its debt dominates its balance-sheet so that it has little equity, and the firm cannot no longer manage its business. Because the secondary effects propagate along the supplier-customer network, a bankruptcy may be regarded as a fracture under a stress strengthened in the neighbor of preceding increase of stress. We assume that the process of stress propagation can be traced by observing such neighboring bankruptcies on the production network.

We employ two datasets by leading credit research agencies in Tokyo, which carefully identified the causes of bankruptcies in exhaustive lists of domestic failures.^{5), 6)} Let us illustrate the method of identification in the dataset of Tokyo Shoko Research, Inc.⁶⁾ It is exhaustive in the sense that all bankruptcies with debt exceeding 10 million yen (roughly equal to 0.1 million dollars or euros) are recorded. Causes for bankruptcies are investigated and classified as follows.

1. “solo” failure
 - poor performance in business, which includes business depression, excessive competition and extrinsic shocks
 - loose management, which includes failure of speculative investment, internal conflict and lack of efficient management
 - long-term accumulated deficit
 - insufficient working capital
 - accidental causes (disasters etc.)
 - deterioration of products in inventory
 - excessive investment in facilities and equipment
2. “link” effect
 - secondary effect from bankruptcy of customer, subsidiary or collateral companies and failure of business-related firms
 - failure of accounts receivable
3. others
 - refusal of credit by financial institutions
 - unclassified

A bankruptcy is caused by two or more combination of these classified effects; the most dominant cause is recorded in such a case. See Ref. 8) for more details.

In addition, they carefully investigated the causes of bankruptcies after the Kobe Earthquake in 1995 and the East Japan Earthquake in 2011. The solo failures and link effects may have been originated from business depression and other bankruptcies in the primary shock. A typical case of solo failures is an abrupt drop of sales due to an exogenous shock in a damaged region, when the firm depends crucially on the region for its sales. Typical link effects are what we already mentioned above. We assume that such cases are *neighboring bankruptcies* in secondary effects due to the primary shock.

The two credit research agencies had accumulated monthly numbers of bankruptcies after the Kobe Earthquake, and are accumulating after the East Japan Earthquake. The data are publicly available, which we shall examine in the following section.

§3. Omori law after large-scale shocks of production network

The Omori law states that the frequency of aftershocks per unit time, $n(t)$ at time t decays as

$$n(t) = K(t + c)^{-p}, \quad (3.1)$$

where K is a positive constant which determines the magnitude of $n(t)$ and c is a positive constant to avoid divergence at the origin, $t = 0$. $p > 0$ is the exponent of the power-law decay. The original proposal corresponds to the case $p = 1$, and was later modified into the above form, a modified Omori formula. This is a power-law decay

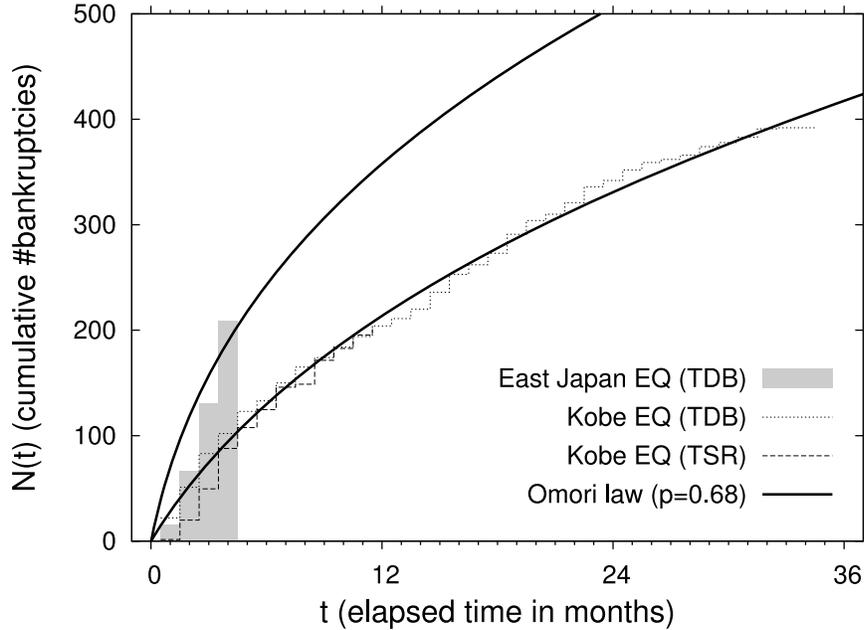


Fig. 1. Cumulative number of bankruptcies $N(t)$ and elapsed time t since the events of two earthquakes, Kobe (1995) and East Japan (2011). Dotted lines are two data $N(t)$ based on investigation of the Kobe case for three years by two independent agencies, TDB and TSR (the latter is scaled with its number for comparison). Gray bars are a few monthly data of TDB for the East Japan case. The two lines are modified Omori formula with the same parameter $p = 0.68$, which is estimated by least-square fit for the Kobe data.

with exponent p , so an extremely long relaxation in the aftershock earthquakes. The value of p ranges from 0.6 to 2.5 with median 1.1 in the survey of more than 200 observations during 33 years of earthquakes (see Ref. 2)).

Equivalently, but more suitable for comparison with data, the cumulative number $N(t)$ can be written by the following:

$$N(t) = \int_0^t dt' n(t') = \begin{cases} K ((t+c)^{1-p} - c^{1-p}) / (1-p) & \text{for } p \neq 1, \\ K \ln(t/c + 1) & \text{for } p = 1. \end{cases} \quad (3.2)$$

Because $N(t)$ is accumulation of $n(t)$, the statistical fluctuation and measurement errors in $N(t)$ are much reduced compared to them in $n(t)$. It is a custom to measure $N(t)$ to estimate the parameters of p and K, c .

Let us put the monthly data of neighboring bankruptcies described in §2 into the cumulative number $N(t)$, where t is the time of occurrence of the extreme shocks. The result is shown in Fig. 1.

For the Kobe Earthquake in 1995, the neighboring bankruptcies were recorded for three years by Teikoku Data Bank, Inc. (TDB),⁵⁾ for a year by Tokyo Shoko Research, Inc. (TSR).⁶⁾ We fit the data of TDB by (3.2) in a nonlinear least-square fitting. The parameter p was estimated as $p = 0.68 \pm 0.18$ with the standard error. Although the two data have different magnitude of $N(t)$ based on independent investigation, they simply give different scales of the parameter K if they obey the

same power-law. The TSR data is rescaled vertically in Fig. 1 which can be fitted well by the power-law with the same exponent p .

Similarly, although the data for the East Japan Earthquake has only a half-year elapsed time, it is observed from the fitting is satisfactory with respect to the same power-law (see gray bars for the actual data and a line for the Omori law). One can also see that the East Japan case has a much larger impact in terms of the overall scale in the number of failures. The parameters p and K are to be determined by new data available in coming months.

Note that while the TDB data was for three years, longer investigation might reveal that the effect is longer than the first three years. Actually, if one naively extrapolates the Omori law, there is a possibility that neighboring bankruptcies had been still going on beyond into the fourth or even longer years.

We have shown that the Omori law holds for the number of aftershocks measured by neighboring bankruptcies after the exogenous shocks of the two large earthquakes. Let us discuss the implication of this finding and also necessary verification to be done in our future work.

§4. Discussion

While our result in the preceding section is encouraging, there are several issues to be investigated. Let us discuss them in this section.

1. Other methods of measuring aftershocks, especially with “magnitudes”:
We used the number of neighboring bankruptcies on the production network in order to measure aftershocks, because a bankruptcy is the most clear evidence that a firm is in a so critically deteriorated state that it fails under a stress caused by the neighboring failures. In a more mild case, such a firm may be affected seriously in its financial state but yet does not go into bankruptcy. If one measures the magnitude of stress by an abrupt drop of sales, a sudden increase of debt and so forth, for example, one is able to quantify the extent of aftershocks including their magnitude. These methods will add the validity of our finding.
2. Other economic shocks:
By using the methods of measuring, the one in this paper and others described above, one will be able to test the Omori law in other large-scale destruction or depression of the economy rather than the exogenous and natural disasters. The Lehman shock and the present and past financial crises are good examples.
3. Implication of Omori law:
The fact that the relaxation process obeys a power-law implies that the influence of secondary effects or “Tsunami” after the primary damage in the network is extremely sluggish. Figure 1 indicates two things — duration of relaxation and its extent. Suppose that one can observe the first half-year, for instance, then by fitting the data by a modified Omori formula with a set of parameters, one can predict how long the relaxation will take and how large effects it will bring about under the hypothesis of power-law relaxation. What is more important than such a prediction is the possibility that one can identify industrial sectors

and geographical regions that are fragile or robust under such a propagation of failures. Such information would be useful to plan a recovery of and investment into sectors and regions.

4. Origin of Omori law:

The presence of a power-law relaxation seems to be ubiquitous in a variety of complex systems. The origin of Omori law even in the case of earthquake is not fully understood. Nevertheless, many proposed mechanisms are based on the idea that the entire region of aftershocks is composed of a number of small and heterogeneous regions containing faults, among which the stress of main earthquake is released subsequently with delayed shocks. This idea can be possibly applied to our case. Because the production network is not homogeneous but very heterogeneous with hierarchical tightly-knitted clusters or modules (see Ref. 9) for more details). Such a hierarchical and heterogeneous network structure quite likely provides an arena where a delayed release of stress propagation in terms of financial deterioration of firms. This approach of understanding the origin of Omori law seems promising, because this is common to other phenomena with power-law relaxation in natural and social systems. These open issues are to be investigated in a future work.

§5. Summary

We investigated “aftershocks” after a large-scale destruction of supplier and customer network. The aftershock is defined and measured by neighboring bankruptcies on the production network. The neighborhood is determined in two independent investigations of credit research agencies on exhaustive lists of bankruptcies by the criterion that the cause of bankruptcy is due to the primarily damaged firms or regional economy. Assuming that the relaxation process in secondary effects after the primary exogenous shock can be traced by such neighboring bankruptcies, we found that the modified Omori law holds for the decay in the number of neighboring bankruptcies in the country by using two cases of the Kobe Earthquake in 1995 and the East Japan Earthquake in 2011.

At this preliminary stage, our finding in this paper implies that the recovery from such a huge shock on production network is much more sluggish than one can naively expect. There are several open problems, which include other methods of quantifying aftershocks, consideration of magnitudes of aftershocks, study of economic disasters rather than natural ones, the origin of Omori law, and so forth. These issues are to be investigated further in future.

Acknowledgements

The author thanks Jiancang Zhuang at the Institute of Mathematical Statistics in Tokyo for references on Omori law in seismology. He would also like to thank the participants of the workshop “Econophysics 2011 — The Hitchhiker’s Guide to the Economy” (YITP-W-11-04) at the Yukawa Institute of Theoretical Physics, Kyoto University, for useful discussion and comments. This work was partially supported

in part by the Program for Promoting Methodological Innovation in Humanities and Social Sciences by Cross-Disciplinary Fusing of the Japan Society for the Promotion of Science.

References

- 1) F. Omori, "On Aftershocks of Earthquakes", Journal of the College of Science, Imperial University of Tokyo **7** (1894), 111.
- 2) T. Utsu, Y. Ogata and R. S. Matsu'ura, J. of Physics of the Earth **43** (1995), 1.
- 3) F. Lilo and R. N. Mantegna, Phys. Rev. E **68** (2003), 016119.
- 4) D. Sornette, F. Deschatres, T. Gilbert and Y. Ageon, Phys. Rev. Lett. **93** (2004), 228701.
- 5) <http://www.tdb.co.jp>
- 6) <http://www.tsr-net.co.jp>
- 7) <http://www.chusho.meti.go.jp>
- 8) Y. Fujiwara, Adv. in Complex Systems **11** (2008), 703.
- 9) Y. Fujiwara and H. Aoyama, Eur. Phys. J. B **77** (2010), 565.

Impact of the Great East Japan Earthquake on Hotel Industry in Pacific Tohoku Prefectures

— *From Spatio-Temporal Dependence of Hotel Availability* —

Aki-Hiro SATO

*Department of Applied Mathematics and Physics, Kyoto University,
Kyoto 606-8501, Japan*

This paper investigates the impact of the Great Japan Earthquake (and subsequent tsunami turmoil) on socio-economic activities by using data on hotel opportunities collected from an electronic hotel booking service. A method to estimate both primary and secondary regional effects of a natural disaster on human behavior is proposed. It is confirmed that temporal variation in the regional share of available hotels before and after a natural disaster may be an indicator to measure the socio-economic impact at each district.

§1. Introduction

Since people in the world are also products of nature, the physical effects of natural environment on our society are remarkable. Specifically, natural disasters often affect our societies significantly. Therefore, we need to understand the subsequent impact of natural disasters on human behavior from both economical and social perspectives.

The first Great East Japan Earthquake hit at 14:46 on 11 March 2011 in Japanese local time (05:46 in UTC). Within 20 minutes after shaking, huge tsunamis had devastated cities along Japan's northeastern coastline. In fact, destruction was largely physical, but social infrastructures have been damaged. Currently, it is considerably significant for us to understand its subsequent impact on our socio-economic activities.

According to Sigma 2/2011 from Swiss Re, significant insurance losses are expected from the 11 March Tohoku Earthquake (and subsequent tsunami), which resulted in the death and disappearance of more than 23,000 people.¹⁾ Economic activities before and after the Great East Japan disaster completely changed due to the physical destruction observed in many industrial sectors. The fiscal cost of this disaster to Japan has been conservatively estimated at \$200 billion. Actually, the short-term impact on Japanese growth is likely to be negative and potentially quite large. Nevertheless, reconstruction efforts are likely to get underway, which will provide a substantial boost to growth by the end of this year.

Monitoring regional human behavior may provide decision-makers with useful insights on the management of re-establishment after disasters. However, in general, it is not so easy to collect large-scale data on human activities before and after the natural disasters with both a high level of detail and a large number of samples. In this case, it is important to find an adequate proxy variable.

Tourism demand is particularly sensitive to security and health concerns. Esti-

mation of the economic impact of changes in the demand for tourism has typically been investigated with several methods. Blake et al. analyzed the effects of crisis using a computational general equilibrium model of the US and also examined potential and actual policy responses to the crisis.²⁾ Moreover, the problem of estimating demand from censored booking data has been recognized for many years in the hotel industry. Patrick et al. developed parametric regression models that consider not only the demand distribution, but also the conditions under which the data were collected.³⁾ Sato investigated regional patterns of Japanese travel behavior by using the EM algorithm for finite mixtures of Poisson distributions.⁵⁾

Furthermore, understanding what motivations influence people's travel habits and destination selections is crucial to predicting their future travel patterns. A review of the literature on tourist motivation reveals that a model that views motivations on the two dimensions of "push" and "pull" factors has been generally accepted.⁴⁾ The idea behind this two-dimensional approach is that people travel because they are pushed by their own internal forces and pulled by the external forces of the destination attributes.⁶⁾ The pull factors originate from the destination properties (supply). The push factors belong to consumers (demand). More recently, Tkaczynski et al. applied the stake-holder theory, a management theory proposed by Freeman (1984),⁸⁾ to a destination in tourism.⁷⁾ The existence of hotel accommodations implies that pull factors are present in the district where they are located.

Therefore, we may assume that changes of demand and supply in the hotel industry can reflect both the social and economic impact of natural disasters on human behavior. In this article, the use of data on hotel availability collected from an electronic hotel booking service is suggested for this purpose. Hotel opportunities are handled in real-time and the geographical coverage of hotel locations seems high. Spatio-temporal dependence of hotel availability can be analyzed from physical point of view.

According to a recent report by Japan Tourism Agency in the Ministry of Land, Infrastructure, Transport and Tourism, the total number of accommodations, including hotels, Japanese inns, and pensions all over Japan, is estimated as 53,468 during 2010.⁹⁾ Specifically, in the prefectures that have been damaged by the earthquake and tsunami (Iwate, Miyagi, and Fukushima), it is estimated that there are 3,846 accommodation facilities. Notably, hotels in this area are so densely located that we can use the data to measure the socio-economic impact. Based upon this idea, the number of available hotels before and after the Great East Japan Earthquake is examined in this study. Hotel availability is assumed to be a proxy variable of human mobility, so comparative analysis before and after the disaster is conducted.

§2. Data and methods

We used data collected from a Japanese hotel booking site named Jalan via a Web API (Application Programming Interface).¹³⁾ The Jalan is one of the most popular hotel reservation services in Japan. The Web API is an interface code set that is designed to simplify development of application programs.

Table I. The ratio of the number of available hotels during the period from 1 to 31 May 2011 to that during the period from 1 to 31 May 2010 (after and before the Great East Japan Earthquake), the number of both completely destroyed houses and partially destroyed houses, as confirmed at the end of September 2011 and the number of evacuees, as confirmed at 1 May 2011.

prefecture	district	ratio	complete collapse	partial collapse	evacuees
Iwate	Shizukuishi	1.970	0	0	372
	Morioka	1.834	0	4	366
	Appi, Hachimantai, Ninohe	2.250	3	0	0
	Hanamaki, Kitakami, Tohno	1.350	27	364	853
	Sanriku Kaigan	0.481	18,098	2,166	12,896
	Oushu, Hiraizumi, Ichinoseki	0.374	83	533	338
Miyagi	Sendai	0.550	21,789	37,522	3,608
	Matsushima, Shiogama	0.345	7,895	12,581	5,115
	Ishinomaki, Kesenuma	0.0	33,661	6,083	23,840
	Naruko, Osaki	1.484	486	1,577	929
	Kurihara, Tome	1.404	224	1,105	1,049
	Shiroishi, Zao	1.608	2,522	1,644	1,612
Fukushima	Fukushima, Nihonmatsu	0.665	168	1,898	1,321
	Soma	0.038	6,279	1,618	1,969
	Urabandai, Bandai Kogen	1.134	0	0	2
	Inawashiro, Omotebandai	1.009	10	12	303
	Aizu	1.352	4	27	266
	Minamiaizu	1.768	0	0	14
	Koriyama	0.604	2,596	12,185	2,489
	Shirakawa	1.915	135	1,820	418
	Iwaki, Futaba	0.195	6,550	17,614	2,115

In order to estimate both economic and social damages in three Tohoku prefectures (Iwate, Miyagi, and Fukushima), we focus on the number of available hotels in each district before and after the Great East Japan Earthquakes and Tsunami. We selected 21 districts in three prefectures, as shown in Table I and two periods, one before and one after the disaster.

The data on hotels in this area cover about 31% of the potential hotels. Therefore, we have to estimate the uncensored states from these censored booking data. If we assume that the hotels in the data are sampled from uncensored data in a homogeneous way, then the relative frequency of the available accommodations from censored data can approximate the true value, computed from uncensored data.

In order to conduct a quantitative study, let $x_i(t, s)$ ($i = 1, \dots, K; t = 1, \dots, T$) be the number of available hotels in district i at day t in period s . Then the relative frequency at district i can be calculated as

$$p_i(s) = \frac{\sum_{t=1}^T x_i(t, s)}{\sum_{i=1}^K \sum_{t=1}^T x_i(t, s)}. \quad (2.1)$$

Let us consider a ratio of the relative frequencies after and before a specific event

$$q_i(a; b) = p_i(a)/p_i(b), \quad (2.2)$$

where $p_i(a)$ and $p_i(b)$ represent the relative frequencies after and before the event,

respectively. Obviously, Eq. (2.2) can be rewritten as

$$q_i(a; b) = \frac{n_i(a)}{n_i(b)} / \frac{N(a)}{N(b)}, \quad (2.3)$$

by using $n_i(s)$ and $N(s)$, which are the number of available hotels in district i within the period s and the total number at that moment, respectively. Since $N(a)/N(b)$ is independent of i , $q_i(a; b)$ should be proportional to the ratio of the number of hotels after and before the event.

§3. Analysis

The 3rd column in Table I shows $q_i(a; b)$, where the term b represents May 2010 (before the disaster), and the term a May 2011 (after the disaster), respectively. Since the value of $q_i(a; b)$ is related to damages to hotels in district i , $q_i(a; b) < 1$ implies that available hotels decreased after the earthquake at i , compared to the ratio of the total number of hotels. Similarly $q_i(a; b) > 1$ means that they increased or maintained at i .

From this it is estimated that Sanriku Kaigan, Oshu, Hiraizumi, Ichinoseki, Sendai, Matsushima, Shiogama, Fukushima, Nihonmatsu, Soma, Koriyama, Iwaki and Futaba were significantly damaged by the earthquake and tsunami. We may assume that the decrease of $q_i(a; b)$ at district i results from both a decrease in supply and an increase in demand. The decrease in supply is caused by the physical destruction of infrastructure in this case. The increase in demand comes from behavior of individuals. The regional dependence of supply can be estimated from the number of destroyed houses in each district.

To do so, we calculated the numbers of both completely destroyed and partially destroyed houses in each district from the data downloaded from a Web page of National Research Institute for Earth Science and Disaster Prevention.¹⁴⁾ The numbers were calculated by summing the number of destroyed houses in a town or city included in each district. Table I shows the numbers of destroyed houses. From this histogram, it can be seen that house damages were concentrated in the maritime area of these prefectures.

We can confirm that the damages to houses were serious in Sanriku Kaigan, Sendai, Matsushima, Shiogama, Ishinomaki, Kesenuma, Soma, Koriyama, Iwaki and Futaba. The highest number of completely destroyed houses is in Ishinomaki and Kesenuma, with 33,661 homes destroyed. The second-highest is 21,789, in Sendai. The third-highest is 18,098, in Sanriku Kaigan. The highest number of partially destroyed houses is 37,522, in Sendai. The second-highest is 17,614, in Iwaki and Futaba. The third-highest is 12,185, in Koriyama.

From an official report by each Prefecture^{10)–12)} the number of evacuees who have stayed at public refuge holes is counted at each district. The 6th column in Table I shows the number of evacuees as confirmed at 1 May 2011. The number of evacuees in each district is proportional to the number of completely destroyed houses. It is found that physical damages were not serious from both the number of destructed houses and evacuees in Oshu, Hiraizumi, Ichinoseki and Fukushima, Nihonmatsu.

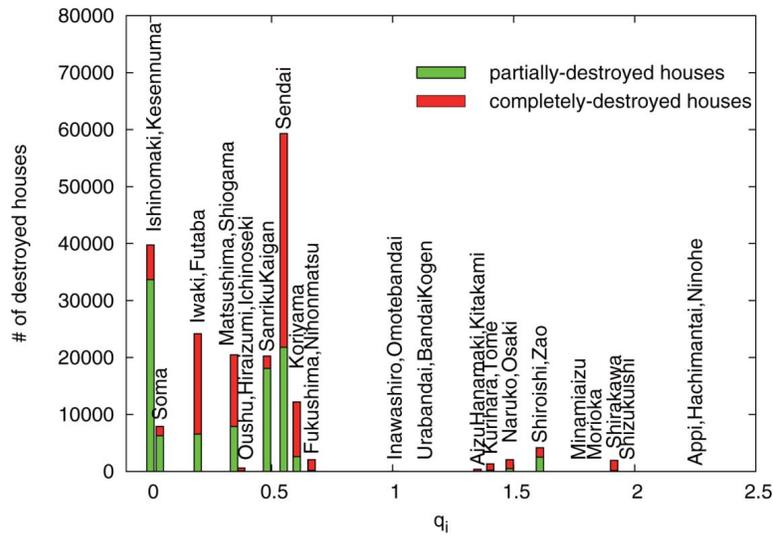


Fig. 1. Relationship between the ratio $q_i(a; b)$ before and after the Great East Japan Earthquake, and the number of destroyed houses, as confirmed at the end of September 2011.

The degree of damages which is measured by physical damages is consistent with a value of q_i computed by the proposed method except these districts.

Figure 1 shows correlation among the number of destroyed houses and q_i . In fact, where the ratio $q_i(a; b)$ is greater than 1, the number of destroyed houses is not significant. We confirmed that the ratio $q_i(a; b)$ may measure the degree of damage to economic activities through the travel industry. However, it was not confirmed that the significant physical damages to houses in Oushu, Hiraizumi, Ichinoseki, Fukushima, and Nihonmatsu even have a ratio less than 1. It may be thought that hotels in Oushu, Hiraizumi, and Ichinoseki were used by works and evacuated victims of the disaster. The decrease in available hotels in Fukushima and Nihonmatsu may be related to the accidents at the Fukushima Daiichi nuclear power plant.

The number of evacuated victims of the disaster in each prefecture according to an official announcement by Japanese Cabinet Office on 3 June 2011 is shown in Table II. In the case of the Fukushima prefecture, 17,874 evacuated people were in hotels at that moment. Although we could not collect correct data on hotel booking

Table II. The number of evacuees of the Great East Japan Earthquake at three prefectures (Iwate, Miyagi, and Fukushima). The data were officially announced by the Japanese Cabinet Office on 3 June 2011.

prefecture	A: public places	B: hotels	C: others	A+B+C
Aomori	0	78	777	855
Iwate	9,039	2,007	14,701	25,747
Miyagi	23,454	2,035	-	25,489
Akita	128	619	909	1,656
Yamagata	305	779	2,366	3,450
Fukushima	6,105	17,874	-	23,979

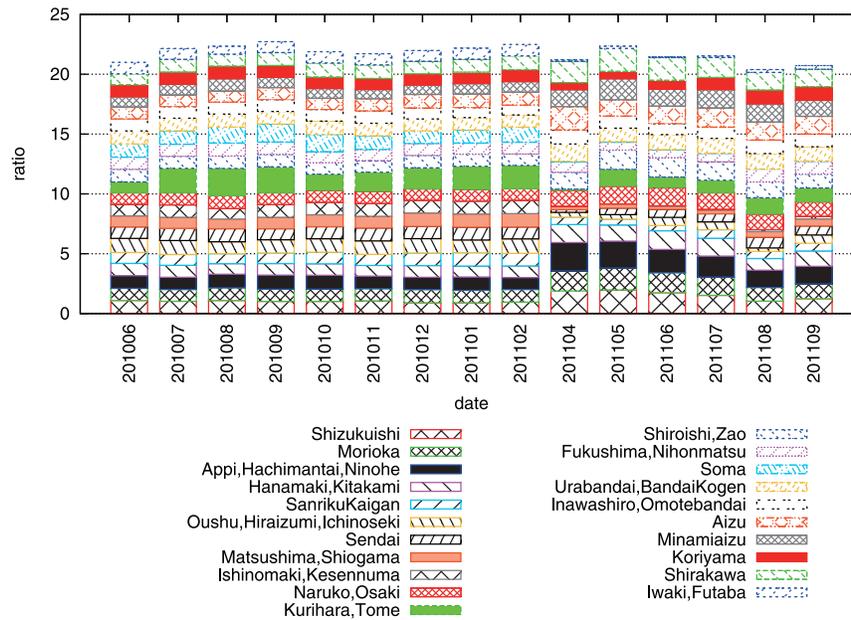


Fig. 2. The ratios of the relative frequencies of available hotels from May 2010 to September 2011. The base period is fixed as May 2010.

on these two districts, we may assume that hotels in these districts were occupied by individuals (refugees, workers, volunteers and civil groups).

Consequently, we concluded that the physical damage to buildings constructions is estimated to be small if the ratio is larger than 1. Moreover, it is found that regional distribution of evacuated victims and workers may influence the ratio. This represents the secondary effects of natural disasters on human behavior.

§4. Transient behavior

In this section, we investigate temporal development of hotel availability before and after the Great East Japan Earthquake. By using Eqs. (2.2) and (2.3), we compute $q_i(a; b)$ on the basis of the regional share of available hotels in May 2010. Figure 2 shows $q_i(a; b)$, where b represents May 2010, and a the period from May 2010 to September 2011. It is found that the ratios are stable and take values around 1 at all the districts except Kurihara and Tome until February 2011.

However, after the earthquake (from April 2011) the ratios drastically changed. Figure 3 shows the ratios of several districts. The ratio in April 2011 reflects the degree of physical damage just after the earthquake, as shown in §3. Therefore, the temporal variation can be assumed to show the degree of both physical and social damage.

In Fig. 3, we see that 22.6% of the hotels in Sendai had recovered by April 2011, 55% by May 2011, 66.9% by June 2011, 67.6% by July, 88.8% by August 2011 and 75.8% by September 2011, in comparison with May 2010. Hotels along the

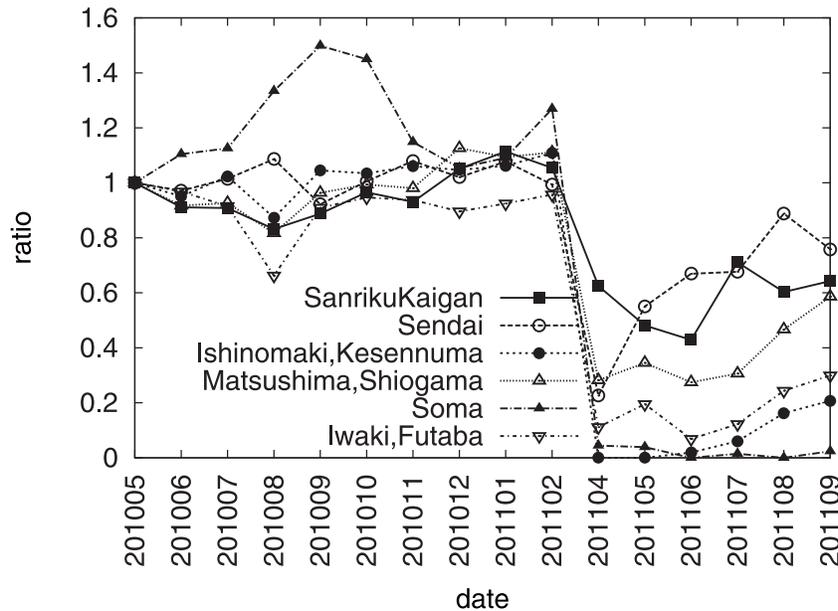


Fig. 3. The ratios of the relative frequencies of available hotels from May 2010 to September 2011 at several districts (Sanriku Kaigan, Sendai, Ishinomaki, Kesennuma, Matsushima, Shiogama, Soma, Iwaki and Futaba). The base period is fixed as May 2010.

coastlines (Sanriku Kaigan, Matsushima, Shiogama, Ishinomaki, Kesennuma and Iwaki) are eventually recovering, though more slowly than at Sendai. In the case of Ishinomaki and Kesennuma, the hotels completely disappeared in April 2011, but 20.1% of the hotels had recovered by September 2011. In contrast to these areas, hotels at Soma had not recovered by the time this text was written, in September 2011.

§5. Conclusion

A method to measure changes in human activities before and after a natural disaster from both social and economic perspectives with data collected from an electronic booking site was proposed. Regional shares of the number of available hotels were introduced and a ratio of the share after the disaster to that before it was defined. Hotel availability before and after the Great East Japan Earthquake and Tsunami was analyzed through a comparison of the number of destroyed houses with the ratio. As a result, we confirmed that the physical damage to buildings constructions is estimated to be small if the ratio is larger than 1. Moreover, it is found that regional distribution of evacuated victims and workers may influence the ratio. This represents the secondary effects of natural disasters on human behavior.

Acknowledgements

The author thanks the Yukawa Institute for Theoretical Physics at Kyoto University. Discussions during the YITP workshop YITP-W-11-04 on “Econophysics 2011: The Hitchhiker’s Guide to the Economy” were useful in completing this work. The author is thankful to Prof. Hideaki Aoyama and Prof. Zdzisław Burda for valuable suggestions. The author also thanks Mr. Sigehiro Kato, Mr. Hiroshi Yoshimura, Mr. Kotaro Sasaki and Ms. Yoko Miura of Recruit Cooperation for their stimulating discussions.

References

- 1) Sigma 2/2011, Swiss Re. <http://www.swissre.com/publications/>
- 2) A. Blake and M. T. Sinclair, *Annals of Tourism Research* **30** (2003), 813.
- 3) P. H. Liu, S. Smith, E. B. Orkin and G. Carey, *J. of Revenue & Pricing Management* **1** (2002), 121.
- 4) S. Cha, K. W. Mcclary and M. Uysal, *J. of Travel Research* **34** (1995), 33.
- 5) A.-H. Sato, *International Review of Financial Analysis* (2011), in press.
- 6) G. M. S. Dann, *Annals of Tourism Research* **4** (1977), 184.
- 7) A. Tkaczynski, S. Rundle-Thiele and N. Beaumont, *J. of Travel Research* **49** (2010), 139.
- 8) R. E. Freeman, *Strategic Management: A Stakeholder Approach* (Harpercollins College Div, Boston/NY, 1984).
- 9) The data is downloaded from a Web page of Japan Tourism Agency in Ministry of Land, Infrastructure, Transport and Tourism:
<http://www.mlit.go.jp/kankocho/siryoutoukei/index.html> (31 Aug 2011).
- 10) The data is downloaded from a Web page of Iwate Prefecture:
<http://www.pref.iwate.jp/~bousai/taioujoukyou/201105011700hinanbasyo.pdf>
- 11) The data is downloaded from a Web page of Miyagi Prefecture:
<http://www.pref.miyagi.jp/kikitaisaku/higasinihondaisinsai/pdf/5011900.pdf>
- 12) The data is downloaded from a Web page of Fukushima Prefecture:
<http://www.pref.fukushima.jp/j/hinanjolist0501.pdf>
- 13) The data is collected from Jalan Web Service:
<http://www.jalan.net/>
- 14) The data is downloaded from a Web page of National Research Institute for Earth Science and Disaster Prevention:
<http://www.j-risq.bosai.go.jp/ndis/> (31 Aug 2011).

An Analysis of Japanese Hotel Plan Availability

— *Regionality and Seasonality* —

Minoru NODA^{*)} and Aki-Hiro SATO^{**)}

*Department of Applied Mathematics and Physics, Graduate School of Informatics,
Kyoto University, Kyoto 606-8501, Japan*

This paper investigates regionality and seasonality of room plan availability in Japan. We employ the singular spectrum analysis to the daily plan availability and extract the trends. We examine the coefficient of a linear relation between a normalized trend at each prefecture and a normalized trend of the whole of Japan. It is concluded that the coefficients, with respect to the four seasons, characterize properties of the prefectures.

§1. Introduction

Many people visit Kyoto in the spring and autumn, but, in the same seasons there are few people in the summer retreats. Thus, the preference of travelers depends on both seasons and districts. How does the preference of travelers visiting each region change in the whole of Japan? According to Haag and Weidlich,¹⁾ migration dynamics are modeled on personal preference and regional characteristics. Moreover, according to the study of tourism management,²⁾ there are push and pull factors, so that tourism motivation is determined by the situation of the travelers (push) and the situation of the destination (pull). In the context of economics, this means that demand and supply are dependent on the area and the season.³⁾

How do demand and supply fluctuate according to the area and the season? To know this, the data about the situation of travelers' migration in Japan must be studied. In fact, there is survey data provided by the Japan Tourism Agency, but it is not sufficient to capture the motivation structure of travelers with a high resolution.

More precise analysis needs more frequency data. To solve this problem, we constructed a system that automatically downloaded hotel data from a Japanese hotel booking site every day. In this paper, we found that the preference of travelers depends on both seasons and districts by using the high frequency data that we collected.

This article is organized as follows. The data of room opportunities is presented in §2. In §3, we briefly explain the singular spectrum analysis, which can extract the trend from the time series data. In §4, we explain a method to analyze the dependence of a trend at each prefecture based on the trend of Japan as a whole and show the results. Finally, §5 makes concluding remarks.

^{*)} E-mail: minoru.noda.527@gmail.com

^{**)} E-mail: sato.akihiro.5m@kyoto-u.ac.jp

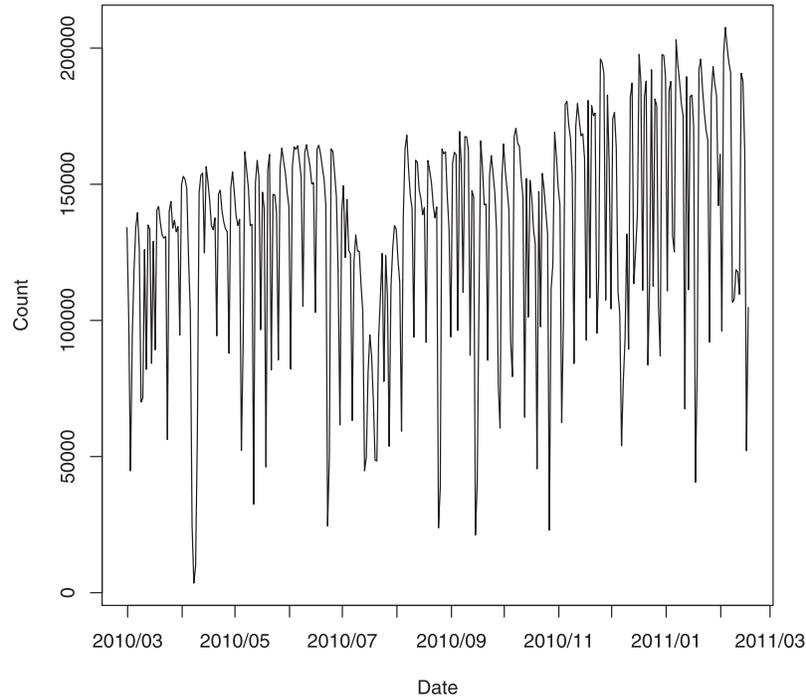


Fig. 1. The daily number of available plans in Japan during the period of 25 March, 2010 to 13 March, 2011.

§2. Data description

In this section, we give a brief explanation of our data. We used data collected from a Japanese hotel booking site named Jalan under the condition that two adults would be able to stay in the room per night. On the Jalan web site, there exists over 100,000 room possibilities from over 14,000 hotels. The plan availability is defined by the number of plans that consumers can reserve from the web site. Figure 1 shows the time series of the daily number of available plans during the period from 25 March, 2010 to 13 March, 2011. This time series consists of 354 data points. Features of the time series are that the number of available plans decreases on Saturdays. This means that periodicity exists in this time series. Moreover, during Golden and New Year holidays, when the long vacation comes, the number of booked plans increases steeply.

§3. Singular spectral analysis

In this section, we explain the method to extract a trend from the original series by using the singular spectrum analysis.⁴⁾ The main purpose of the singular spectrum analysis is to decompose the original series into a sum of orthogonal series, so that each orthogonal component in the sum can be identified as either a trend, periodicity or quasi-periodicity (perhaps, amplitude-modulated), or noise. This method

generates a reconstruction of the original series.

The main idea of the singular spectrum analysis is as follows. Consider the real-valued time series $Y_T = (y_1, \dots, y_T)$ of sufficient length T . Assume that Y_T is a nonzero series. The singular spectrum analysis consists of two complementary stages: decomposition and reconstruction.

3.1. Decomposition

The embedding procedure maps the original time series to a sequence of multi-dimensional lagged vectors. Let K be $T - L + 1$, where L ($1 < L < T$) is some integer called the window length. The embedding procedure forms K lagged vectors:

$$X_i = (y_i, y_{i+1}, \dots, y_{i+L-1})^T, \quad 1 \leq i \leq K,$$

which have L elements.

The trajectory matrix of the series Y_T is defined as

$$\mathbf{X} = [X_1, X_2, \dots, X_K],$$

which has lagged vectors as its columns. \mathbf{X} is an $L \times K$ matrix. In other words, the trajectory matrix is described as

$$\mathbf{X} = (x_{ij})^{L,K} = \begin{pmatrix} y_1 & y_2 & y_3 & \dots & y_K \\ y_2 & y_3 & y_4 & \dots & y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \dots & y_T \end{pmatrix}. \quad (i = 1, \dots, L; j = 1, \dots, K)$$

Obviously, $x_{ij} = y_{i+j-1}$ and the element x_{ij} is equal on $i + j = \text{const}$. Thus, the trajectory matrix is a *Hankel matrix*.

The next step is the singular value decomposition of the trajectory matrix. Define the matrix $\mathbf{S} = \mathbf{X}\mathbf{X}^T$. The singular spectrum analysis of \mathbf{S} provides the eigenvalues of \mathbf{S} ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L \geq 0$) and the eigenvectors \mathbf{S} (U_1, U_2, \dots, U_L). Let $d = \max\{i, \text{such that } \lambda_i > 0\}$.

If we denote $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$, ($i = 1, \dots, d$), then the singular value decomposition of \mathbf{X} can be written as

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d, \tag{3.1}$$

where $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i$. We call the collection $(\sqrt{\lambda_i}, U_i, V_i)$ i -th eigentriple.

3.2. Reconstruction

The reconstruction step in the singular spectrum analysis transforms each decomposition matrix shown in Eq. (3.1) into a new series with the length T . Let $\mathbf{X}_1 = \sqrt{\lambda_1} U_1 V_1$ be an $L \times K$ ($L \leq K$) matrix with elements \hat{y}_{ij} ($1 \leq i \leq L, 1 \leq j \leq K$). To represent the diagonal averaging step, we employ the *Hankelization* operator \mathcal{H} . The operator \mathcal{H} acts on an arbitrary $(L \times K)$ -matrix \mathbf{X}_1 in the following way:

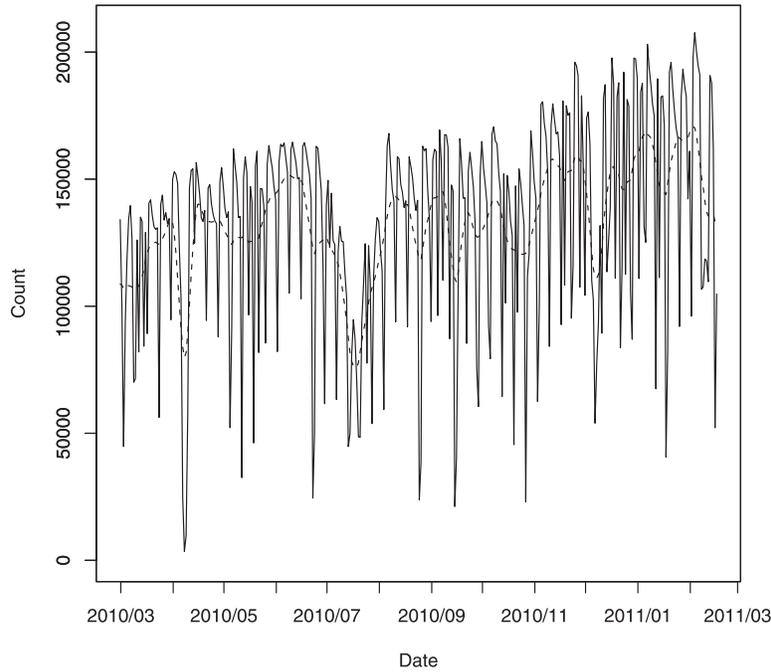


Fig. 2. A solid curve represents the daily number of available plans in the whole Japan during the period from 25 March, 2010 to 13 March, 2011. A dashed curve represents the trend extracted by singular spectrum analysis under the condition the first eigentriple and the window length $L = 7$.

for $i + j = s$ and $T = L + K - 1$, the element \tilde{y}_{ij} of the matrix $\mathcal{H}\mathbf{X}_1$ is

$$\tilde{y}_{ij} = \begin{cases} \frac{1}{s-1} \sum_{l=1}^{s-1} \hat{y}_{l,s-l} & \text{for } 2 \leq s \leq L, \\ \frac{1}{L} \sum_{l=1}^L \hat{y}_{l,s-l} & \text{for } L+1 \leq s \leq K+1, \\ \frac{1}{K+L-s+1} \sum_{l=s-K}^L \hat{y}_{l,s-l} & \text{for } K+2 \leq s \leq K+L. \end{cases}$$

Then, we obtain the reconstructed time series as

$$\tilde{Y}_T = (\tilde{y}_{1,1}, \tilde{y}_{2,1}, \dots, \tilde{y}_{L,1}, \tilde{y}_{L,2}, \tilde{y}_{L,3}, \dots, \tilde{y}_{L,K}).$$

§4. Method and analysis

Let $G(t)$ be the total number of available plans on the date t in the whole Japan. $F(t)$ is denoted as the trend extracted from the first eigentriple of $G(t)$ and the window length $L = 7$. According to Hassani et al., due to smoothing, correlation coefficients have increased.⁵⁾ Figure 2 shows the original series $G(t)$ for the whole of Japan and the extracted trend $F(t)$.

Let us divide the region of Japan into M prefectures. We define $g_k(t)$ as the number of available plans for the prefecture k on the date t . Because there are 47 prefectures in Japan, M is equal to 47. We extracted the trend $f_k(t)$ from $g_k(t)$ by using the singular spectrum analysis. We normalize all data series in order to deal with them in the same scale:⁵⁾

$$\frac{f_k(t)}{\|f_k\|}, \quad \frac{F(t)}{\|F\|}, \quad (k = 1, \dots, M; t = 1, \dots, T)$$

where

$$\begin{aligned} \|f_k\|^2 &= \frac{1}{T} \sum_{t=1}^T f_k(t)^2, \\ \|F\|^2 &= \frac{1}{T} \sum_{t=1}^T F(t)^2. \end{aligned}$$

Here, we consider how plan availability $f_k(t)/\|f_k\|$ is explained by $F(t)/\|F\|$. Suppose a linear relation between $f_k(t)/\|f_k\|$ and $F(t)/\|F\|$:

$$\frac{f_k(t)}{\|f_k\|} = a_k \frac{F(t)}{\|F\|} + b_k, \quad (k = 1, \dots, M) \tag{4.1}$$

where a_k and b_k represent real number coefficients on the prefecture k , respectively. We define squares error E of Eq. (4.1) as

$$E = \sum_{t=1}^T \sum_{k=1}^M \left(\frac{f_k(t)}{\|f_k\|} - a_k \frac{F(t)}{\|F\|} - b_k \right)^2,$$

and set its partial differential in terms of a_k and b_k into zero. Then, we have

$$a_k = \frac{\text{Cov} \left[\frac{f_k}{\|f_k\|}, \frac{F}{\|F\|} \right]}{\text{Var} \left[\frac{F}{\|F\|} \right]}, \tag{4.2}$$

$$b_k = \frac{\langle f_k \rangle}{\|f_k\|} - a_k \frac{\langle F \rangle}{\|F\|}, \tag{4.3}$$

where we use a notation $\langle x \rangle$ with T observations $x(t)$ ($t = 1, \dots, T$), defined as $\langle x \rangle = T^{-1} \sum_{t=1}^T x(t)$. Note that

$$\frac{\langle f_k(t) \rangle}{\|f_k\|} \simeq 1, \quad \frac{\langle F(t) \rangle}{\|F\|} \simeq 1, \tag{4.4}$$

from the empirical analysis of the trends, as shown in Fig. 3. From Eqs. (4.3) and (4.4), the following relation between a_k and b_k exists:

$$b_k \simeq 1 - a_k. \tag{4.5}$$

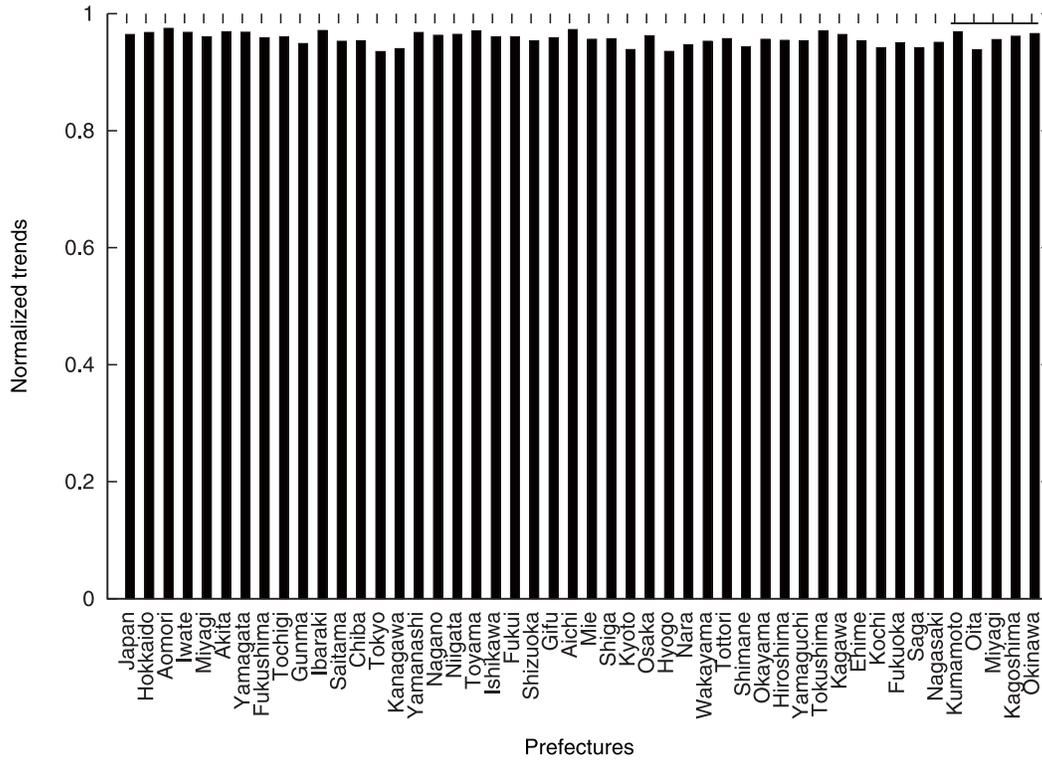


Fig. 3. Values of the normalized trend at each prefecture and of the normalized general trend in Japan. The values computed from the trends of the winter season. The x -axis represents the prefectures, and the y -axis their values.

In order to examine the dependence of the prefecture on seasons, we divided the time series $G(t)$ and $g_k(t)$ into four seasons: spring (from 1 March, 2010 to 31 May, 2010), summer (from 1 June, 2010 to 31 August, 2010), autumn (from 1 September, 2010 to 30 November, 2010), and winter (from 1 December, 2010 to 28 February, 2011). After this operation, we extracted the trends from the time series of each season, and then found (a_k^s, b_k^s) ; $s = 1$ (spring), 2 (summer), 3 (autumn), and 4 (winter).

As shown in Fig. 4, it is found that the relationship between a_k^s and b_k^s follows Eq. (4.5) for every season. The larger a_k^s places on the more right-hand side, and the smaller a_k^s places on the more left-hand side. The points indicating Kyoto lie on the right-hand side of the plot area in the spring, left-hand in the summer, right-hand in the autumn, and left-hand in the winter. Because the values of (a_k^s, b_k^s) indicate the dependence of the trend $f_k(t)$ at the prefecture k on $F(t)$, in the spring and winter of 2010 to 2011, Kyoto follows the general Japanese trend of vacationing in the spring and autumn, but not for summer and winter.

Figure 5 shows the relations between $F(t)/\|F\|$ and $f_k(t)/\|f_k\|$ for the four seasons in Kyoto. From Figs. 5 (a) and (c) (indicating the spring and autumn), it is found that a_k is small and b_k is large. Figures 5 (b) and (d) (indicating summer and winter) show that a_k is large and b_k is small. We may conclude that the prefecture

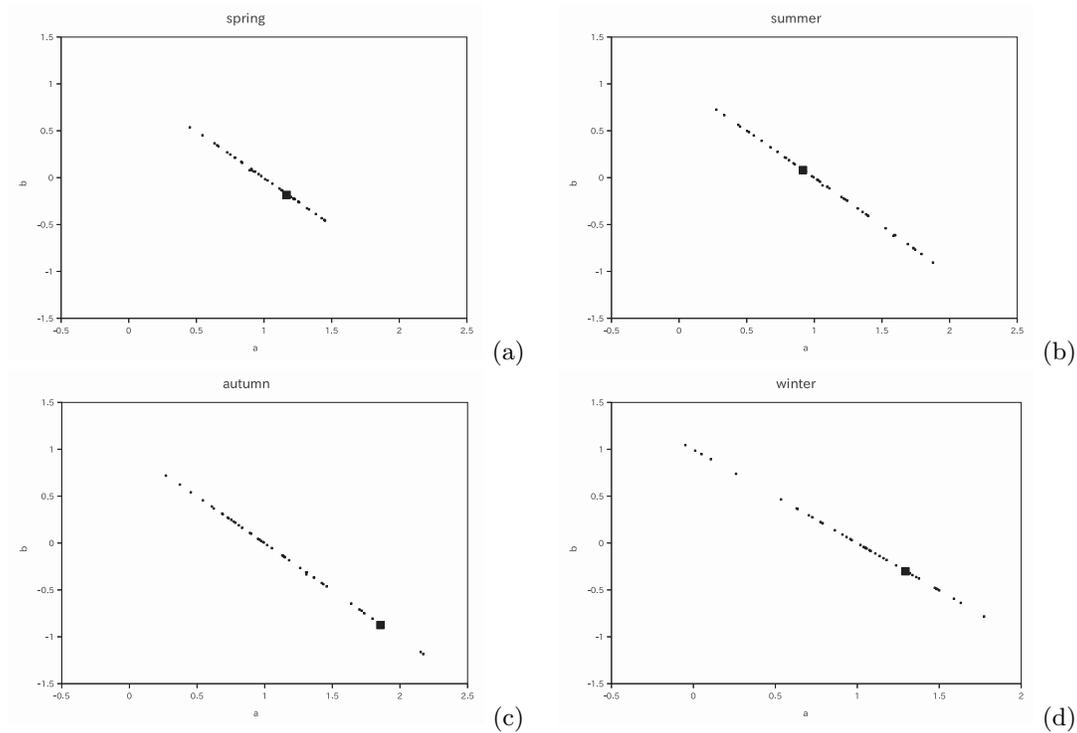


Fig. 4. Relationship of (a_k, b_k) in the four seasons; (a) spring, (b) summer, (c) autumn, and (d) winter. An unfilled circle represents the relation of each prefecture. A filled square represents the relation of Kyoto.

k with larger a_k^s is popular in the season s .

§5. Conclusions

We analyzed data on a Japanese hotel reservation site and showed that the preference of travelers depends on both seasons and districts. We quantified the degree of the dependence using a trend series extracted by the singular spectrum analysis. We examined seasonal dependence of each prefecture, and found that there are different preferred prefectures for each season.

This large-scale set of data on hotel opportunities provides us with invisible properties on human activities related to travel and tourism in Japan.

Acknowledgements

The authors thank the Yukawa Institute for Theoretical Physics at Kyoto University. Discussions during the YITP workshop YITP-W-11-04 on “Econophysics 2011” were useful to complete this work. Furthermore, we are thankful to the Kyoto University Global COE Program: Informatics Education and Research Center for Knowledge-Circulating Society for financial support.

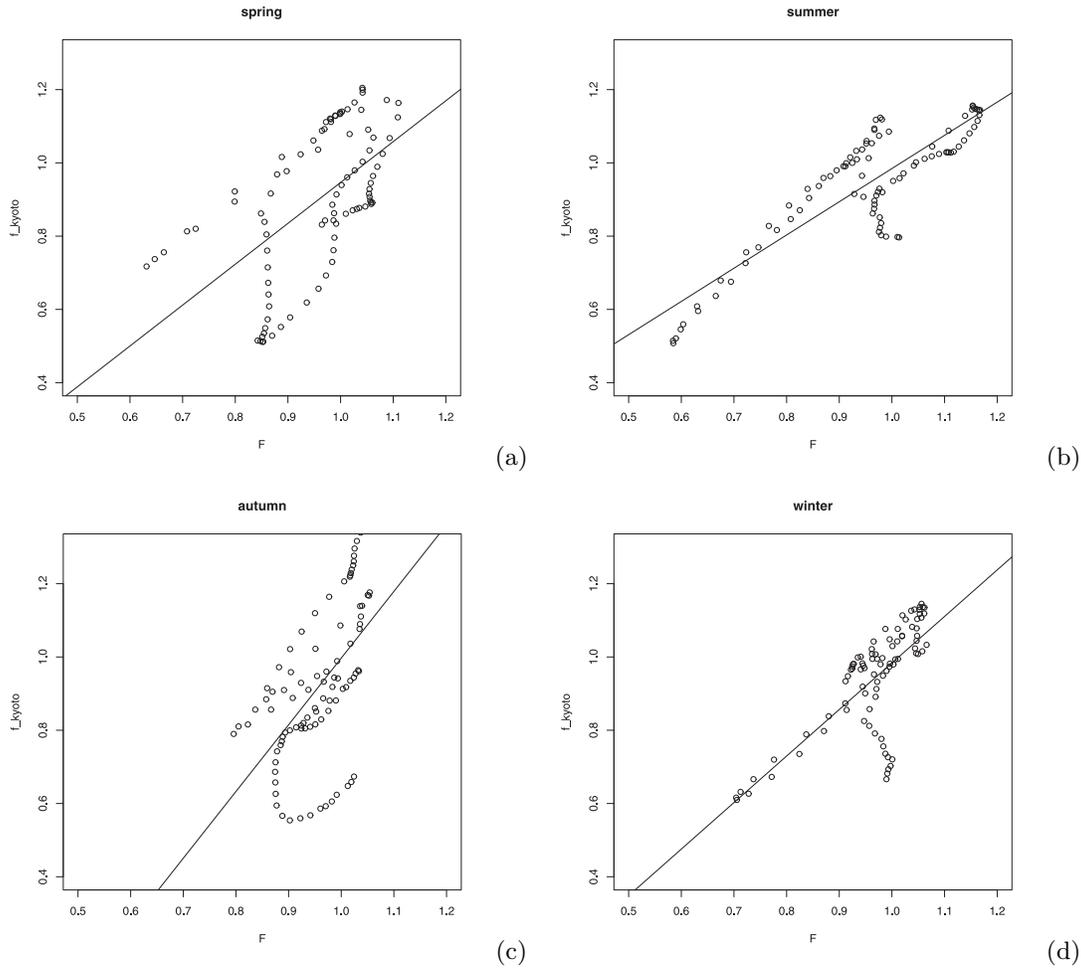


Fig. 5. The relations of $\left(\frac{F(t)}{\|F\|}, \frac{f_k(t)}{\|f_k\|}\right)$ in terms of t in Kyoto for the four seasons; (a) spring, (b) summer, (c) autumn, and (d) winter. Each line represents a linear regression between $\frac{F(t)}{\|F\|}$ and $\frac{f_k(t)}{\|f_k\|}$ computed from Eqs. (4.2) and (4.3).

References

- 1) G. Haag and W. Weidlich, *Regional Science and Urban Economics* **16** (1986), 57.
- 2) S. Cha, K. W. Mcclary and M. Uysal, *J. of Travel Research* **34** (1995), 33.
- 3) T. Cuccia and I. Rizzo, *Tourism Management* **32** (2011), 589.
- 4) N. Golyandina, V. Nekrutkin and A. A. Zhigljavsky, *Analysis of Time Series Structure* (Chapman & Hall/CRC, USA, 2001).
- 5) H. Hassani, A. S. Soofi and A. A. Zhigljavsky, *Nonlinear Analysis: Real World Applications* **11** (2010), 2023.

Temporal and Cross Correlations in Business News

Takayuki MIZUNO,^{1,2} Kazumasa TAKEI,³ Takaaki OHNISHI^{2,4}
and Tsutomu WATANABE^{2,4}

¹*Division of Information Engineering, Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba 305-8573, Japan*

²*Canon Institute for Global Studies, Tokyo 100-6511, Japan*

³*The Norinchukin Bank, Tokyo 100-8420, Japan*

⁴*Graduate School of Economics, The University of Tokyo, Tokyo 113-0033, Japan*

We empirically investigate temporal and cross correlations in the frequency of news reports on companies, using a dataset of more than 100 million news articles reported in English by around 500 press agencies worldwide for the period 2003–2009. Our first finding is that the frequency of news reports on a company does not follow a Poisson process, but instead exhibits long memory with a positive autocorrelation for longer than one year. The second finding is that there exist significant correlations in the frequency of news across companies. Specifically, on a daily time scale or longer the frequency of news is governed by external dynamics, while on a time scale of minutes it is governed by internal dynamics. These two findings indicate that the frequency of news reports on companies has statistical properties similar to trading volume or price volatility in stock markets, suggesting that the flow of information through company news plays an important role in price dynamics in stock markets.

§1. Introduction

News is the communication of information about important events. In macroeconomics, quantitative finance, and econophysics, the impact of news on prices and trading volumes in stock markets has previously been studied.^{1),2)} Some financial economists have shown that there is only a weak relationship between the number of news reports each day, the trading volume, and the price return in stock markets.³⁾ On the other hand, in the area of econophysics, it has been found by using tick-by-tick data that market volatility and volume increases immediately after particular news has been reported.^{4)–6)} The influence of exogenous shocks, including news reports, on pricing in financial markets has been examined using numerical models.⁷⁾ Another strand of research has attempted to detect patterns in the flow of information. For instance, it has been suggested that the frequency of use of specific words in blogs on the internet does not follow a Poisson process,^{8),9)} while Ref. 10) shows that using latent Dirichlet allocation, news articles appearing in the New York Times can be classified into several topics.

The aim of this paper is to empirically identify certain statistical properties of the frequency of news, with a special focus on the temporal correlation of news frequency for a specific company as well as the cross correlation of news across companies. For this purpose, we use a dataset of news articles reported by around 500 press agencies worldwide. The dataset – “Reuters NewsScope Archive” – was obtained from Thomson Reuters Corporation. The rest of the paper is organized

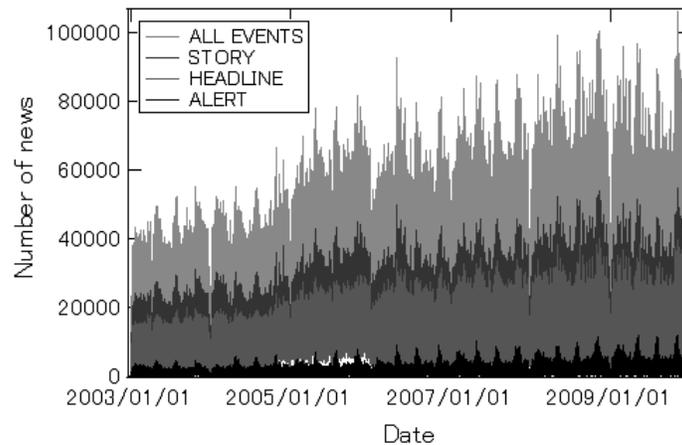


Fig. 1. Time series of the daily number of news reports. The number of news reports in English is counted. The top line is for all news; the second line is for story news; the third line is for headline news; and the bottom line is for alert news.

as follows. Section 2 describes our dataset and shows that there are periodicities in the frequency of news. Section 3 analyzes the autocorrelations of the frequency of news on a particular company and shows that the autocorrelation function follows a power law. Section 4 examines the cross correlations for the frequency of news across companies. We show that the coupling of the average number of news items on a company with its fluctuations obeys a scaling law, and that the frequency of news on a company is not governed solely by internal dynamics (i.e., a Poisson process) but is also affected by external dynamics, such as an increase in the number of news items due to the outbreak of an economic crisis. In §5, we extract common movements across companies using random matrix theory techniques. Section 6 concludes the paper.

§2. Overview of the news data

Thomson Reuters Corporation is a world-famous provider of information for businesses and professionals, providing, among other things, “Reuters 3000 Xtra”, an electronic trading platform typically used by professional traders and investment analysts in trading rooms. “Reuters 3000 Xtra” offers real-time streaming news, comprehensive economic indicators, and financial data, and displays news from not only Thomson Reuters but also around 500 third parties. From 2003 to 2009, approximately 165 million news reports were provided. While these reports were in several languages, about 65 percent of them (107 million) were in English. In this study, we use only the English news reports, all of which are available in the Reuters NewsScope Archive database.

There are three news types in the database. The first type is “alert” news, which covers an urgent newsworthy event and is 80–100 characters long. Alert news is normally followed by another news type. The second type is “headline” news, consisting of the headline of a news report for an event. The third type, finally, is

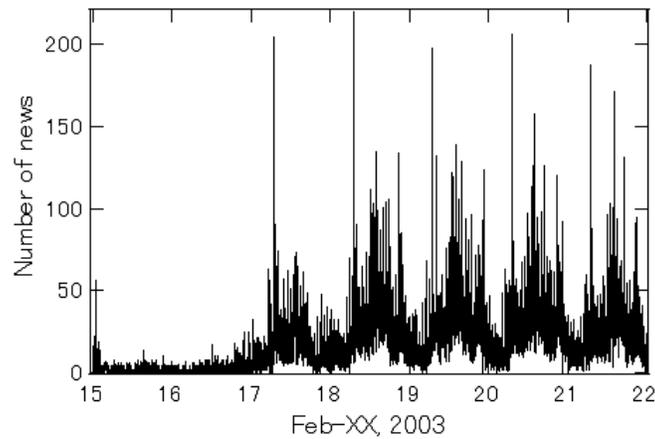


Fig. 2. Number of news reports per minute for a particular week, February 15–22, 2003. The number of news reports in English is counted.

Table I. Mean of the number of news reports on each day of the week. The number of news reports in English is counted.

	Sat.	Sun.	Mon.	Tue.	Wed.	Thu.	Fri.
ALERT	101	148	3010	3807	3960	4366	2666
HEADLINE	2398	3967	20565	23115	23111	23453	20553
STORY	2614	4316	27811	31607	31633	32428	27804

Table II. Example of news.

Date	Time	News type	Text
2010-04-05	00:03:14.307	STORY	...It topped Credit Suisse Group, which jumped from ninth place a year ago to second,...

“story” news, which contains the text that provides further information about the event. If the event is important, story news is often updated.

Figure 1 shows the time series of the number of news reports for each news type. We find that the number of news reports delivered by Reuters increases every year. There were 9.8 million news reports in 2003, but 18.6 million in 2009. Figure 2 shows the time series of the number of news reports per minute for the week starting February 15, 2003. We clearly see that the frequency of news has intraday seasonality, as has been observed previously.^{11),12)} Table I presents the mean number of news reports for each news type on each day of the week. There are fewer news reports on the weekend, indicating that before proceeding to a detailed analysis we need to deal with the nonstationarity of the time series. This is discussed in the next section.

To investigate the frequency of company news, we first need to construct time series for the number of news reports for each company. We do so using the following steps. First, we focus on the top 100 companies in the world in terms of market capitalization in 2003 and search the database by company name. For example, we find that the Credit Suisse Group is mentioned in the text of a news report published

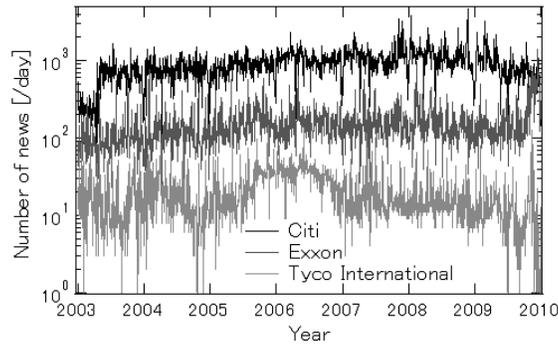


Fig. 3. Time series for the number of company news reports per day excluding the weekend. The top, middle, and bottom lines are for news on Citi, Exxon, and Tyco International, respectively.

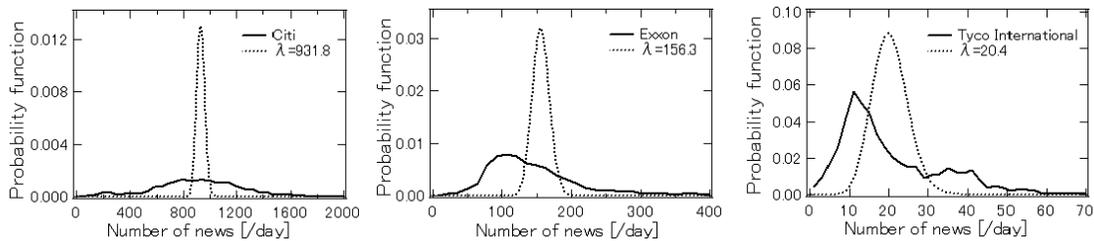


Fig. 4. Probability functions for the frequency of news on Citi, Exxon, and Tyco International. The dashed lines show Poisson distributions with the same mean as in the data. News reports on the weekend are excluded.

at 00:03:14 on April 5 (see Table II). Next, we define company news as news that mentions the name of the company. Finally, by counting the number of news reports for each company, we obtain the relevant time series.

§3. Autocorrelations for the frequency of news reports

We investigate the probability functions and autocorrelations of the frequency of company news for each news type. We focus on the time series for the number of news reports for three companies, Citi, Exxon, and Tyco International, which are shown in Fig. 3. The daily mean number of reports on Citi, Exxon, and Tyco International, excluding the weekend, is 932, 156, and 20, respectively. Figure 4 shows the probability function of the daily number of news reports for each company. Compared to the Poisson distribution that has the same mean as the data, each probability function has a fatter tail, suggesting that the time series for company news do not follow a Poisson process.

As mentioned in the previous section, the news frequency time series are not stationary due to the time trend and daily periodicity. It may be that the fat tails of the probability functions observed in Fig. 4 come from the nonstationarity of the time series. To transform our data into stationary time series, we introduce the concept of “tick time” for news. Tick time refers not to actual time, but is measured in terms of the appearance of news reports, where each news report corresponds to

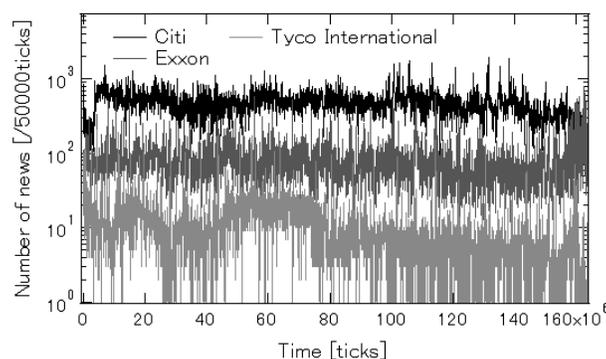


Fig. 5. Time series of the number of company news reports per 50,000 ticks. The top, middle, bottom lines are for Citi, Exxon, and Tyco International, respectively.

Table III. Augmented Dickey-Fuller test for the Citi, Exxon, and Tyco International time series.

		Trend	Drift	None
Citi	t -values	-19.7	-18.6	-1.08
	p -values	0.00	0.00	0.28
Exxon	t -values	-27.5	-26.8	-2.5
	p -values	0.00	0.00	0.01
Tyco International	t -values	-18.4	-18.0	-9.0
	p -values	0.00	0.00	0.00

a unit of “time”. That is, the tick time increases by one whenever a fresh news item in any language is reported. Note that because news reports are less frequent, and the interval between “ticks” in actual time therefore longer, on weekends, tick time passes more slowly on Saturdays and Sundays than during the weekdays, when the number of news reports is larger. We set the tick time to zero at the beginning of our sample period (January 1, 2003). Thus, using tick time allows us to eliminate the periodicity and trends observed in the original data. Figure 5 shows the time series of the number of news reports measured by tick time. In this figure, we count the number of news reports per 50,000 ticks, corresponding to about half a day, for each of the three companies. Comparing this with Fig. 3, we see that the upward trend and daily periodicity have been eliminated.

To check the stationarity of the time series with tick time, we use the Augmented Dickey-Fuller (ADF) test, which is a test for a unit root of a time series.^{13),14)} We choose the lag order of the ADF test using Akaike’s Information Criterion and conduct three types of ADF test (“none”, “drift”, and “trend”) for the time series for Citi, Exxon, and Tyco International. If the type is set to “none”, neither an intercept nor a trend is included in the test regression; if it is set to “drift”, an intercept is added; and if it is set to “trend”, both an intercept and a trend are added. Table III presents the results of the ADF test for each type, showing that the null hypothesis that the time series measured by tick time are not stationary is rejected for eight of the nine cases. In the rest of the paper, we will use tick time unless otherwise indicated.

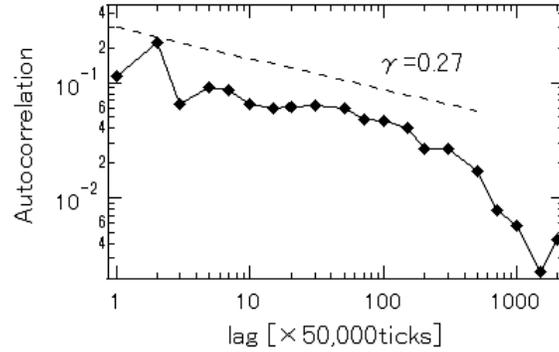


Fig. 6. Autocorrelation function of news reports. The frequency of news reports is defined as the number of news reports per 50,000 ticks, which is about half a day. The dashed reference line represents $\gamma = 0.27$.

We now turn to estimating the autocorrelation for the news frequency $f_{i,t}$ of company i using an autocorrelation function of the form

$$\rho_i(\tau) = \frac{\langle f_{i,t} f_{i,t+\tau} \rangle - \langle f_{i,t} \rangle \langle f_{i,t+\tau} \rangle}{\sigma(f_{i,t}) \sigma(f_{i,t+\tau})}, \quad (3.1)$$

where τ is a time lag, $\langle \cdot \rangle$ denotes the time average over the sample period, and σ is the standard deviation. We continue to measure the frequency by the number of news reports per 50,000 ticks, and pool observations for the top 100 companies. Figure 6 presents the estimated autocorrelation function, showing that it follows a power law of the form

$$\rho(\tau) \propto \tau^{-\gamma}, \quad (3.2)$$

where $\rho(\cdot)$ is the average of $\rho_i(\cdot)$ over the 100 companies. Note that the exponent γ is about 0.27, as represented by the reference line in the figure, and that the estimated autocorrelation decays along the reference line up to $\tau = 600$, which is equivalent to approximately one year. This indicates the presence of long memory in the frequency of news reports. Similar long memory properties have also been observed for price volatility and trading volumes in stock markets (e.g., Ref. 15)).

§4. Scaling laws for the frequency of news

In this section, we investigate correlations in the frequency of news across different companies. A useful method for examining such cross correlations in the context of complex networks, such as the internet, is to look at the average flux and fluctuations at individual nodes.^{16)–18)} It has been found that the coupling of the flux fluctuations with the total flux on individual nodes obeys a unique scaling law for a wide variety of complex networks, including the internet (i.e., a network of routers linked by physical connections), highways, river networks, and the World Wide Web of web pages and links.¹⁷⁾ Specifically, it has been shown that the average flux $\langle f \rangle$

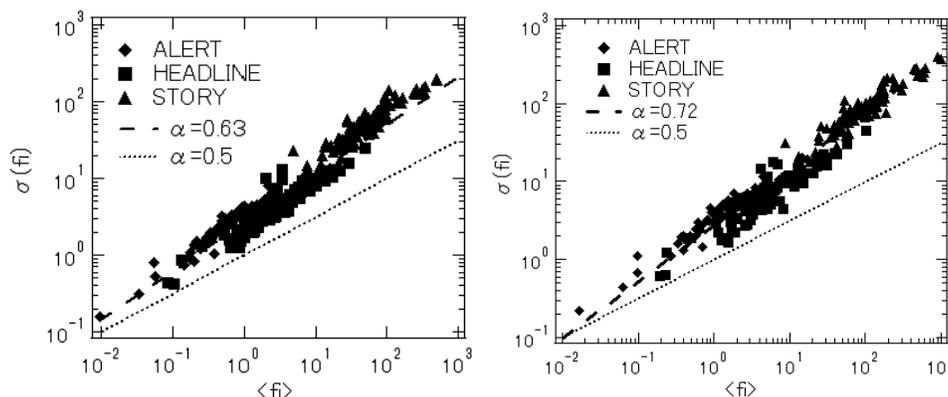


Fig. 7. The relationship between the mean and the standard deviation of the frequency of news reports for the top 100 companies in the world in terms of market capitalization in 2003. The frequency of news is defined as the number of news reports per 50,000 ticks in the left panel, while it is defined as the number of news reports per day in the right panel.

and the standard deviation σ of those individual nodes are related by¹⁷⁾

$$\sigma \sim \langle f \rangle^\alpha, \quad (4.1)$$

where α is a scaling exponent. The scaling exponent is equal to $1/2$ if the flux on individual nodes follows a Poisson process or is governed mainly by internal dynamics. On the other hand, the scaling exponent is not $1/2$ if the flux does not follow a Poisson process, and is equal to 1 if the flux on individual nodes is governed completely by external dynamics. For example, for river networks, the exponent α has been found to be quite close to unity, because the stream of rivers in different locations is mainly driven by weather patterns.

We apply this method to the frequency of news on individual companies by calculating the mean and standard deviation of the frequency of news for each company. Figure 7 plots $\sigma(f_i)$ for each of the top 100 companies as a function of the average $\langle f_i \rangle$ of the company. The frequency of news is defined as the number of news reports per 50,000 ticks in the left panel and as the number of news reports per day in the right panel. We see that in both cases the dots are not on the dotted line denoted by $\alpha = 1/2$. The estimate for α is 0.63 in the case of tick time (left panel) and even higher in the case of actual time (right panel). These results suggest that the frequency of news is governed, at least partially, by external dynamics, such as the outbreak of an economic crisis that results in a simultaneous increase in the number of news reports for each company. Note that the higher estimate of α in the right panel can be interpreted as reflecting a closer co-movement across companies due to intraday seasonality.

To see whether the scaling exponent α depends on the time scale, we estimate α for different time scales. Specifically, we count the number of news reports per s ticks, with s ranging between 5 and $100,000$. Figure 8 shows that α is close to $1/2$ for small values of s , indicating that the frequency of news is governed by internal dynamics on shorter time scales such as minutes. However, α increases monotonically with the time scale s and exceeds 0.6 for sufficiently large values of s , indicating that

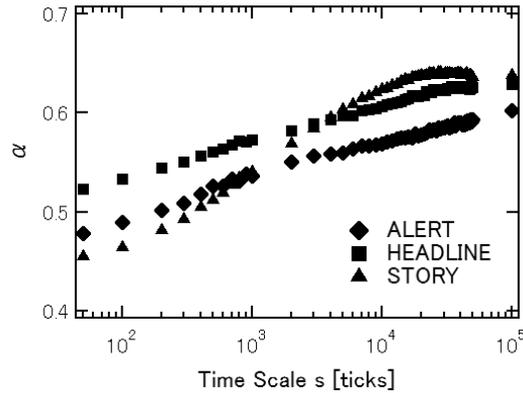


Fig. 8. Scaling exponent α for different time scales. The value of α is estimated using the observations for the top 100 companies in the world in terms of market capitalization in 2003.

the frequency of news is governed, at least partially, by external dynamics on a daily or longer time scale. Interestingly, a similar statistical property was found for transaction values on the New York Stock Exchange, namely that α is close to $1/2$ on a time scale of minutes, while it is higher and close to unity on a daily or longer time scale.¹⁶⁾

§5. Extraction of common movements across companies

To learn more about the cross correlation detected in the previous section, we extract common movements across companies by applying random matrix theory (RMT) techniques to the cross-correlation matrix for the frequency of news reports. The cross-correlation matrix \mathbf{C} is defined by

$$C_{i,j} = \frac{\langle f_{i,t} f_{j,t} \rangle - \langle f_{i,t} \rangle \langle f_{j,t} \rangle}{\sigma(f_{i,t}) \sigma(f_{j,t})}, \quad (5.1)$$

and can be decomposed as

$$\mathbf{C} = \sum_{n=1}^N \lambda_n \mathbf{A}_n \mathbf{A}_n^T, \quad (5.2)$$

where λ_n is the n -th largest eigenvalue and \mathbf{A}_n is the associated eigenvector. It has been shown that, if a cross-correlation matrix is generated from finite uncorrelated time series, the eigenvalue distribution of \mathbf{C} is given by

$$p(\lambda) = \begin{cases} \frac{Q}{2\pi} \frac{\sqrt{(\lambda_{\max} - \lambda)(\lambda - \lambda_{\min})}}{\lambda} & \text{if } \lambda_{\min} \leq \lambda \leq \lambda_{\max}, \\ 0 & \text{otherwise,} \end{cases} \quad (5.3)$$

where Q is defined as the ratio between the length of a time series L and the cross sectional dimension N (namely, $Q = L/N$), $\lambda_{\min} = \left(1 - \sqrt{1/Q}\right)^2$, and $\lambda_{\max} = \left(1 + \sqrt{1/Q}\right)^2$.^{19),20)}

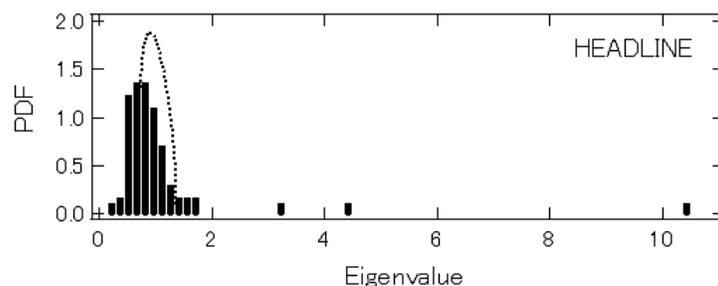


Fig. 9. The eigenvalue distribution for the case of headline news. This is estimated using the observations for the top 100 companies in the world in terms of market capitalization in 2003. The frequency of news is defined by the number of news reports per 50,000 ticks. The dotted line represents the eigenvalue distribution predicted for finite uncorrelated time series, as given by Eq. (5.3).

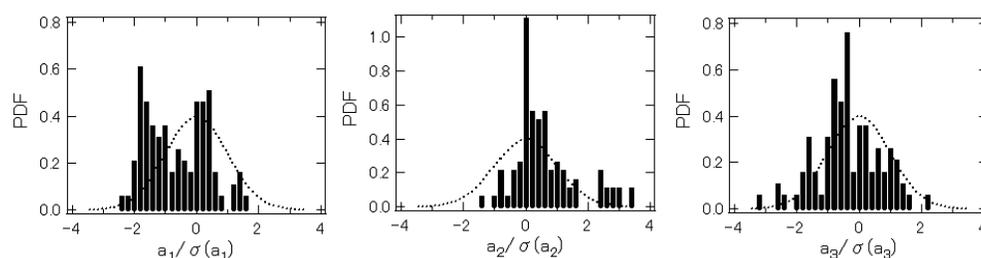


Fig. 10. The probability density functions of the eigenvector components associated with the first, second, and third largest eigenvalues. The horizontal axis shows the normalized component size (i.e., the size of the component divided by the standard deviation). The dotted line represents the standard normal distribution, which is predicted for finite uncorrelated time series.

The sample period we analyze covers seven years (January 2003 to December 2009), so that the length L of the time series is 3,274 (i.e. $3,274 \times 50,000$ ticks). As before, we pick the top 100 companies in terms of market capitalization in 2003. Given that $L = 3,274$ and $N = 100$, we have $\lambda_{\min} = 0.68$ and $\lambda_{\max} = 1.38$. Figure 9 shows the probability density function for the eigenvalues estimated from the cross-correlation matrix for the frequency of headline news, with the dotted line representing the eigenvalue distribution predicted for finite uncorrelated time series, as given by Eq. (5.3). There are eight eigenvalues exceeding λ_{\max} , with three of them exceeding λ_{\max} by a large margin.

Figure 10 presents the probability density functions for the eigenvector components associated with the largest, second largest, and third largest eigenvalues. We see that they deviate significantly from a standard normal distribution, which is predicted for finite uncorrelated time series. Figure 11 shows the degree to which each company contributes to each of the eigenvectors associated with the three largest eigenvalues. The horizontal axis represents the 100 companies sorted by industry code. The three panels, each of which corresponds to the three largest eigenvalues, show that companies belonging to the financial sector contribute greatly to the eigenvector for the second largest eigenvalue (see the middle panel), and companies belonging to the information technology sector contribute greatly to the eigenvector

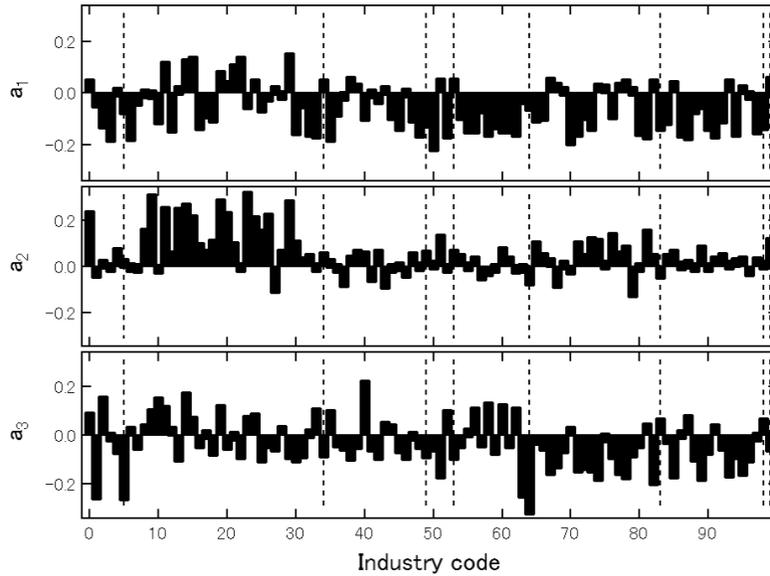


Fig. 11. Contributions of each company to the eigenvector components associated with the three largest eigenvalues of the correlation matrix. The upper, middle, and lower panels present the eigenvector components for the first, second, and third largest eigenvalues. The horizontal axis represents the 100 companies sorted by industry code. Industry codes 0-4 represent basic materials industries; 5-33 financial services industries; 34-48 consumer goods industries; 49-52 conglomerates; 53-63 services industries; 64-82 information technology industries; 83-97 health-care industries; 98 the industrial goods industry; and 99 utilities. The industry coding we use is available at http://biz.yahoo.com/ic/ind_index.html.

for the third largest eigenvalue (the bottom panel). On the other hand, the top panel shows that almost all non-financial companies contribute evenly to the eigenvector for the largest eigenvalue, which is similar to the result in Ref. 21) that the largest eigenvalue of the stock return correlation matrix is attributed to the “market mode” in financial markets.

Finally, we examine how the scaling exponent α changes when we eliminate common movement across companies. We start by defining F_t as follows:

$$F_t = \sum_i a_{1,i} f_{i,t}, \quad (5.4)$$

where $a_{1,i}$ denotes the eigenvector component i for the largest eigenvalue. A similar variable has been used to summarize common movement of stock prices (see, e.g. Ref. 20)). We then eliminate the common movement by regressing $f_{i,t}$ on F_t :

$$f_{i,t} = b_i + d_i F_t + \epsilon_{i,t}, \quad (5.5)$$

where b_i and d_i are regression coefficients. Using the residual term $\epsilon_{i,t}$ rather than $f_{i,t}$ itself, we estimate the scaling exponent α' that satisfies a relationship of the form

$$\sigma(\epsilon_{i,t}) \propto \langle f_i \rangle^{\alpha'}, \quad (5.6)$$

where $\sigma(\epsilon_{i,t})$ is the standard deviation of the residual term $\epsilon_{i,t}$. We find that the scaling exponent, which is equal to 0.63 when estimated using the original data,

decreases to 0.61 when the common movement represented by F_t is removed. This result suggests that the deviation of α from $1/2$ shown in the previous section stems, at least partially, from the common movement across companies captured by the largest eigenvalue of the correlation matrix. It is natural to suggest that the scaling exponent would approach $1/2$ when one further eliminates common movements represented by other eigenvalues. One of our future research tasks is to see whether or not this is true by developing a method to eliminate the common movement represented by these eigenvalues.

§6. Conclusion

We have empirically investigated temporal and cross correlations in the frequency of news reports on companies using a dataset of more than 100 million news articles reported in English by around 500 press agencies during the period 2003–2009. Our main findings are as follows. First, the frequency of news reports on a company does not follow a Poisson process, but is instead characterized by long memory with a positive autocorrelation lasting more than a year. Second, there exist significant correlations in the frequency of news across companies. Specifically, on a daily or longer time scale, the frequency of news is governed by external dynamics such as an increase in the number of news reports due, for example, to the outbreak of an economic crisis, while it is governed by internal dynamics on a time scale of minutes. These two findings indicate that the frequency of news on a company has similar statistical properties as trading activities in stock markets, measured by trading volumes or price volatility, suggesting that the flow of information through news on companies plays an important role in price dynamics in stock markets.

Acknowledgements

We would like to thank Aki-Hiro Sato for helpful comments on an earlier version of this paper. We also thank the Yukawa Institute for Theoretical Physics at Kyoto University, where this work was completed during the YITP-W-11-04 on “Econophysics 2011”. This work was supported in part by a Grant-in-Aid for Young Scientists (A) from the Ministry of Education, Culture, Sports, Science and Technology (No. 23686019).

References

- 1) D. M. Cutler, J. M. Poterba and L. H. Summers, *J. of Portfolio Management* **15** (1989), 4.
- 2) P. Balduzzi, E. J. Elton and T. C. Green, *J. of Financial and Quantitative Analysis* **36** (2001), 523.
- 3) M. L. Mitchell and J. H. Mulherin, *J. of Finance* **49** (1994), 923.
- 4) J.-P. Bouchaud, Y. Gefen, M. Potters and M. Wyart, *Quantitative Finance* **4** (2004), 176.
- 5) J.-P. Bouchaud, J. Kockelkoren and M. Potters, *Quantitative Finance* **6** (2006), 115.
- 6) A. Joulin, A. Lefevre, D. Grunberg and J.-P. Bouchaud, arXiv:0803.1769.
- 7) G. Harras and D. Sornette, Swiss Finance Institute Research Paper Series No. 08-16 (2008).
- 8) R. Lambiotte, M. Ausloos and M. Thelwall, *J. of Informetrics* **1** (2007), 277.

- 9) Y. Sano, K. Kasaki and M. Takayasu, *Proceedings of the 9th Asia-Pacific Complex Systems Conference, 2009*, p. 195.
- 10) D. Newman, C. Chemudugunta, P. Smyth and M. Steyvers, “Analyzing Entities and Topics in News Articles Using Statistical Topic Models”, in *Lecture Notes in Computer Science* Vol. 3975 (Springer, 2006), p. 93.
- 11) D. Leinweber and J. Sisk, “Relating News Analytics to Stock Returns”, in *The Handbook of News Analytics in Finance* (John Wiley & Sons, 2011), Chap. 6.
- 12) R. Cahan, Y. Luo, J. Jussa and M. Alvarez, Deutsche Bank Quantitative Strategy Report, July 2010.
- 13) D. A. Dickey and W. A. Fuller, *J. of American Statistical Association* **74** (1979), 427.
- 14) E. S. Said and D. A. Dickey, *Biometrika* **71** (1984), 599.
- 15) J.-P. Bouchaud and M. Potters, *Theory of Financial Risks and Derivative Pricing*, First edition (Cambridge University Press, 2000).
- 16) Z. Eisler, J. Kertesz, S. H. Yook and A. L. Barabasi, *Europhys. Lett.* **69** (2005), 664.
- 17) M. Argollo de Menezes and A. L. Barabasi, *Phys. Rev. Lett.* **92** (2004), 028701.
- 18) M. Argollo de Menezes and A. L. Barabasi, *Phys. Rev. Lett.* **93** (2004), 068701.
- 19) V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr and H. E. Stanley, *Phys. Rev. Lett.* **83** (1999), 1471.
- 20) V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr and H. E. Stanley, *Phys. Rev. E* **65** (2002), 066126.
- 21) C. Biely and S. Thurner, *Quantitative Finance* **8** (2008), 705.

Statistical Regularities of Seismic Noise

Zeyu ZHENG,^{1,*} Tetsuya TAKAISHI,³ Naoko SAKURAI,¹
Xiangguang ZHANG¹ and Kazuko YAMASAKI^{1,2}

¹*Department of Environmental Information,*

Tokyo University of Information Sciences, Chiba 265-8501, Japan

²*Center for Polymer Studies and Department of Physics, Boston University,
Boston, MA 02215, USA*

³*Hiroshima University of Economics, Hiroshima 731-0192, Japan*

Seismic noise includes the information that comes from crust of earth and hypocenter. The study of seismic noise is an essential topic for earthquake predictability and investigation of crustal structure. In this paper we analyze the seismic noise from 46 stations in all around Japan of Full Range Seismograph Network of Japan (F-net). To investigate the influence from earthquake we analyze two types of seismic noise data, one is the data after earthquake shake, which from 2,000 seconds after the max shake of an earthquake, and others are without any earthquake effects. We find that (i) the magnitude of seismic noise values follows a log-normal distribution and (ii) the conditional mean growth rate and the standard deviation of the growth rate of the magnitude value both obey power-law with the magnitude value. Furthermore we also find that the Hurst exponent depends on the max amplitude of earthquake, and shows a linear correlation with logarithmic max amplitude.

§1. Introduction

Many physical, physiological, biological, and social systems are characterized by complex interactions between a large number of individual components, which manifest in power-law correlations.¹⁾⁻¹⁴⁾ For example, in financial field, several stylized facts have been found for the equity price data in temporal field, such as 1, the distribution of the stock price changes (return) has a power-law tail. 2, the absolute value of stock price change (volatility) is long-term power-law correlated. 3, The spectral density of stock price is well described by power-law function.¹⁵⁾

In seismology temporal and spatial clustering are considered to be important properties of seismic occurrences and, together with the Omori law (dictating after-shock timing) and the Gutenberg-Richter law (specifying the distribution of earthquake size), comprise the main starting requirements to construct reasonable seismic probabilistic models. Analyzing the timing of individual earthquakes, Ref. 5) introduces the scaling concept to statistical seismology. The recurrence times are defined as the time intervals between consecutive events, $\tau_i = t_i - t_{i-1}$. In the case of stationary seismicity, the probability density $P(\tau)$ of the occurrence times was found to follow a universal scaling law

$$P(\tau) = Rf(R\tau), \quad (1.1)$$

where f is a scaling function and R is the rate of seismic occurrence, defined as the mean number of events with $M \geq M_c$.¹⁶⁾ References 17) and 18) have demonstrated

*) E-mail: zhengzy@edu.tuis.ac.jp

how the structure of seismic occurrence in time and magnitude can be treated within the framework of critical phenomena.

For seismic wave, there exist many articles which indicate that part of the spectral density follow a power-law function.^{19)–22)} This property is very similar to the spectral density of stock price. Thus it might be interesting to see whether other properties in financial markets can also be observed for temporal seismic wave, e.g. long-term correlations in the absolute value of seismic wave series and power-law tails of the distributions of the seismic wave series. In order to reveal the statistical regularities of seismic wave we analyze the broad-band seismic wave from Full Range Seismograph Network of Japan (F-net) and aim to understand the physical dynamics of seismic wave.^{5), 16)–18), 23)–28)}

The paper is organized as follows: In §2 we introduce the database and investigate the distribution of the magnitude of seismic noise magnitude values. We find that the distribution follows a log-normal function. In §3 we investigate the conditional mean growth rate and the conditional standard deviation of the growth rate. We find power-law correlations for both. Section 4 deals with the long-term correlations in the time series of the magnitude of seismic noise analyzed by using detrended fluctuation analysis (DFA) method which is characterized by the Hurst exponent H . We find the Hurst exponents $H = 0.66 \pm 0.05$ for none-earthquake data and $H = 0.69 \pm 0.1$ for after-earthquake data. We also find that the relation between exponent β in the growth rate and Hurst exponent H does not follow the expected function $H = 1 - \beta$. Furthermore we also show that the Hurst exponent H is linearly correlated with logarithmic max amplitude. In §4 we summarize our findings.

§2. Data and PDF of magnitude

In this study we use the seismic waveform database from the National Research Institute for Earth Science and Disaster Prevention (NIED) F-net, which records continuous seismic waveform data w_i (where i is the time accurate to seconds) by using broadband sensors in 46 selected stations in all over Japan. Seismic signals are recorded in three directions: (1) U (up-down with up positive), N (north-south with north positive), and E (east-west with east positive).²⁹⁾ We report results from the vertical dimension only (U data), since the results for the horizontal data (N and E) data are expected to be similar. We study two types of time series, one is surface wave data without earthquake called “none-earthquake data”, and the other is surface wave data after a earthquake called “after-earthquake data” (see Fig. 1). Sampling intervals have five recording frequencies: 80 Hz, 20 Hz, 1 Hz, 0.1 Hz, and 0.01Hz. We study surface wave data with 1Hz sampling interval for the year 2003, together with 11 March, 2011. We note that because of the interaction between earthquakes, not all earthquake data can be used in our analysis (see Appendix A). The data from 11 March, 2011 is selected because it contains the notable 2011 Tohoku earthquake (“Great East Japan Earthquake”) which resulted in the tsunami that caused a number of nuclear accidents.

For after-earthquake surface wave series data, we analyze the data as follows.

For the raw seismic acceleration waveform data of each selected earthquake (Appendix A), denoted as w_t , we create a new time series—the normalized waveform $w_{norm} \equiv (w_t - \bar{w})/\sqrt{w_t^2 - \bar{w}^2}$. Then we define a new sub-series fixed length (here the length = 5,000) series $w'_t = |w_{norm}|$, starting at time coordinate 2,000 seconds after the amplitude of signal reaches maximum. Here in order to avoid large errors which come from highly non-stationary signals, we do not analyze the series immediately after the maximum amplitude occurs, but instead analyze the data of the magnitude series 2,000 seconds after the amplitude of the signal reaches maximum. Then, we calculate the time series of the magnitude $V_t \equiv abs(w'_t)$. (Fig. 1 (b)) For none-earthquake series data, we select the data with the time length ≥ 30000 seconds which do not include any earthquakes. Then we calculate the magnitude series $V_t \equiv abs(w'_t)$.

We first examine the distributions of the magnitude. In Fig. 2, we plot the probability density function (PDF) of earthquake series of all earthquakes ($M \geq 3$) and none earthquake series of all none earthquake periods. The distribution of earthquake series is in the right of none earthquake series.

Both have a similar shape and can be well fitted by a log-normal function (as shown by the dashed line in Fig. 2),

$$f(V) \approx \frac{1}{V\delta\sqrt{2\pi}} \exp\left(\frac{-(\ln V - \mu)^2}{2\delta^2}\right), \tag{2.1}$$

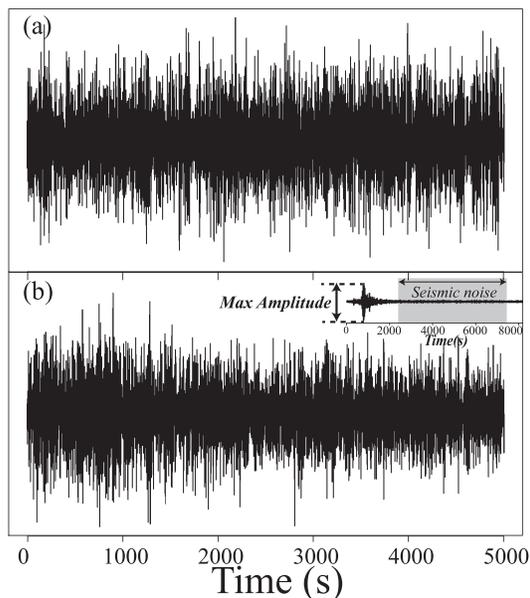


Fig. 1. Samples of seismic noise (a) none-earthquake data (b) after-earthquake data 2,000 seconds after maximum amplitude occurs.

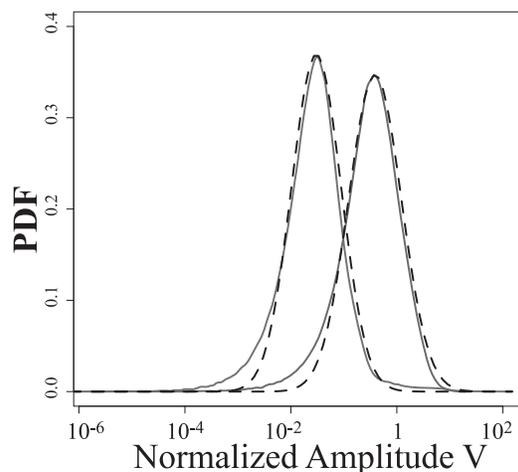


Fig. 2. Distributions of the magnitude of seismic noise. The solid lines plot the probability density functions of the seismic noise magnitude value, left one is from none-earthquake data and right one is from after-earthquake data. Both can be well fitted by a log-normal distribution (as shown by the dashed line). These results suggest that the seismic noise magnitude value follows a log-normal distribution.

where δ and μ are fitting parameters.

Here, the δ does not represent the standard deviation and μ does not represent the average. This result supports that the distribution of the magnitude values of surface waves follows before equation.

§3. Power-law of growth rate

The growth rate is often used to characterize the dynamics of the financial markets in econophysics and economics. Since the surface wave time series are very similar to the price return time series and the magnitude series are also similar to the volatility of price, it is suggested that the growth rate may also be used for our study of the surface wave magnitude series. To avoid huge fluctuations we define the growth rate at time t , $g(t)$ as the logarithmic change of two consecutive cumulative values.³⁶⁾

$$g(t) \equiv \ln\left(\frac{V(t-1) + V(t)}{V(t-1)}\right) = \ln\left(1 + \frac{V(t)}{V(t-1)}\right), \quad (3.1)$$

where the $V(t)$ stands for the magnitude value at time t . For simplicity, we denote the initial magnitude value of the growth rate, $V(t-1)$ as V_i in the following.

In this study, we examine two features of memory by calculating the growth rate:

- (i) the conditional mean growth rate $\langle g|V_i \rangle$, which quantifies the average growth rate of the magnitude value given at the initial magnitude value.

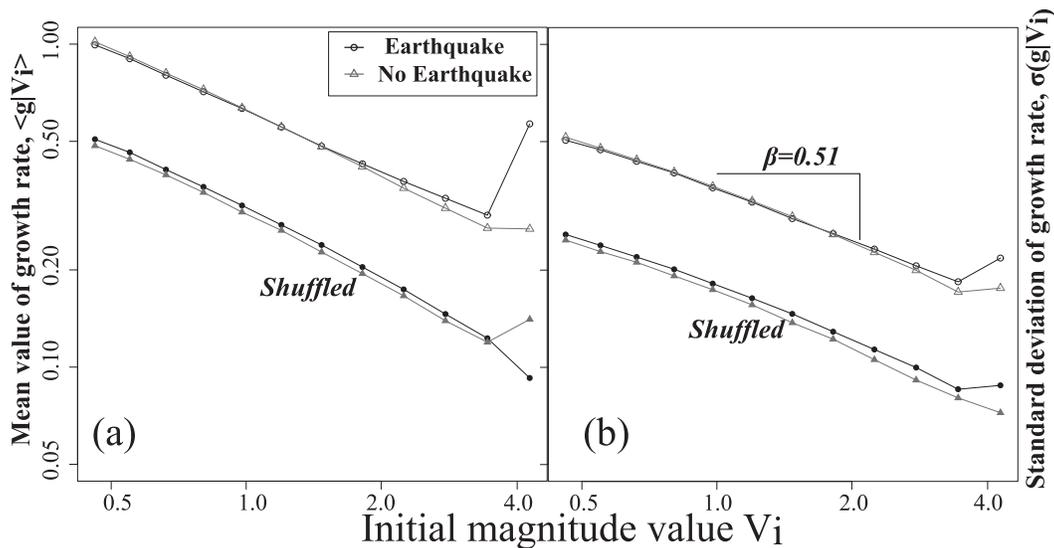


Fig. 3. (a) The conditional mean growth rate $\langle g|V_i \rangle$ and (b) conditional standard deviation of the growth rate $\sigma(g|V_i)$ as a function of the initial magnitude value V_i . Shuffled data are shown in the same panel. Remarkably, both (a) and (b) show definite power-law correlations. Also note that after-earthquake and none-earthquake data show similar power-law behavior. In (b), σ results are fitted to $\sigma \sim V_i^{-\beta}$ and the exponent $\beta = 0.51$ is obtained. The slope of shuffled data also takes the same value. These results suggest that no long-term correlations at temporal field of seismic noise exist.

- (ii) the conditional standard deviation of growth rate $\sigma(g|V_i) = \sqrt{\langle g^2|V_i \rangle - \langle g|V_i \rangle^2}$, which characterizes the fluctuation of the growth rate conditional on a given initial magnitude value V_i .

In Fig. 3(a) we plot the conditional mean magnitude value $\langle g|V_i \rangle$ as a function of initial magnitude value V_i . We analyze all time series data at once. Interestingly, both after-earthquake data and none-earthquake data show similar power-law behavior. The curves of shuffled value are also shown in the figure and they are almost parallel to the curves from the original data.

Next we plot the conditional standard deviation $\sigma(g|V_i)$ as a function of V_i in Fig. 3(b). The curves are similar to $\langle g|V_i \rangle$ and well fitted by a power-law function,

$$\sigma(g|V_i) \sim V_i^{-\beta}, \tag{3.2}$$

where the exponent β can be used to describe the memory of time series.³⁵⁾ We find that both after-earthquake and none-earthquake data show similar scaling behavior for the growth rate. The scaling exponent is found to be $\beta = 0.51 \approx 0.5$ which means that there is no definite long memory for the seismic time series.

§4. Long-term memory and dependence on max amplitude

Many physical time series have memory, where a value in the sequence depends on the previous values. Thus, the temporal structures in the seismic noise are also interesting. Figure 3 shows the scaling exponent β in the growth rate for seismic noise, which indicates the temporal correlations. From Eq. (3.1) we can find that the growth rate can only reflect the correlation of the closest following two data, though in many cases the correlation may be influenced by much more remote values. However the β value seems to show that there is no long-term memory in seismic surface data, it is necessary to check the long-term memory by other method. For further test the long-term memory in the seismic noise, we employ the DFA, a wide-used method to examine the long-range correlations in the magnitude time series.³⁰⁾⁻³²⁾ In the DFA method, the time series is partitioned into pieces of equal size n . For each piece, the local trend is subtracted and the resulting standard deviation over the entire series is obtained. In general, the standard deviation $F(l)$ of the detrended fluctuations depends on n , with smaller n resulting in trends that more closely match the data. The dependence of F on n can generally be represented as a power law such that

$$F(n) \propto n^H, \tag{4.1}$$

where H is as the Hurst exponent. DFA therefore can conceptually be understood as characterizing the motion of a random walker whose steps are given by the time series. $F(n)$ gives the walker's deviation from the local trend as a function of the trend window. Because the root mean square displacement of a walker with no correlations between his steps scales like \sqrt{n} , we can expect a time series with no autocorrelations to yield an H of 0.5. Similarly, long-range power-law correlations in the signal (i.e. large values follow large values and small values follow small

values) manifest as $H > 0.5$. Power-law anti-correlations within a signal will result in $H < 0.5$. Additionally, DFA can be related to the autocorrelation as follows: if the autocorrelation function $C(L)$ can be approximated by a power law with exponent γ such that

$$C(L) \propto L^{-\gamma}, \quad (4.2)$$

where L indicate the lag-time, γ is related to α by³⁰⁾

$$H \approx 1 - \gamma/2. \quad (4.3)$$

Figure 4 plots the PDF of scaling exponent H values (a) is from none earthquake data $H = 0.66 \pm 0.05$ (b) is from after earthquake data $H = 0.69 \pm 0.01$. Similar to PDF of β , none earthquake and after earthquake data shows similar probability density function, but not as β the H value from shuffled data and original data shows completely different PDFs. Almost all scaling exponents H are bigger than 0.5, it indicates definite long-range correlated correlations. Because Hurst exponent $H \gg 0.5$ that means correlated correlation, it suggests that seismic surface wave data show a definite long-range correlated correlations. Previous studies have indicated that the relation between the scaling in the growth rate and the long-range correlation is consistent with $\beta = 1 - H$ in Human interaction³⁵⁾ and Equity market activity.³⁶⁾ This equation is based on a series of physical assumptions, such as stationarity of time series, power-law attenuation of autocorrelation, uniform distribution of Growth rate, etc, for details refer to Ref. 35). The β and H of seismic noise do not satisfy the equation that is founded in human interaction activity.

As we know, the long-range correlations may be affected by many factors, such as the trading value, size, activity, risk and return of the stock.^{11),15),36)} In after earthquake data, we note that the max amplitude varies in a very huge field. Therefore we test the relation between H and the max amplitude. Figure 5 plots

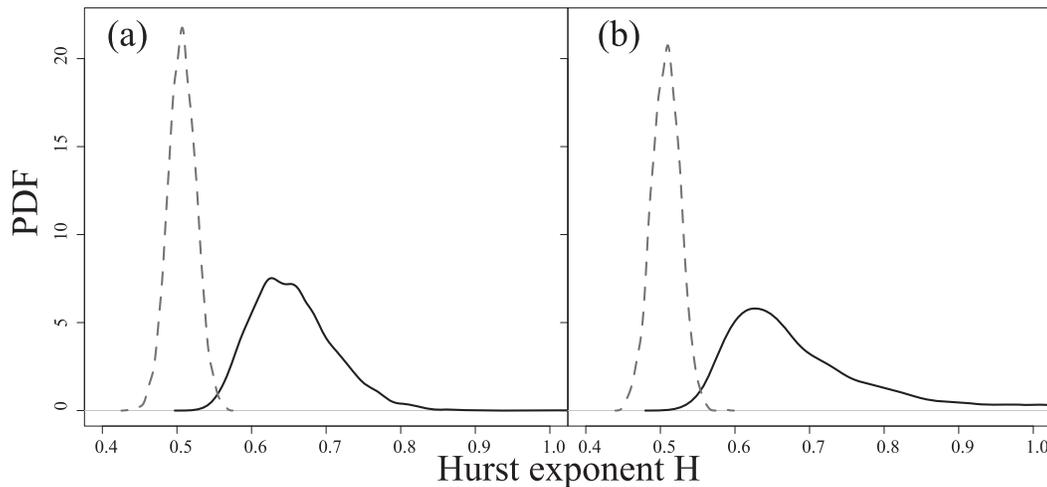


Fig. 4. Probability density functions of the Hurst exponent H values. (a) shows the none earthquake data, (b) shows the after earthquakes data. The PDF of Shuffled data are plotted by the dashed lines.

the correlation of H vs logarithmic max amplitude. The plot shows that the H is correlated with max amplitude, and shows a linear correlation with logarithmic max amplitude. Since the max amplitude refers to the energy contained in an earthquake that means the size of earthquake, the results also indicate that Hurst exponent H depends on the size of earthquake.

§5. Summary

We studied the seismic noise from 46 station of F-net. We found that the distribution of magnitude of surface noise follows a log-normal function for both none-earthquake and after-earthquake data. We also found that the conditional standard deviation of growth rate has a power-law dependence on the initial magnitude value. The scaling exponent of growth rate is found to be $\beta = 0.51$, which indicates that no significant autocorrelations exist in seismic surface noise data. On the other hand most of the Hurst exponents H of seismic surface noise data are close to 0.7, which indicates the existence of long-memory. This result does not satisfy the linear correlation function $\beta = 1 - H$ as seen in financial data and human interaction data.^{35),36)} Therefore there must be some differences between the seismic noise data and others, e.g. some assumptions needed for the relation of $\beta = 1 - H$ might not be fulfilled. Since this is the unique features of seismic wave, in order to understand the dynamics of the seismic noise more comprehensive studies are needed.

By using DFA method we find that both none-earthquake data and after-earthquake data show similar long-term correlation behaviors. We also find that the Hurst exponents are logarithmically dependent on the Max amplitude in after-earthquake surface wave data.

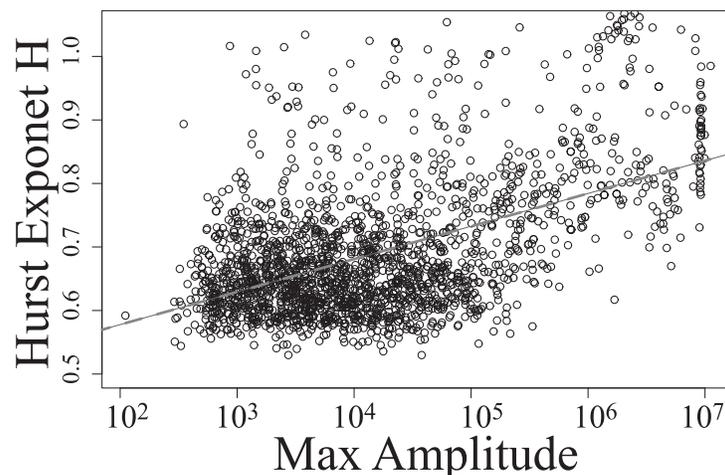


Fig. 5. Hurst exponent H vs max amplitude. The relation coefficient is equal to 0.507, the slope of linear regression analysis (least squares method) is 0.052 ± 0.002 , t value is 26.75 and P value is smaller than 0.01. It suggests that Hurst exponents H are dependent on Max amplitudes.

Acknowledgements

We thank S. Havlin for his constructive suggestions, and thank JSPS for grant of “Research project for a sustainable development of economic and social structure dependent on the environment of the eastern coast of Asia” that made it possible to complete this study. We also thank the Yukawa Institute for Theoretical Physics at Kyoto University. Discussions during the YITP workshop YITP-W-11-04 on “The Hitchhiker’s Guide to the Economy” were useful to this study.

Appendix A

— The Selection of Earthquakes —

In some regions it is common for multiple earthquakes to occur in short succession. In many cases, because the inter occurrence times are so short, the surface waves can be derived from more than one earthquake. This is especially true for large earthquakes with many aftershocks.³⁴⁾ In order to make sure that the surface waves we study are the effects of only one earthquake, we need a way of determining which earthquakes are independent. We use the following two functions to determine the sphere of influence and duration of each earthquake by using the Richter magnitude scale M .³⁴⁾ We select only those earthquakes that have no larger earthquake in their spatiotemporal sphere of influence,

$$t \approx 10^{(M-4.71)/1.67} \quad (\text{A}\cdot 1)$$

and

$$R \approx 2 \times 10^{(M+1)/2.7}, \quad (\text{A}\cdot 2)$$

where t is the duration and R is the sphere radius of influence, t and R units are hours and kilometers respectively. The two functions are empirical formulas based on an analysis of earthquakes in Japan.³⁴⁾ The $10^{M+1.0}/2.7$ is an empirical formula that indicates the maximum radius that a human can feel an earthquake, especially for the earthquakes in Japan.

References

- 1) M. F. Shlesinger, Ann. NY Acad. Sci. **504** (1987), 214.
- 2) A. Bunde and S. Havlin (Ed.), *Fractals in Science* (Springer, Berlin, 1994).
- 3) Y. Ashkenazy, P. C. Ivanov, S. Havlin, C. K. Peng, A. L. Goldberger and H. E. Stanley, Phys. Rev. Lett. **86** (2001), 1900.
- 4) R. F. Engle, Econometrica **50** (1982), 987.
- 5) P. Bak, K. Christensen, L. Danon and T. Scanlon, Phys. Rev. Lett. **88** (2002), 178501.
- 6) K. Yamasaki, A. Gozolchiani and S. Havlin, Phys. Rev. Lett. **100** (2008), 228501.
- 7) A. Gozolchiani, K. Yamasaki, O. Gazit and S. Havlin, Europhys. Lett. **83** (2008), 28005.
- 8) B. Podobnik, P. Ch. Ivanov, V. Jazbinsek, Z. Trontelj, H. E. Stanley and I. Grosse, Phys. Rev. E **71** (2005), 025104(R).
- 9) K. Yamasaki, L. Muchnik, S. Havlin, A. Bunde and H. E. Stanley, Proc. Natl. Acad. Sci. **102** (2005), 9424.
- 10) F. Wang, K. Yamasaki, S. Havlin and H. E. Stanley, Phys. Rev. E **73** (2006), 26117.
- 11) F. Wang, K. Yamasaki, S. Havlin and H. E. Stanley, Phys. Rev. E **79** (2009), 016103.
- 12) H. E. Stanley, V. Plerou and X. Gabaix, Physica A **387** (2008), 3967.

- 13) B. Podobnik, D. Horvatic, A. M. Petersen and H. E. Stanley, Proc. Natl. Acad. Sci. **106** (2009), 22079.
- 14) B. Podobnik, D. Horvatic, A. M. Petersen, Branko Urosevic and H. E. Stanley, Proc. Natl. Acad. Sci. **107** (2010), 18325.
- 15) R. Mantegna and H. E. Stanley, *An Introduction to Econophysics Correlations and Complexity in Finance* (Cambridge University Press, New York, 2000).
- 16) A. Corral, Phys. Rev. Lett. **92** (2005), 108501.
- 17) A. Corral, Phys. Rev. Lett. **95** (2005), 028501.
- 18) E. Lippiello, C. Godano and L. de Arcangelis, Phys. Rev. Lett. **98** (2008), 098501.
- 19) K. Mayeda and W. Walter, J. of Geophysical Research **101** (1996), 11195.
- 20) Y. Hisada, Bull. Seismol. Soc. Am. **90** (2000), 387.
- 21) Y. Hisada, A. Shibaya and M. R. Ghayamghamian, Bulletin Earthquake Research Institute, University of Tokyo **79** (2004), 81.
- 22) K. Aki, J. of Geophysical Research **72** (1967), 1217.
- 23) Y. Y. Kagan, Bull. Seismol. Soc. Am. **94** (2004), 1207.
- 24) E. Lippiello, L. de Arcangelis and C. Godano, Phys. Rev. Lett. **100** (2008), 038501.
- 25) M. Bottiglieri, L. de Arcangelis, C. Godano and E. Lippiello, Phys. Rev. Lett. **104** (2010), 158501.
- 26) S. Lennartz, A. Bunde and D. L. Turcotte, Geophys. J. Int. **184** (2010), 1214.
- 27) V. L. Livina, S. Havlin and A. Bunde, Phys. Rev. Lett. **95** (2005), 208501.
- 28) S. Lennartz, V. N. Livia, A. Bunde and S. Havlin, Europhys. Lett. **81** (2008), 69001.
- 29) Y. Okada, K. Kasahar, S. Hori, K. Obara, S. Sekiguchi, H. Fujiwara and A. Yamamoto, Earth Planets Space **56** (2004), xv.
- 30) C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley and A. L. Goldberger, Phys. Rev. E **49** (1994), 1685.
- 31) K. Hu, Z. Chen, P. C. Ivanov, P. Carpena and H. E. Stanley, Phys. Rev. E **64** (2001), 011114.
- 32) Z. Chen, P. C. Ivanov, K. Hu and H. E. Stanley, Phys. Rev. E **65** (2002), 041107.
- 33) H. Takayasu, *Fractals in the Physical Sciences* (Manchester University Press, Manchester, 1997).
- 34) T. Utu, *Seismicity studies: a comprehensive review* (University of Tokyo Press, Tokyo, 1999).
- 35) D. Rybski, S. V. Buldyrev, S. Havlin, F. Liljeros and H. A. Makse, Proc. Natl. Acad. Sci. **106** (2009), 12640.
- 36) F. Wang, K. Yamasaki, S. Havlin and H. E. Stanley, arXiv:0911.4258.

Zipf's Law and Heaps' Law Can Predict the Size of Potential Words

Yukie SANÓ,¹ Hideki TAKAYASU^{2,3} and Misako TAKAYASU⁴

¹*College of Science and Technology, Nihon University, Funabashi 274-8501, Japan*

²*Sony Computer Science Laboratories, 3-14-13 Higashi-Gotanda,
Tokyo 141-0022, Japan*

³*Meiji Institute for Advanced Study of Mathematical Sciences,
Kawasaki 214-8571, Japan*

⁴*Department of Computational Intelligence and Systems Science,
Interdisciplinary Graduate School of Science and Engineering,
Tokyo Institute of Technology, Yokohama 226-8502, Japan*

We confirm Zipf's law and Heaps' law using various types of documents such as literary works, blogs, and computer programs. Independent of the document type, the exponents of Zipf's law are estimated to be approximately 1, whereas Heaps' exponents appear to be dependent on the observation size, and the estimated values are scattered around 0.5. By definition, randomly shuffled documents reproduce Zipf's law and Heaps' law. However, artificially generated documents using the empirically observed Zipf's law and number of distinct words do not reproduce Heaps' law. We demonstrate that Heaps' law holds for artificial documents in which a certain number of distinct words are added to empirically observed distinct words. This suggests that the number of potential distinct words considered in the creation of a given document can be predicted.

§1. Introduction

Zipf's law¹⁾ is an empirical law stating that word frequency in documents is inversely proportional to word rank in descending order of occurrence. Zipf's law is consistent with the following power law cumulative distribution of the number of word appearances:

$$P(\geq x) \propto x^{-\alpha}, \quad (1.1)$$

where the exponent α is a positive constant. Zipf's law is applicable not only to word frequencies in documents but also to incomes of firms and individuals, the sizes of gypsum fragments, and the abundances of expressed genes.²⁾⁻⁴⁾ As such, Zipf's law has attracted a great deal of attention, but the general mechanism of Zipf's law remains unclear. In quantitative linguistics, Zipf's exponent, α , is evaluated for various languages, such as English and Russian,⁵⁾ and for programming languages, such as Java and C,⁶⁾ and is known in many cases to be approximately 1.

Heaps' law⁷⁾ states that the number of distinct words increases nonlinearly as the total number of words in a document increases. The number of distinct words $D(n)$ among the first n words of a document is approximated by the following power law:

$$D(n) \propto n^\beta. \quad (1.2)$$

Heaps' exponent, β , is smaller than 1, and Araújo et al. estimating from newspaper articles in the Wall Street Journal and from scientific papers concluded β is 0.4

to 0.6.⁹⁾ Based on several thousands of web pages, Baldi et al. estimated that $\beta = 0.76$,¹⁰⁾ and Zhang reported that, in computer programming tokens, which include several identifiers, such as “:”, β takes a value of between 0.540 and 0.869.⁶⁾

Baeza-Yates et al. reported that Zipf's law and Heaps' law are equivalent and that $\beta = \alpha$ when $\alpha < 1$.¹¹⁾ Leijenhorst et al. developed Heaps' law from the Zipf-Mandelbrot law in a more sophisticated manner and obtained the same result.¹³⁾ Lü et al. showed by analytical and simple numerical simulation that, in the case of $\alpha \geq 1$, the value of β is equal to 1.¹²⁾ Cattuto et al. observed these two empirical laws in social bookmark data that focused on word tag co-occurrence distributions for Zipf's law and Heaps' law curve for number of distinct word tags. Zipf's law and Heaps' law yielded the same results, i.e., $\beta = \alpha = 0.7$, for both empirical data and simple network simulations.^{14),15)} Furthermore, when applied in ecology especially in island biogeography, Zipf's law is valid for species abundance distributions and Heaps' law describes species-area relationships in which the number of species found within an area increases nonlinearly for increasing area.^{16)–18)} In this case, $\beta \simeq \alpha \simeq 0.5$.¹⁷⁾ Note that, although the total number of species can be estimated based on the size of the area, species-area relationships are not fully equivalent linguistics with respect to Heaps' law. Heaps' law is also known as rarefaction curve in ecology.^{19),20)} In this case, one can plot the number of species as a function of individuals sampled. Although it does not assume Zipf's law, it allows the calculation of the species richness for a given number of sampled individuals. Consequently, it is an important issue to preserve biodiversity.

In this paper we describe our data and show empirical results from the data in §2. In §3 we check the validity of Heaps' law under the condition of words from shuffled real documents and artificially generated documents by comparing empirical results. Finally in §4 we indicate predictability of the size of potential words in a certain type of document from the simulation.

§2. Empirical results

In this section, we confirm the validity of Zipf's law and Heaps' law for various types of real documents. We chose and analyzed a range of documents, covering various languages, professional and amateur authors and human spoken natural languages and computer processing formal languages.

- “The Adventures of Sherlock Holmes” was written in English by Sir Arthur Conan Doyle during the 1890s and is one of most frequently downloaded documents from Project Gutenberg, which freely provides public domain book content on the web.²¹⁾
- “Don Quixote” was written in Spanish by Miguel de Cervantes during the 16th century and is the most frequently downloaded book in the Spanish category of Project Gutenberg.
- “Light and Darkness (Meian)” was written in Japanese by Soseki Natsume during the 1910s. This unfinished novel is the longest work of Soseki Natsume, who is one of the most popular novelists in Japanese history. Although this novel is unfinished, we got similar results from his novel “Kokoro” and “I am a

Table I. Data description of documents.

	Language	Total words N	α	β	β_{rand}
Sherlock Holmes	English	107,219	0.97	0.54	0.52
Don Quixote	Spanish	390,402	0.98	0.65	0.57
Light and Darkness	Japanese	180,623	1.04	0.49	0.46
Blog	Japanese	308,687	0.98	0.57	0.50
NumPy	Python	850,699	1.10	0.68	0.50

Cat (Wagahai wa neko de aru)”.

- Blog entries posted by a single anonymous blogger in Japanese who was randomly selected from among 20 thousand bloggers. We collected 5,000 blog entries posted by the author from 2007 to 2010 and confirmed that the blog was not a typical spam blog. We accumulated his 5,000 blog entries into one document in sequence of time.
- “NumPy”, which is an open-source Python program developed by several contributors to provide support for large, multi-dimensional arrays and matrices, along with a large library of high-level mathematical functions to operate on these arrays. NumPy includes 397 small programs.²²⁾ We ignored flows of program procedures, such as loops and concatenated all Python programs in the library into one document. We also removed all comments written in natural language sentences but did not remove symbols.

The details of the documents are listed in Table I. In the case of Japanese, it is necessary to separate words, which are not separated by spaces. We use the “MeCab” morphological analyzer with a dictionary that added new words from Wikipedia titles and Hatena keywords’ titles on March 2011 to accurately separate words, including new words from blogs. In all of the documents except NumPy, we removed all symbols, such as periods and colons, from the documents.

Figures 1 and 2 demonstrate the validity of Zipf’s law and Heaps’ law based on empirically obtained data and estimated exponents are shown in Table I using the Gauss-Newton algorithm to minimize the sum of the squares of the errors in the whole area under the graph. As expected, Zipf’s exponent α is approximately 1 in all cases, which is confirmed to be universal. On the other hand, Heaps’ exponent β takes values in the range of from 0 to 1. Although our results are estimated by linear scale in whole areas, the area of large n dominates and thus the Heaps’ exponent tends to be close to 0.5. On the other hand, previous studies, for example,⁵⁾ are often estimated with an accuracy close to 1 by using log-log scale and the areas of small n are playing a central role to the estimation. Note that NumPy consists primarily of small programs developed by numerous different authors, the same names of variables tend to appear together in the same programs. Therefore, unlike for other spoken languages, the curve is not smooth and depends on the order of concatenation. In order to avoid this, we averaged 10 different patterns of Heaps’ curves for concatenation.

Finally, we want to clarify the importance of having a limit of words on Heaps’ law which states that distinct words grow by power functions in empirical data.

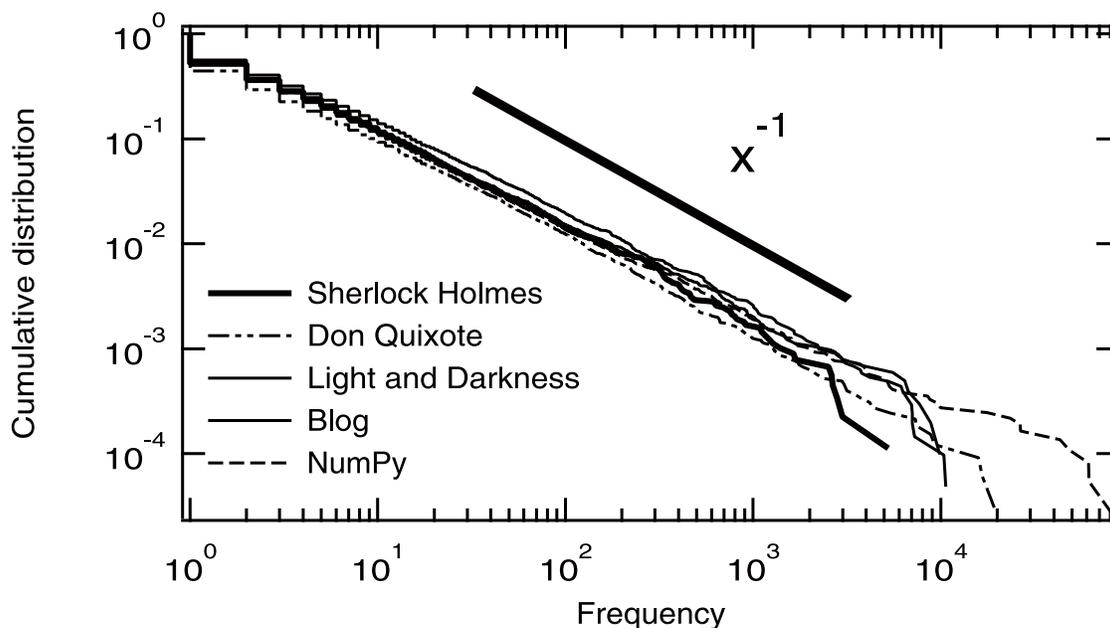


Fig. 1. Zipf's law for an English novel, a Spanish novel, a Japanese novel, a Japanese blog, and a computer program. Zipf's exponent α is approximately 1 in all cases.

While the number of words is limited in languages, there should be a point of saturation in very long documents. Therefore, when $n \rightarrow \infty$, $\beta \rightarrow 0$. On the other hand, at the beginning of the documents, new distinct words tend to appear one after another. Thus, $\beta \rightarrow 1$. In other words, Heaps' exponent depends on the value of n , and, for very long documents, a single Heaps' exponent cannot describe the entire document.

§3. Validation of Heaps' law

In this section, using the empirical data of the present study, we confirm that Heaps' law cannot be simply derived from Zipf's law.

3.1. Shuffled documents

As mentioned in previous studies,^{10)–13)} if Heaps' law can be derived from Zipf's law alone, then we should be able to reproduce Heaps' law from randomly shuffled documents that ignore correlations between word occurrences. The circles in Fig. 3 indicate the shuffled results for the case of a blog which we shuffled 10 times with different random seeds and used the average. Since Heaps' exponents for a randomized document β_{rand} are almost the same as the value shown in Table I, Heaps' law can be regarded as the derivative of Zipf's law.

3.2. Simulation using empirical Zipf's law

Here, we generated artificial words using Zipf's law with the same empirically observed values of α and the same total number of words N and number of distinct

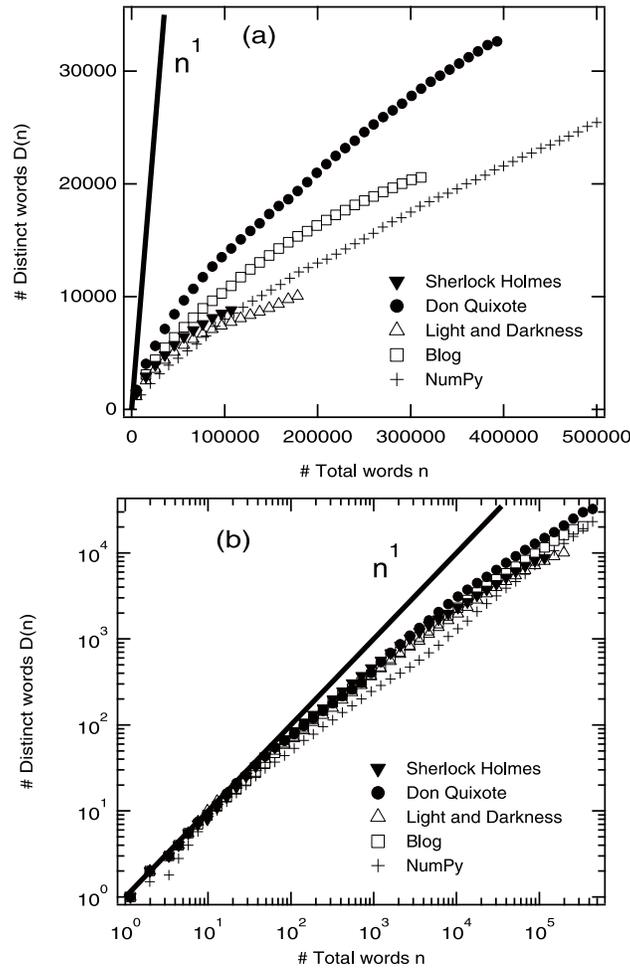


Fig. 2. Heaps' law for an English novel, a Spanish novel, a Japanese novel, a Japanese blog, and a computer program. Data is plotted (a) linearly and (b) on a log-log scale.

words $D(N)$. The author of the document is assumed to have $D(N)$ distinct words in his mind, and these words are assumed to be selected randomly with a probability based on the observed Zipf's law. The triangles in Fig. 3 (Simulation 1) show the results of this simulation using blog data. There is clearly a significant difference from the real data.

Table II. Number of observed distinct words $D(N)$ in documents containing a total of N words and $D^*(N)$ optimized distinct words in simulations.

	Distinct words $D(N)$	Simulated distinct words $D^*(N)$
Sherlock Holmes	8,910	14,800
Don Quixote	32,687	77,800
Light and Darkness	10,132	13,100
Blog	19,324	34,300
NumPy	36,676	101,900

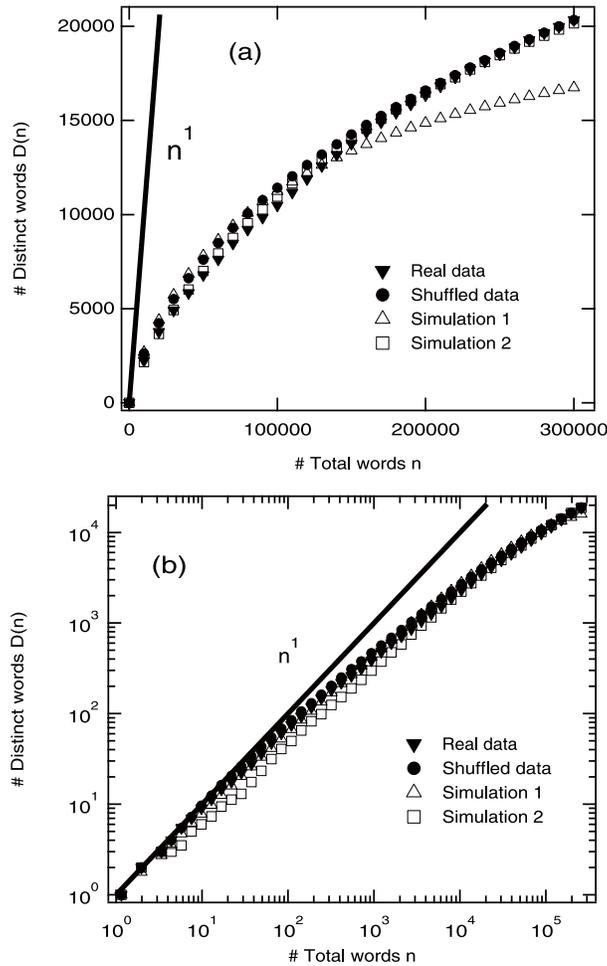


Fig. 3. Heaps' law for a Japanese blog (downward triangles), a randomly shuffled document (circles), simulation using a real number of distinct words $D(N)$ (triangles, Simulation 1), and simulation using a number of optimized distinct words $D^*(N)$ (squares, Simulation 2). Data is plotted (a) linearly and (b) on a log-log scale.

In order to clarify the reason for this discrepancy, we vary number of distinct words $D(N)$ while leaving Zipf's exponent α and the total number of words N unchanged. We select the number of distinct words $D^*(N)$ so as to minimize the sum of the squares of the residual errors. The squares in Fig. 3 (Simulation 2) indicate the modified simulation results, which are much closer to the results for the real data. The potential number of distinct words, $D^*(N)$, can be estimated in this manner for other examples, as summarized in Table II and Fig. 4. In any case, Heaps' law can be well reproduced for all examples by simulation using the optimized parameters, $D^*(N)$.

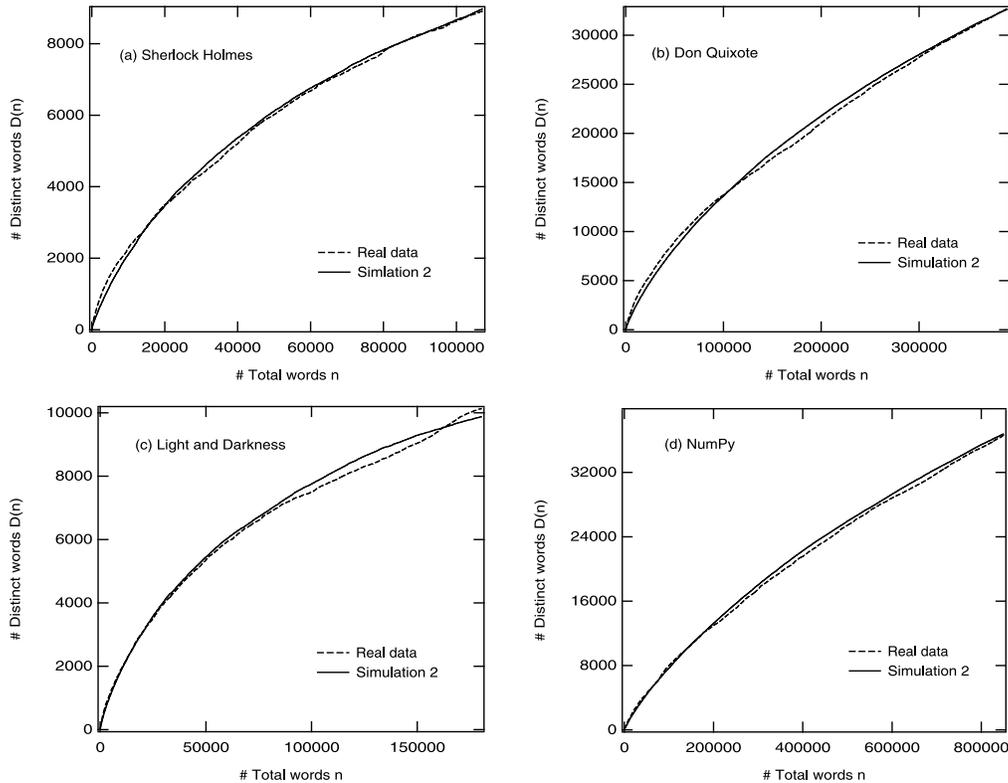


Fig. 4. Comparison of real data (dashed line) and simulated data obtained using the optimized parameters listed in Table II (solid line) for Heaps' law. (a) "The Adventures of Sherlock Holmes", (b) "Don Quixote", (c) "Light and Darkness (Meian)", and (d) NumPy.

§4. Summary and discussion

In the present study, we examined the validity of Zipf's law and Heaps' law for various types of documents. With respect to the shapes of the functions, Heaps' law is regarded as a deviation of Zipf's law, as reported in a previous study. We showed that a random sampling of N words from $D(N)$ distinct words cannot reproduce Heaps' law. However, Heaps' law can be revised using an estimated potential number of distinct words, $D^*(N)$. The obtained results indicate that we can estimate the potential number of words considered in the creation of given documents using Zipf's law and Heaps' law with real data.

Acknowledgements

The authors would like to thank Dentsu Kansai Inc. and Hottolink Inc. for providing the blog data. The present study was supported in part by a Grant-in-Aid for Scientific Research No. 22656025 (M.T.) from MEXT, Japan.

References

- 1) G. K. Zipf, *Human behavior and the principle of least effort* (Addison-Wesley, Cambridge, 1949).
- 2) K. Okuyama, M. Takayasu and H. Takayasu, *Physica A* **269** (1999), 125.
- 3) C. Furusawa and K. Kaneko, *Phys. Rev. Lett.* **90** (2003), 088102.
- 4) M. E. J. Newman, *Contemporary Physics* **46** (2005), 323.
- 5) A. Gelbukh and G. Sidorov, *Lecture Notes in Computer Science* **2004** (2001), 332.
- 6) H. Zhang, *Information Processing and Management* **45** (2009), 477.
- 7) H. S. Heaps, *Information Retrieval: Computational and Theoretical Aspects* (Academic Press, Orlando, 1978).
- 8) R. Baeza-Yates and B. Ribeiro-Neto, *Modern information retrieval* (ACM Press, 1999).
- 9) M. D. Araújo, G. Navarro and N. Ziviani, in *Proc. Fourth South American Workshop on String Processing*, Carleton University Press International Informatics Series **8** (1997), 2.
- 10) P. Baldi, P. Frasconi and P. Smyth, *Modeling the internet and the web: Probabilistic methods and algorithms* (Wiley, 2003).
- 11) R. Baeza-Yates and G. Navarro, *J. of the American Society for Information Science* **51** (2000), 69.
- 12) L. Lü, Z.-K. Zhang and T. Zhou, *PLoS ONE* **5** (2010), e14139.
- 13) D. C. van Leijenhorst and Th. P. van der Weide, *Inf. Sci.* **170** (2005), 263.
- 14) C. Cattuto, V. Loreto and L. Pietronero, *Proc. Natl. Acad. Sci.* **104** (2007), 1461.
- 15) C. Cattuto, A. Barrat, A. Baldassarri, G. Schehr and V. Loreto, *Proc. Natl. Acad. Sci.* **106** (2009), 10511.
- 16) O. Arrhenius, *J. Ecol.* **9** (1921), 95.
- 17) H. Irie, K. Tokita and H. Habara, *Publ. RIMS, Kyoto Univ.* **1432** (2005), 116 (in Japanese).
- 18) H. Irie and K. Tokita, q-bio/0609012.
- 19) R. Colwell and J. Coddington, *Philos. Trans. R. Soc. London B* **345** (1994), 101.
- 20) M. Arumugam, J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D. R. Mende, G. R. Fernandes et al., *Nature* **473** (2011), 174.
- 21) Digital libraries providing public domain books.
Gutenberg Project, <http://www.gutenberg.org/>
and Aozora bunko, <http://www.aozora.gr.jp/>
- 22) Extension to the Python programming language.
NumPy (Version1-6-1), <http://numpy.scipy.org/>
- 23) List of Japanese common new words.
Wikipedia titles, <http://ja.wikipedia.org>
and Hatena keywords, <http://d.hatena.ne.jp/keyword/>

Conservation Laws and Symmetries in Competitive Systems

Lisa UECHI*) and Tatsuya AKUTSU**)

*Bioinformatics Center, Institute for Chemical Research, Kyoto University,
Uji 611-0011, Japan*

We investigate a conservation law of a system of symmetric $2n$ -dimensional nonlinear differential equations. We use Lagrangian approach and Noether's theorem to analyze Lotka-Volterra type of competitive system. We observe that the coefficients of the $2n$ -dimensional nonlinear differential equations are strictly restricted when the system has a conserved quantity, and the relation between a conserved system and Lyapunov function is shown in terms of Noether's theorem. We find that a system of the $2n$ -dimensional first-order nonlinear differential equations in a symmetric form should appear in a binary-coupled form (BCF), and a BCF has a conserved quantity if parameters satisfy certain conditions. The conservation law manifests characteristic properties of a system of nonlinear differential equations and can be employed to check the accuracy of numerical solutions in the BCF.

§1. Introduction

Nonlinear dynamical systems characterized by self-interactions, self-organizations, spontaneous emergence of order,¹⁾ dissipative structure²⁾ and nonlinear cooperative phenomena have shown essential roles in natural sciences, as well as economy, ecology, and environmental sciences.^{3),4)} Nonlinear dynamical systems are difficult to handle unlike linear dynamical systems because their complex interactions and structures make it so hard to understand a response of a system, which may exhibit no simple laws or orders. However, in terms of natural sciences, conservation laws, symmetries and orders in nonlinear dynamical systems are expected to exist even in biology,⁵⁾ ecology, and economy. The important examples in the field of ecology are those of Malthus (1959) for a population analysis, Alfred Lotka (1925)⁶⁾ and Vito Volterra (1926)⁹⁾ for predator-prey differential equations known as Lotka-Volterra (LV) equation. Also, in the field of economy, a conserved quantity in the system of business cycle⁷⁾ is studied and this business cycle model is regarded as a predator-prey type competitive system, and also a mathematical model for Lanchester Strategic Management is known as a predator-prey type competitive system.⁸⁾ We employed the Lagrangian approach to examine a system of $2n$ -dimensional nonlinear differential equations which contains linear interactions and Lotka-Volterra type of nonlinear interactions.

The research on symmetries and the first integrals of the LV system attracted attention in the 20th century and Lagrangian approach to equations of motion is known as a remarkable method to analyze conservation laws in nonlinear dynamics¹⁰⁾ and generalized to space-time 4-dimensional and multi-dimensional systems.^{11),12)} The concept of symmetry provides us with conservation laws and enables us to find sta-

*) E-mail: uechir@kuicr.kyoto-u.ac.jp

***) E-mail: takutsu@kuicr.kyoto-u.ac.jp

ble solutions of nonlinear differential equations.¹⁴⁾ Several methods such as the Lie method,^{15),16)} Painlevé analysis,¹⁷⁾ have been used in order to search conservation laws or symmetries. For instance, José Fernández-Núñez studied symmetries of two-dimensional LV system.¹³⁾ They discussed a Lagrangian structure in LV system with Lagrangian linear in velocities. However, we found a conservation law which is velocity-independent, and the symmetric nonlinear differential equations of $2n$ -dimensional ND system will be derived from conservation laws by Noether's theorem.¹⁸⁾ The symmetry and conservation law in the $2n$ -dimensional first-order differential equations require a binary coupled form, and hence, the system of the nonlinear differential equations becomes the even dimensions $(2, 4, 6, \dots, 2n)$. We denote the requirement of the symmetric form as the binary-coupled form (BCF).

In this paper, we discuss the existence of conservation laws in a $2n$ -dimensional competitive system with general nonlinear interactions. Competitive systems are well known in ecology and biology, as well as in the fields of engineering and information systems.²⁵⁾ We discuss that symmetries, conservation laws of nonlinear interactions are important in nature by analyzing a general nonlinear dynamical (ND) system using Noether's theorem.

In §2, we introduce notations, Euler-Lagrange equation, and Noether's theorem to define conservation laws. In §3, we derive Lagrangian and conservation laws of a general competitive system for the $2n$ -dimensional ND system. Then we discuss that a symmetric and nonlinear dynamical system should exist in a form of $2n$ -dimensional nonlinear differential equations when velocity-independent conservation law exists. In §4, we will illustrate examples of conservation laws in two or three variables in order to explain the results discussed in §3. In §5, we shall discuss our results.

§2. Noether's theorem and conserved quantities

The general formulation of necessary condition for extrema in Lagrangian formulation, $\mathcal{L}(t, x^k(t), \dot{x}^k(t))$, is given by

$$\delta J = \delta \int \mathcal{L}(t, x^k(t), \dot{x}^k(t)) dt = 0, \quad (k = 1, \dots, n) \quad (2.1)$$

and all functions, $x(t) = (x^1(t), \dots, x^n(t))$, $t \in [a, b]$, belong to $C^2[a, b]$, which denotes the set of all continuous functions on the interval $[a, b]$ and the second derivatives of all functions are continuous. If $x(t)$ is a relative minimum of the functional J , the condition (2.1) generates,

$$E_k \equiv \frac{\partial \mathcal{L}}{\partial x^k} - \frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{x}^k} \right) = 0. \quad (2.2)$$

This is the Euler-Lagrange equation which determines equations of motion of a system.

Noether's theorem describes that the Lagrangian with Euler-Lagrange equation is invariant under certain space-time transformations, and invariances of Lagrangian generate respective conservation laws. For example, the time-translation invariance

of Lagrangian corresponds to the conservation of energy of a system. Let us consider r -parameter transformations in general that will be regarded as transformations of configuration space, (t, x^1, \dots, x^n) -space, depending upon r real, independent parameters $\varepsilon^1, \dots, \varepsilon^r$. The transformations are defined by

$$\begin{aligned}\bar{t} &= t + \tau_s(t, x)\varepsilon^s + o(\varepsilon), \\ \bar{x}^k &= x^k + \xi_s^k(t, x)\varepsilon^s + o(\varepsilon),\end{aligned}\tag{2.3}$$

where s ranges over $1, \dots, r$, and $o(\varepsilon)$ denotes the terms which go to zero faster than $|\varepsilon|$, $\lim_{|\varepsilon| \rightarrow 0} o(\varepsilon)/|\varepsilon| = 0$. The functions $\tau_s(t, x)$ and $\xi_s^k(t, x)$ for linear parts of \bar{t} and \bar{x}^k with respect to ε are commonly called the infinitesimal generators of the transformations. Classical theorem of Emmy Noether on invariant variational problems can be derived under the hypotheses of extrema (2.1), and transformation (2.3).¹⁹⁾ The result is the r identities produced by the transformation (2.3),

$$-E_k(\xi_s^k - \dot{x}^k \tau_s) = \frac{d}{dt} \left(\left(\mathcal{L} - \dot{x}^k \frac{\partial \mathcal{L}}{\partial \dot{x}^k} \right) \tau_s + \frac{\partial \mathcal{L}}{\partial \dot{x}^k} \xi_s^k - \Phi_s \right),\tag{2.4}$$

where $k = 1, \dots, n$ is summed, and Φ_s is an arbitrary function defined as a gauge function of the transformation. Note that the arbitrary choice of a gauge function will not change equations of motion of a system, which can be usually used for a convenient expression of conservation laws. If the fundamental integral (2.1) is divergence-invariant under the r parameter group of transformation (2.3), and if $E_k = 0$ for $k = 1, \dots, n$, then following r expressions are obtained:

$$\Psi_s \equiv \left(\mathcal{L} - \dot{x}^k \frac{\partial \mathcal{L}}{\partial \dot{x}^k} \right) \tau_s + \frac{\partial \mathcal{L}}{\partial \dot{x}^k} \xi_s^k - \Phi_s = \text{constant}.\tag{2.5}$$

This defines the conserved quantities of a system. Since the expressions Ψ defined in (2.5) are constant with the condition: $E_k = 0 (k = 1, \dots, n)$, they are the first integrals of the differential equations of motion, that is, the conserved quantities. In physical applications, the first integral (2.5) is interpreted as the energy of the system, whose governing equations are $E_k = 0$. In general, the expression Ψ is constant with respect to time and any extremal solutions of a system. By employing the formalism, the conservation law and symmetry of nonlinear dynamical differential equations are discussed in §3, and we will show the first integrals, the symmetry of Ψ and its solutions in an explicit coupled two-variables system in §4, as an example for the current approach.

§3. Conserved quantities and symmetric BCF in $2n$ -dimensional nonlinear dynamical (ND) systems

Nonlinear dynamical systems are well known in feed-back systems, such as biology, environmental sciences and computer systems, and they are important to understand many complicated interactive structures. Consider a nonlinear dynamical system that has $2n$ variables. The system has self-interactions, mixing interactions of quadratic forms of all combinations, such as $x_1^2, x_2^2, \dots, x_1 x_2, x_3 x_4, \dots$. In this

section, we will show the following. (1) We begin to examine a nonlinear dynamical system of $2n$ variables in symmetric form. The variables $(x_1, x_2, x_3, \dots, x_{2n})$ are, for example, $2n$ species of competing creatures in LV system,⁴⁾ or a cell-structured organism which cooperates together.⁵⁾ We will write down a quadratically interacting system as general as possible and investigate the properties of nonlinear interactions as a whole. (2) We will discuss the solution of the $2n$ -dimensional ND differential system with respect to the conservation law and Noether's theorem by calculating Ψ explicitly. (3) We will show that a coupled nonlinear dynamical system in BCF has a conserved quantity, or Hamiltonian of the BCF, which is velocity-independent.

The $2n$ -dimensional ND system in BCF having $2n$ variables (x_1, \dots, x_{2n}) is generally described as

$$\begin{aligned} \dot{x}_{2k-1} &= \sum_{i=1}^n (a_{2k-1,2i-1}x_{2i-1} + a_{2k-1,2i}x_{2i}) + \sum_{j=1}^{2n} a_{2k-1,2n+j}x_jx_{2k-1} \\ &\quad + \sum_{i=1}^{n'} a_{2k-1,4n+i}x_{2i-1}x_{2i}, \\ \dot{x}_{2k} &= \sum_{i=1}^n (a_{2k,2i-1}x_{2i-1} + a_{2k,2i}x_{2i}) + \sum_{j=1}^{2n} a_{2k,2n+j}x_jx_{2k} + \sum_{i=1}^{n'} a_{2k,4n+i}x_{2i-1}x_{2i}, \end{aligned} \tag{3.1}$$

where $k = 1, 2, \dots, n$. The first summation expresses the linear part of interaction which contains all variables x_1, \dots, x_{2n} . The second term is the competitive interaction expressed by coupled two variables, x_kx_1, \dots, x_kx_{2n} . This term is typical in classical LV equations. The third term with prime, $\sum_{i=1}^{n'} x_{2i-1}x_{2i}$, expresses all the mixing interactions other than $i = k$.

The Lagrangian of (3.1) is given by the following form,

$$\begin{aligned} \mathcal{L} &= \sum_{i=1}^n (\alpha_{2i-1}\dot{x}_{2i-1}x_{2i} + \alpha_{2i}x_{2i-1}\dot{x}_{2i}) + \sum_{i=1}^n \sum_{j=1}^{2n} \alpha_{\{2n+2n(i-1)+j\}}x_{2i-1}x_j \\ &\quad + \sum_{i=1}^n \sum_{j=1}^{2n} \alpha_{\{2n^2+2n+2n(i-1)+j\}}x_{2i}x_j + \sum_{i=1}^n \sum_{j=1}^{2n} \alpha_{\{4n^2+2n+2n(i-1)+j\}}x_{2i-1}x_{2i}x_j. \end{aligned} \tag{3.2}$$

Applying Euler-Lagrange equation (2.2) to Eq. (3.2), we can get ordinary differential equations; the ordinary differential equation for \dot{x}_{2k-1} is derived as

$$\begin{aligned} d_{2k,2k-1}\dot{x}_{2k-1} &= \sum_{i=1}^n (\alpha_{2n+2n(i-1)+2k} + \alpha_{2n^2+2n+2n(k-1)+2i-1})x_{2i-1} \\ &\quad + \sum_{i=1}^n (\alpha_{2n^2+2n+2n(i-1)+2k} + \alpha_{2n^2+2n+2n(k-1)+2i})x_{2i} \end{aligned}$$

$$+ \sum_{j=1}^{2n} \alpha_{4n^2+2n+2n(k-1)+j} x_j x_{2k-1} + \sum_{i=1}^n \alpha_{4n^2+2n+2n(i-1)+2k} x_{2i-1} x_{2i}, \quad (3.3)$$

and the ordinary differential equation for \dot{x}_{2k} is derived as

$$\begin{aligned} d_{2k-1,2k} \dot{x}_{2k} &= \sum_{i=1}^n (\alpha_{2n+2n(i-1)+2k-1} + \alpha_{2n+2n(k-1)+2i-1}) x_{2i-1} \\ &+ \sum_{i=1}^n (\alpha_{2n^2+2n+2n(i-1)+2k-1} + \alpha_{2n+2n(k-1)+2i}) x_{2i} \\ &+ \sum_{j=1}^{2n} \alpha_{4n^2+2n+2n(k-1)+j} x_j x_{2k} + \sum_{i=1}^n \alpha_{4n^2+2n+2n(i-1)+2k-1} x_{2i-1} x_{2i}, \end{aligned} \quad (3.4)$$

where $d_{2k,2k-1} = \alpha_{2k} - \alpha_{2k-1}$ and $d_{2k,2k-1} = -d_{2k-1,2k}$ for all k . The conditions of parameters can be obtained by comparing (3.1) and (3.3) when the coupled ND differential equations have a conserved quantity. The conditions of coefficients for \dot{x}_{2k-1} are given by

$$\begin{aligned} a_{2k-1,2i-1} &= \frac{1}{d_{2k,2k-1}} (\alpha_{2n+2n(i-1)+2k} + \alpha_{2n^2+2n+2n(k-1)+2i-1}), \\ a_{2k-1,2i} &= \frac{1}{d_{2k,2k-1}} (\alpha_{2n^2+2n+2n(i-1)+2k} + \alpha_{2n^2+2n+2n(k-1)+2i}), \\ a_{2k-1,2n+j} &= \begin{cases} \frac{2}{d_{2k,2k-1}} \alpha_{4n^2+2n+2n(k-1)+2k}, & (j = 2k) \\ \frac{1}{d_{2k,2k-1}} \alpha_{4n^2+2n+2n(k-1)+j}, & (j \neq 2k) \end{cases} \\ a_{2k-1,4n+i} &= \frac{1}{d_{2k,2k-1}} \alpha_{4n^2+2n+2n(i-1)+2k}, \quad (i \neq k) \end{aligned} \quad (3.5)$$

and also by comparing (3.1) and (3.4), we get conditions of parameters for \dot{x}_{2k} ND differential equations,

$$\begin{aligned} a_{2k,2i-1} &= \frac{1}{d_{2k-1,2k}} (\alpha_{2n+2n(i-1)+2k-1} + \alpha_{2n+2n(k-1)+2i-1}), \\ a_{2k,2i} &= \frac{1}{d_{2k-1,2k}} (\alpha_{2n^2+2n+2n(i-1)+2k-1} + \alpha_{2n+2n(k-1)+2i}), \\ a_{2k,2n+j} &= \begin{cases} \frac{2}{d_{2k-1,2k}} \alpha_{4n^2+2n+2n(k-1)+2k}, & (j = 2k - 1) \\ \frac{1}{d_{2k-1,2k}} \alpha_{4n^2+2n+2n(k-1)+j}, & (j \neq 2k - 1) \end{cases} \\ a_{2k,4n+i} &= \frac{1}{d_{2k-1,2k}} \alpha_{4n^2+2n+2n(i-1)+2k-1}. \quad (i \neq k) \end{aligned} \quad (3.6)$$

One should note that if the parameters satisfy these conditions, then the ND differential equations have a conserved quantity. The conserved quantity for time trans-

formation for the coupled ND equations is given by

$$\Psi \equiv \sum_{i=1}^n \sum_{j=1}^{2n} \alpha_{\{2n+2n(i-1)+j\}} x_{2i-1} x_j + \sum_{i=1}^n \sum_{j=1}^{2n} \alpha_{\{2n^2+2n+2n(i-1)+j\}} x_{2i} x_j + \sum_{i=1}^n \sum_{j=1}^{2n} \alpha_{\{4n^2+2n+2n(i-1)+j\}} x_{2i-1} x_{2i} x_j. \quad (3.7)$$

The conserved quantity, $\Psi(x_1, x_2, \dots, x_{2n})$, that produces the first order coupled nonlinear differential equations in BCF is conserved and constant with respect to time.

In this ND system, we define the stable solutions as follows. When one substitutes the solutions $(x_1, x_2, \dots, x_{2n})$ into $\Psi(x_1, x_2, \dots, x_{2n})$ obtained in certain time range and Ψ becomes strictly constant, we say that the solutions are stable in the time range. If Ψ is not constant, solutions are unstable or may not exist. In addition, one can conclude that the coefficients of the ND system are strictly constrained by the conservation law. This is also one of the important results of the conserved ND system. When the coefficients satisfy the relations (3.5) and (3.6), the ND differential equations have solutions that maintain the conserved quantity $\Psi(x_1, x_2, \dots, x_{2n})$. If the ND system does not meet conditions (3.5) and (3.6), Ψ may not be constant with respect to time. It will be shown explicitly in examples in §4.

However, because nonlinear coefficients can be assumed to take any values although they are restricted by some conditions, there exist cases that the general nonlinear dynamical system does not have solutions despite the fact that it has formally the conserved quantity Ψ . For this reason, we mainly concern ourselves to analyze cases in our BCF competitive system which has converged solutions and a conserved quantity Ψ .

The conserved quantity Ψ is invariant under exchanges of any variables $x_i \leftrightarrow x_j$;

$$\Psi(x_1, \dots, x_i, \dots, x_j, \dots, x_{2n}) = \Psi'(x_1, \dots, x_j, \dots, x_i, \dots, x_{2n}), \quad (3.8)$$

where $\Psi'(x_1, \dots, x_j, \dots, x_i, \dots, x_{2n})$ is obtained from Ψ by renaming of coefficients. Because of the symmetry, the addition of other BCF, for example, $\Psi(x_{2n+1}, x_{2n+2})$ assuming the similar nonlinear interaction to the whole system, changes only $n \rightarrow n + 1$ in (3.7), which means the conservation law of the system is unchanged. It indicates that the property of ND system may be kept unchanged even when the nonlinear differential equations are changed to as $n \rightarrow n + 1$; $\Psi(x_1, \dots, x_{2n}) = \text{constant}$ can be maintained as $\Psi(x_1, \dots, x_{2n}, x_{2n+1}, x_{2n+2}) = \text{constant}$, if the coefficients of nonlinear interactions of $\Psi(x_{2n+1}, x_{2n+2})$ with others are not extremely different. The condition (3.8) could be interpreted in terms of conservation law to maintain stability, homeostasis of a biological system.

The conserved quantity $\Psi(x_1, \dots, x_{2n}) = \text{constant}$, can be used to check the accuracy of numerical solutions to the $2n$ differential equations. Let us suppose that solutions to (3.1), x_1, \dots, x_{2n} are obtained. Then, one should substitute all solutions to (3.7) to check if Ψ becomes constant or not. In Figs. 1 and 2, the solutions to classical LV equation in the example of §4.2 are shown. The parameters of solutions

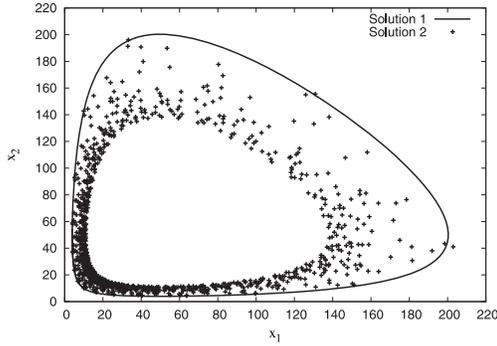


Fig. 1. Numerical simulations of classical LV equation. Solutions 1 and 2 have different increment.

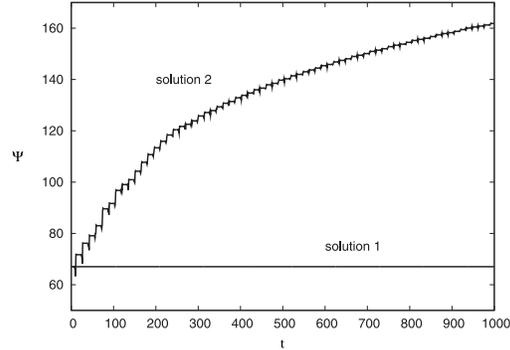


Fig. 2. Numerical simulation of conserved quantity Ψ . Solution 1 are constant but Solution 2 oscillates and increases with respect to time.

1 and 2 are $\alpha_1 = -0.02$, $\alpha_2 = -0.02$, $\alpha_3 = 50.0$, $\alpha_4 = 100.0$, $\alpha_5 = 1.0$, and initial values of solutions 1 and 2 are $x_1 = 10.0$, $x_2 = 10.0$ (see (4.12)). The size of interval ΔH in Runge-Kutta methods of solution 1 is $\Delta H = 0.01$, and solution 2 is $\Delta H = 0.5$.

If Ψ is exactly constant in the assumed interval $t \in [a, b]$, the solutions, x_1, \dots, x_{2n} , are expected to be exact. If $\Psi \approx \text{constant}$, x_1, \dots, x_{2n} are approximate solutions that depend on the accuracy of $\Psi = \text{constant}$.

§4. Examples

The conservation law (3.7) in the BCF is a velocity-independent form. In the work of J. Fernández-Núñez, some functions are velocity-dependent and coupled with derivative terms such as $\dot{x}_{2i} \log x_{2i-1}/x_{2i}$ and $\dot{x}_{2i-1} \log x_{2i}/x_{2i-1}$ to obtain the classical LV system, and we also found a velocity-independent form of the exponential type, $\exp\{f(x_1, x_2, \dots)\}$, shown in the example in §4.2.

4.1. 2-variable system

Here, we show an example of $2n$ -dimensional ND system in the case of $n = 1$ in (3.1). The ND system of (3.1) is defined by

$$\begin{aligned} \dot{x}_1 &= a_{11}x_1 + a_{12}x_2 + a_{13}x_1^2 + a_{14}x_1x_2, \\ \dot{x}_2 &= a_{21}x_1 + a_{22}x_2 + a_{23}x_1x_2 + a_{24}x_2^2. \end{aligned} \tag{4.1}$$

The mixing interactions and self-interactions are expressed as x_1x_2 , x_1^2 and x_2^2 . Since this system has two variables, we have only x_1x_2 mixing interaction. The Lagrangian of this system is described from (3.2) as

$$\begin{aligned} \mathcal{L} &= \alpha_1 \dot{x}_1x_2 + \alpha_2 x_1\dot{x}_2 + \alpha_3 x_1^2 + (\alpha_4 + \alpha_5)x_1x_2 \\ &\quad + \alpha_6 x_2^2 + \alpha_7 x_1^2x_2 + \alpha_8 x_1x_2^2. \end{aligned} \tag{4.2}$$

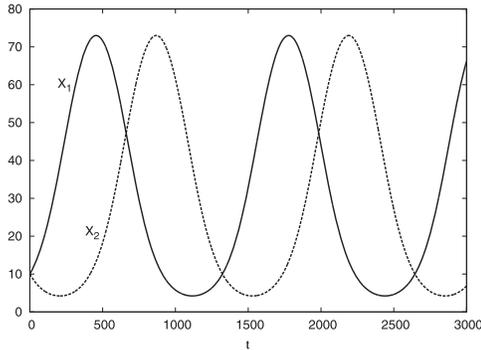


Fig. 3. Numerical simulation of 2-variable ND 1. The parameters of 2-variable ND 1: $\alpha_1 = 1.0, \alpha_2 = 2.0, \alpha_3 = -0.01, \alpha_4 = 0.5, \alpha_5 = 0.5, \alpha_6 = -0.01, \alpha_7 = -0.01, \alpha_8 = -0.01$, initial values are $x_1 = 10.0, x_2 = 10.0$.

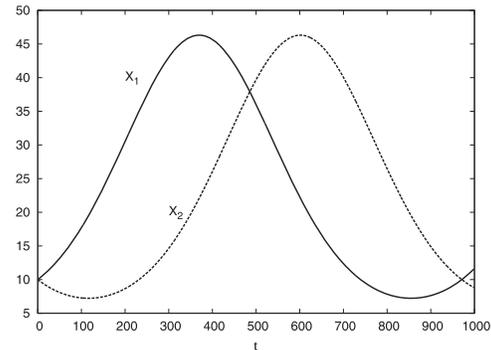


Fig. 4. Numerical simulation of 2-variable ND 2. The parameters of 2-variable ND 2: $\alpha_1 = 1.0, \alpha_2 = 2.0, \alpha_3 = -0.1, \alpha_4 = 0.5, \alpha_5 = 0.5, \alpha_6 = -0.1, \alpha_7 = -0.01, \alpha_8 = -0.01$, initial values are $x_1 = 10.0, x_2 = 10.0$.

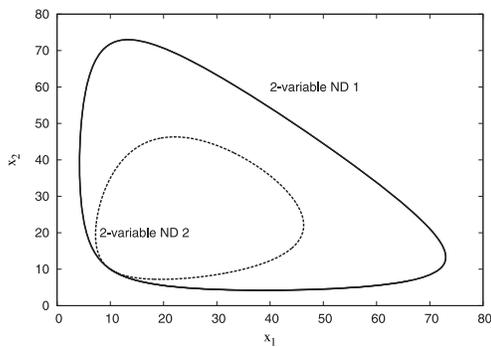


Fig. 5. Numerical solutions of 2-variable ND 1 and ND 2.

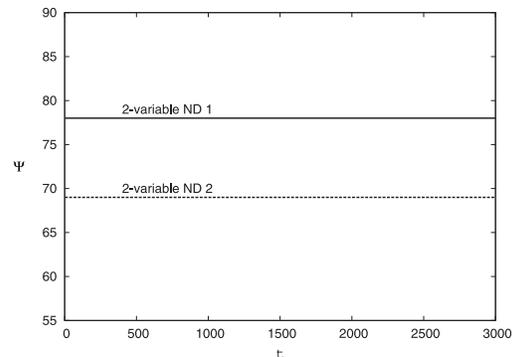


Fig. 6. Numerical solutions of conserved quantity Ψ of 2-variable ND 1 and ND 2. Note that Ψ is constant in both cases.

From (4.2), we get the following nonlinear differential equation,

$$\dot{x}_1 = \frac{1}{d_{21}} \{(\alpha_4 + \alpha_5)x_1 + 2\alpha_6x_2 + 2\alpha_8x_1x_2 + \alpha_7x_1^2\}, \tag{4.3}$$

$$\dot{x}_2 = \frac{1}{d_{12}} \{2\alpha_3x_1 + (\alpha_4 + \alpha_5)x_2 + 2\alpha_7x_1x_2 + \alpha_8x_2^2\}. \tag{4.4}$$

The parameter d_{21} is given by $d_{21} = \alpha_2 - \alpha_1 = -d_{12}$. The conserved quantity Ψ of this system is obtained as follows:

$$\Psi \equiv \alpha_3x_1^2 + (\alpha_4 + \alpha_5)x_1x_2 + \alpha_6x_2^2 + \alpha_7x_1^2x_2 + \alpha_8x_1x_2^2. \tag{4.5}$$

We show numerical simulations of 2-variable ND system in Figs. 3–6. Note that 2-variable ND 1 in Fig. 3 and 2-variable ND 2 in Fig. 4 with different parameters are periodically stable and have the same conserved quantity Ψ defined as (4.5). The parameters of both 2-variable ND 1 and ND 2 are given in the caption of Figs. 3

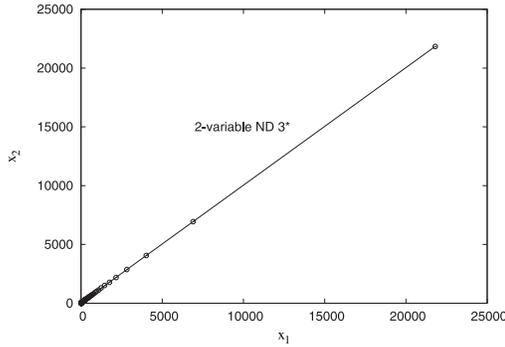


Fig. 7. Numerical solutions of 2-variable ND 3* system with parameters which do not satisfy conditions.

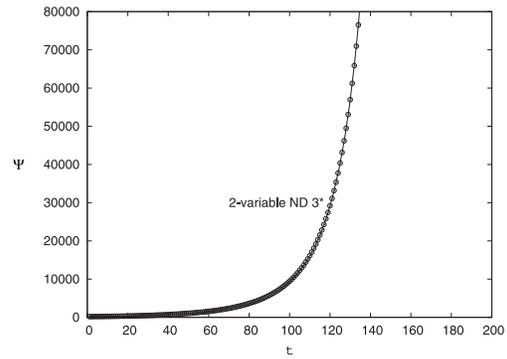


Fig. 8. Numerical solutions of Ψ of 2-variable ND 3*. Note that the solutions are not constant and diverge with respect to time.

and 4. The solutions of 2-variable ND 1 and ND 2 are periodic and steady state in Fig.5, and the conserved quantity defined in (4.5) is given in Fig. 6. The conserved quantity of 2-variable ND 1 and ND 2 is strictly constant with respect to time.

In Figs. 7 and 8, we show an example in the case that parameters of $2n$ -ND system (3.1) do not satisfy the conditions (3.5) and (3.6). The conserved quantity Ψ is not equal to a constant in time. In the simulation of 2-variable ND 3* in Fig. 7, the coefficient of x_1^2 in (4.3) changed from α_7 to $-\alpha_7$, and the coefficient d_{12} in (4.4) changed from d_{12} to d_{21} and the coefficient x_2^2 in (4.4) also changed from α_8 to $-\alpha_8$ so that the system of differential equations does not satisfy the conditions (3.5), (3.6) and the conservation law (4.5). The parameters are determined as $\alpha_1 = 1.0$, $\alpha_2 = 2.0$, $\alpha_3 = 0.5$, $\alpha_4 = 0.5$, $\alpha_5 = 0.5$, $\alpha_6 = 0.1$, $\alpha_7 = 0.01$, $\alpha_8 = 0.01$, and initial values are $x_1 = 10.0$, $x_2 = 10.0$. In the simulation 8, quantity Ψ increased exponentially and diverged. Ψ is not constant with respect to time in this case. It indicates that Ψ is not constant when the coefficients do not meet the conditions (3.5) and (3.6), and Ψ diverged when the solutions of the system diverged with respect to time.

Next, let us assume that the 3-variable system has a similar conservation law as (4.5), and we obtain the Lagrangian of the following type,

$$\begin{aligned} \mathcal{L} = & \alpha_1 \dot{x}_1 x_2 + \alpha_2 \dot{x}_2 x_3 + \alpha_3 \dot{x}_3 x_1 + \alpha_4 x_1 x_2 + \alpha_5 x_1 x_3 \\ & + \alpha_6 x_2 x_3 + \alpha_7 x_1^2 + \alpha_8 x_2^2 + \alpha_9 x_3^2 + \alpha_{10} x_1^2 x_2 \\ & + \alpha_{11} x_1 x_2^2 + \alpha_{12} x_1^2 x_3 + \alpha_{13} x_1 x_3^2 + \alpha_{14} x_2^2 x_3 + \alpha_{15} x_2 x_3^2. \end{aligned} \tag{4.6}$$

The conserved quantity of (4.6) is given by

$$\begin{aligned} \Psi = & \alpha_4 x_1 x_2 + \alpha_5 x_1 x_3 + \alpha_6 x_2 x_3 + \alpha_7 x_1^2 + \alpha_8 x_2^2 + \alpha_9 x_3^2 \\ & + \alpha_{10} x_1^2 x_2 + \alpha_{11} x_1 x_2^2 + \alpha_{12} x_1^2 x_3 + \alpha_{13} x_1 x_3^2 + \alpha_{14} x_2^2 x_3 + \alpha_{15} x_2 x_3^2. \end{aligned} \tag{4.7}$$

The Lagrangian (4.6) produces, the following ordinary differential equations,

$$\begin{aligned}
 \alpha_1 \dot{x}_2 - \alpha_3 \dot{x}_3 &= 2\alpha_7 x_1 + \alpha_4 x_2 + \alpha_5 x_3 + 2\alpha_{10} x_1 x_2 + \alpha_{11} x_2^2 \\
 &\quad + 2\alpha_{12} x_1 x_3 + \alpha_{13} x_3^2, \\
 \alpha_2 \dot{x}_3 - \alpha_1 \dot{x}_1 &= \alpha_4 x_1 + 2\alpha_8 x_2 + \alpha_6 x_3 + 2\alpha_{11} x_1 x_2 + 2\alpha_{14} x_2 x_3 \\
 &\quad + \alpha_{10} x_1^2 + \alpha_{15} x_3^2, \\
 \alpha_3 \dot{x}_1 - \alpha_2 \dot{x}_2 &= \alpha_5 x_1 + \alpha_6 x_2 + 2\alpha_9 x_3 + \alpha_{12} x_1^2 + 2\alpha_{13} x_1 x_3 \\
 &\quad + \alpha_{14} x_2^2 + 2\alpha_{15} x_2 x_3.
 \end{aligned}
 \tag{4.8}$$

The type of differential equations is a little different from that of the $2n$ variables. The coupled linear equation (4.8) can be solved for x_3 , resulting in

$$\begin{aligned}
 (\alpha_2 \alpha_{13} + \alpha_3 \alpha_{15}) x_3^2 &+ \{(2\alpha_1 \alpha_{15} + 2\alpha_3 \alpha_{14}) x_2 + (2\alpha_1 \alpha_{13} + 2\alpha_2 \alpha_{12}) x_1 \\
 &\quad + (2\alpha_1 \alpha_9 + \alpha_2 \alpha_5 + \alpha_3 \alpha_6)\} x_3 + (\alpha_1 \alpha_5 + 2\alpha_2 \alpha_7 + \alpha_3 \alpha_4) x_1 \\
 &\quad + (\alpha_1 \alpha_6 + \alpha_2 \alpha_4 + 2\alpha_3 \alpha_8) x_2 + (\alpha_1 \alpha_{12} + \alpha_3 \alpha_{10}) x_1^2 \\
 &\quad + (2\alpha_3 \alpha_{11} + 2\alpha_2 \alpha_{10}) x_1 x_2 + (\alpha_1 \alpha_{14} + \alpha_2 \alpha_{11}) x_2^2 = 0.
 \end{aligned}
 \tag{4.9}$$

This is another time-independent relation of the 3-variable ND system. However, the time-independent relation is obtained from the differential equations, not from the conservation law Ψ . Hence, it indicates that there are two types of time-independent relations: (1) the time-independent relation that is strictly determined by Lagrangian or Noether’s theorem; (2) the time-independent relation that is derived from a particular structure of differential equations, but is not necessarily related to the conservation law. The conserved quantity, Ψ , is strictly constant with respect to time, which is more general than the relation (4.9).

4.2. Exponential types of 2 variable coupled LV system

We show another type of conserved quantity of 2 variables which is similar to Lyapunov function. The classical type of LV equation is defined as

$$\begin{aligned}
 \dot{x}_1 &= a_{11} x_1 + a_{12} x_1 x_2, \\
 \dot{x}_2 &= a_{21} x_2 + a_{22} x_1 x_2.
 \end{aligned}
 \tag{4.10}$$

x_1 and x_2 are interpreted as a prey and a predator, respectively. The Lagrangian of (4.10) is described as follows:

$$\mathcal{L} = \exp\{\alpha_1 x_1 + \alpha_2 x_2\} \{\alpha_3 \dot{x}_1 + \alpha_4 \dot{x}_2 + \alpha_5 x_1 x_2\}.
 \tag{4.11}$$

We can get the conditions of parameter from (4.10) and (4.11),

$$\begin{aligned}
 \dot{x}_1 &= \frac{1}{d_{12}} \alpha_5 \{\alpha_2 x_1 x_2 + x_1\}, \\
 \dot{x}_2 &= \frac{1}{d_{21}} \alpha_5 \{\alpha_1 x_1 x_2 + x_2\},
 \end{aligned}
 \tag{4.12}$$

where $d_{12} = \alpha_1\alpha_4 - \alpha_2\alpha_3 = -d_{21}$. This is the condition for parameters to have a conserved quantity. The conserved quantity Ψ is obtained as

$$\Psi \equiv \alpha_5 \exp\{\alpha_1 x_1 + \alpha_2 x_2\} x_1 x_2. \quad (4.13)$$

The logarithm of (4.13) is similar to a well known Lyapunov function $V(x_1, x_2)$ which has the property $dV(x_1, x_2)/dt = 0$. It can be directly explained that the time derivative of conserved quantity should vanish, $d\Psi/dt = 0$, and therefore, Lyapunov function has a strong relation with Noether's theorem.

One can see that the conservation law and Lyapunov function are explicitly related to each other in the case of this simple example. However, it is not clear whether the connections between the $2n$ -dimensional ND system and Lyapunov function are directly related to each other. Let us assume 3-variable type of differential equations of exponential type. The Lagrangian of linear 3-variable first-order equation is assumed as

$$\begin{aligned} \mathcal{L} = \exp\{\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3\} \{ & \alpha_4 \dot{x}_1 + \alpha_5 \dot{x}_2 + \alpha_6 \dot{x}_3 + \alpha_7 x_1 + \alpha_8 x_2 \\ & + \alpha_9 x_3 + \alpha_{10} x_1 x_2 + \alpha_{11} x_1 x_3 + \alpha_{12} x_2 x_3 \}. \end{aligned} \quad (4.14)$$

Applying Eq. (2.2) to Eq. (4.16), we get three equations as

$$\begin{aligned} & (\alpha_1\alpha_5 - \alpha_2\alpha_4)\dot{x}_2 + (\alpha_1\alpha_6 - \alpha_3\alpha_4)\dot{x}_3 + \alpha_1\alpha_7 x_1 + (\alpha_1\alpha_8 + \alpha_{10})x_2 \\ & + (\alpha_1\alpha_9 + \alpha_{11})x_3 + \alpha_1\alpha_{10}x_1x_2 + \alpha_1\alpha_{11}x_1x_3 + \alpha_1\alpha_{12}x_2x_3 + \alpha_7 = 0, \\ & (\alpha_2\alpha_4 - \alpha_1\alpha_5)\dot{x}_1 + (\alpha_2\alpha_6 - \alpha_3\alpha_5)\dot{x}_3 + (\alpha_2\alpha_7 + \alpha_{10})x_1 + \alpha_2\alpha_8 x_2 \\ & + (\alpha_2\alpha_9 + \alpha_{12})x_3 + \alpha_2\alpha_{10}x_1x_2 + \alpha_2\alpha_{11}x_1x_3 + \alpha_2\alpha_{12}x_2x_3 + \alpha_8 = 0, \\ & (\alpha_3\alpha_4 - \alpha_1\alpha_6)\dot{x}_1 + (\alpha_3\alpha_5 - \alpha_2\alpha_6)\dot{x}_2 + (\alpha_3\alpha_7 + \alpha_{11})x_1 + (\alpha_3\alpha_8 + \alpha_{12})x_2 \\ & + \alpha_3\alpha_9 x_3 + \alpha_3\alpha_{10}x_1x_2 + \alpha_3\alpha_{11}x_1x_3 + \alpha_3\alpha_{12}x_2x_3 + \alpha_9 = 0. \end{aligned} \quad (4.15)$$

The conserved quantity is velocity-independent and given by

$$\begin{aligned} \Psi = \exp\{\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3\} \{ & \alpha_7 x_1 + \alpha_8 x_2 + \alpha_9 x_3 + \alpha_{10} x_1 x_2 \\ & + \alpha_{11} x_1 x_3 + \alpha_{12} x_2 x_3 \}. \end{aligned} \quad (4.16)$$

The nonlinear equation (4.15) can be solved for x_3 , resulting in

$$\begin{aligned} x_3 = \frac{1}{\sigma_1 + \sigma_2 x_1 + \sigma_3 x_2} [& \{\phi_3 \alpha_1 \alpha_7 + \phi_1 (\alpha_2 \alpha_7 + \alpha_{10}) + \phi_2 (\alpha_3 \alpha_7 + \alpha_{11})\} x_1 \\ & + \{\phi_3 (\alpha_1 \alpha_8 + \alpha_{10}) + \phi_1 \alpha_2 \alpha_8 + \alpha_2 (\alpha_3 \alpha_8 + \alpha_{12})\} x_2 \\ & + \{\phi_3 \alpha_1 \alpha_{10} + \phi_1 \alpha_2 \alpha_{10} + \phi_2 \alpha_3 \alpha_{10}\} x_1 x_2 + \phi_3 \alpha_7 + \phi_1 \alpha_8 + \phi_2 \alpha_9], \end{aligned} \quad (4.17)$$

where $\phi_1 = \alpha_3\alpha_4 - \alpha_1\alpha_6$, $\phi_2 = \alpha_1\alpha_5 - \alpha_2\alpha_4$, $\phi_3 = \alpha_2\alpha_6 - \alpha_3\alpha_5$, $\sigma_1 = -\phi_3(\alpha_1\alpha_9 + \alpha_{11}) - \phi_1(\alpha_2\alpha_9 + \alpha_{12}) - \phi_2\alpha_3\alpha_9$, $\sigma_2 = -\phi_3\alpha_1\alpha_{11} - \phi_1\alpha_2\alpha_{11} - \phi_2\alpha_3\alpha_{11}$, $\sigma_3 = -\phi_3\alpha_1\alpha_{12} - \phi_1\alpha_2\alpha_{12} - \phi_2\alpha_3\alpha_{12}$. As discussed in the example 4.1, the conserved quantity is strictly constant with respect to time, which is more general than Eq. (4.17).

§5. Conclusion

In this paper, we have investigated a system of $2n$ -dimensional, coupled first-order differential equations which contains self-interactions, and mixing interactions by using Noether's theorem. We discussed that the nonlinear differential equations with a conserved quantity, Ψ , calculated by Noether's theorem have converged stable solutions, and the coefficients of nonlinear interactions are strictly confined by the conservation law of the system. The conserving converged solutions are shown by a closed curve in (x_1, x_2) -coordinates for 2-dimensional case, and in general, it should be discussed by a closed hyper-surface in $(x_1, x_2, \dots, x_{2n})$ -coordinates for $2n$ -dimensional case.

Conventionally, the analysis of conserved quantity of solutions to a system of coupled differential equations is discussed with Lyapunov function.²⁰⁾ There are two main types of Lyapunov functions that are strict Lyapunov and non-strict Lyapunov functions. Our conserved quantity would correspond to the strict Lyapunov function, due to the global property of Lagrangian approach. The theorem relates stability and conserved quantity analogous to conservation laws of energy and momentum in physics, hence it is helpful to understand a system of nonlinear differential equations in view of scientific or physical terms.

A system of symmetric nonlinear coupled first order differential equations has a conservation law and exists in the form of the $2n$ -independent variables $(x_1, x_2, \dots, x_{2n})$. We termed the system of symmetric first-order differential equations composed of the $2n$ -dimensional nonlinear interactions as the binary-coupled form (BCF). If a coupled nonlinear system exists in a form of symmetric first-order differential equations, which has a velocity-independent conserved quantity as discussed in the paper, the system tends to be composed of the binary-coupled form $(2, 4, 6, \dots, 2n)$. The predator-prey type system (2-coupled form), food-web system,^{21), 22)} and gene regulation network system^{23), 24)} may be typical examples, and also the computer network seems to be expressed by $2n$ -coupled form.²⁵⁾ If a competitive system is the odd-variable type, the system of the first-order differential equations becomes little different as shown in the example in §4.2.

When a binary-coupled system with a conservation law $\Psi(x_1, \dots, x_{2n})$ and another system with $\Psi(x_{2n+1}, x_{2n+2})$ interact with each other and result in constructing a new binary-coupled system, the system should have the conservation law in the form of $\Psi(x_1, \dots, x_{2n}) + \Psi(x_{2n+1}, x_{2n+2}) \rightarrow \Psi(x_1, \dots, x_{2n+2})$ as in the form (3.7), even if they are nonlinearly interacting as (3.1). This may indicate an addition law for the same conserving ND systems. The addition law may be interpreted as the restoration or rehabilitation phenomena known in a large system of neural network or computer network when a small disordered device or part of a large network system is replaced by a normal device.

The conservation law is also used to check the accuracy of numerical solutions of nonlinear differential equations. The binary-coupled differential equations have interesting properties as shown in this paper, and the system of BCF $(2, 4, 6, \dots, 2n)$ seems to be found in social, natural and biological sciences. If a system of BCF is naturally discovered in nature, the consequences discussed in BCF system would

help understand structures, interactions, evolution mechanism of biological systems as well as social, economic and environmental systems.

We have investigated and discussed the property of $2n$ -dimensional ND system in general. Although $2n$ -dimensional ND system is discussed as a continuous system by employing differential equations, it is useful to examine their qualitative character of the classical LV system. However, most of biological phenomena may be mesoscopic, and a discrete system would fit actual phenomena more than a continuous system. In the future work, we would like to extend our method to discrete systems²⁶⁾ and examine the connection between the pattern and time scale of biological or mesoscopic phenomena and gene-related data sets. We understand that the conservation law Ψ depends sensitively on the values of nonlinear coefficients, and we also hope to investigate the relations between coefficients and stability of solutions of $2n$ -ND systems.

Acknowledgements

We would like to thank Professor Osamu Gotoh and Assistant Professor Natsuhiko Ichinose for introducing us this area of research during a course they provided us at Kyoto University as well as for discussions at different stages of this study.

References

- 1) S. A. Kauffman, *The Origins of Order* (Oxford University Press, New York, 1993).
- 2) I. Prigogine and I. Stengers, *Order Out of Chaos* (Bantam, 1984).
- 3) R. Fujie and T. Odagaki, *Physica A* **389** (2010), 1471.
- 4) T. Shimada, S. Yukawa and N. Ito, *Artificial Life and Robotics* **6** (2002), 78.
- 5) G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems* (John Wiley & Sons, New York, 1977).
- 6) A. Lotka, *Elements of Physical Biology* (Williams and Wilkins, Baltimore, 1925).
- 7) M. Taniguchi, M. Bando and A. Nakayama, *J. Phys. Soc. Jpn.* **77** (2008), 114001.
- 8) E. González and M. Villena, *Euro. J. of Operational Research* **210** (2011), 707.
- 9) V. Volterra, "Variazioni e fluttuazioni del numero di individui in specie animali conviventi", *Mem. Acad. Linc.* **2** (1926), 31.
- 10) N. H. Ibragimov and T. Kolsrud, *Nonlinear Dynamics* **36** (2004), 29.
- 11) N. M. Ivanova, *Nonlinear Dynamics* **49** (2007), 71.
- 12) A. H. Kara and F. M. Mahomed, *Nonlinear Dynamics* **45** (2006), 367.
- 13) J. Fernández-Núñez, *Int. J. Theor. Phys.* **37** (1998), 2457.
- 14) A. G. Johnpillai, A. H. Kara and F. M. Mahomed, *J. Comput. Appl. Math.* **223** (2009), 508.
- 15) R. Cherniha and V. Davydovych, *Mathematical and Computer Modelling* **54** (2011), 1238.
- 16) F. X. Mei, *Acta Mechanica* **141** (2000), 135.
- 17) K. Constandinides and P. A. Damianou, *Regular and Chaotic Dynamics* **16** (2011), 311.
- 18) M. A. Tavel's English translation of 'Invariant Variation Problemes' *Nachr. d. König. Gesellsch. d. Wiss. zu Göttingen, Math-phys. Klasse* (1918), 235.
- 19) J. D. Logan, *Invariant Variational Principles* (Academic Press, New York, 1977).
- 20) F. Mazenc and M. Malisoff, *IEEE Trans. Automatic Control* **55** (2010), 841.
- 21) I. Ispolatov and M. Doebeli, *Theor. Ecol.* **4** (2011), 55.
- 22) Y. Murase, T. Shimada and N. Ito, *New J. Phys.* **12** (2010), 063021.
- 23) A. Mochizuki, *J. of Theoretical Biology* **250** (2008), 307.
- 24) L. Ironi et al., *Physica D* **240** (2011), 779.
- 25) L. Qian, E. Winfree and J. Bruck, *Nature* **475** (2011), 368.
- 26) J. D. Logan, *Aequationes Mathematicae* **9** (1973), 210.