

A Comparative Study of Seasonally Adjusted and Unadjusted Economic Data by the Correlation Integral Method

TOSHIO INABA*, YOSHINORI NAGAI** and HIROSHI WAKO***

(Received 14 Jan. 1997)

Abstract: The correlation integral method was applied to six kinds of economic variables, for each of which six types of data (trended/detrended and unadjusted/seasonally-adjusted/moving-averaged ones) were prepared to calculate their correlation dimensions. The differences in the correlation dimensions calculated for different types of data of a given economic variable are assumed to reflect its dynamics. From a comparison of the correlation dimensions calculated, we found that the six kinds of economic variables can be classified into three classes: that is, finite correlation dimensions are obtained (a) for five of the six types of data, but not for trended unadjusted data (e.g., GNP and private equipment investment), (b) for four of them, but not for detrended seasonally-adjusted and detrended moving-averaged data (e.g., exports and private housing investment; private final consumption expenditure is likely to be classified into this class), and (c) only for trended moving-averaged data (e.g., imports). The implication of this classification (e.g. which is essential to dynamics of a given economic variable, a total amount or a growth rate?) is discussed.

1. Introduction

Recently, an economic system has been investigated from a viewpoint of deterministic chaos to explore some dynamical aspects in it (Benhabib and Day, 1981; Day, 1982; Grandmont, 1985; Brock, 1986; Chen, 1988; Owase, 1991; see also the references in Table I). One of the interests in such an approach is to seek determinisity in economic data with erratic behavior rather than to regard them as a stochastic process. Some researchers suggested that they found some evidences of deterministic chaos in economic data, whereas others did not. The difficulty to verify the determinisity arises mainly from small-sized data. Novel methods have been proposed to overcome it (Barnett, et al., 1992; Barnett and Hinich, 1993), besides the method with correlation dimension (the references in Table I), Kolmogorov entropy (Frank and Stengos, 1989), and Lyapunov exponent (Brock, 1986; Barnett and Chen, 1988). However, whether or not there exists the determinisity inherent in economic data is still controversial.

Here, we study Japanese economic data using the correlation integral method proposed by Grassberger and Procaccia (1983a, b). Recent studies on economic variables with the correlation integral method by various authors are summarized in Table I. In contrast to those

* School of Education, Waseda University

** Center for Information Science and School of Political and Economic Sciences, Kokushikan University

*** School of Social Sciences, Waseda University

Table I Correlation dimension estimated for economic data by various authors

Economic variable	Author	Correlation dimension ^{a)}	Country examined
GNP	Brock (1988)	ca. 3 ($d=20$)	U.S.A.
	Frank & Stengos (1988)	3.6($d=20$)	Canada
	Frank et. al (1988)	not saturated	Italy, Japan, U.K., West Germany
	Tanaka (1993)	3.3 ($d=15$)	U.S.A
		3.6 ($d=15$)	West Germany
		2.3 ($d=15$)	Japan
Unemployment rate	Brock & Sayers (1988)	2.5–3.5 ($d=20$)	U.S.A
	Frank & Stengos (1988)	not saturated	Canada
	Tanaka (1993)	4.3 ($d=15$)	U.S.A
		3.3 ($d=15$)	West Germany
		2.5 ($d=15$)	Japan
Gold/silver returns	Frank & Stengos (1989)	6–7 ($d=25$)	U.K.
Stock return rates	Scheinkman & LeBaron (1989)	5.7 ($d=13, 14$)	U.S.A.
Stock price	Peters (1991)	2.33 ($d=7, 8$)	U.S.A
Money supply	Barnett & Chen (1988)	1.5($d=6$)	U.S.A.

^{a)} d in the parenthesis an embedding dimension, at which correlation dimension is determined.

studies, we do not have much concern for verifying the determinisity of economic data here. Our interest is in classifying the economic variables with respect to their dynamics. From such a purpose we examined trended/detrended and unadjusted/seasonally-adjusted/moving-averaged data for each of six kinds of economic variables. It is usual to investigate seasonally-adjusted data, when the original data have seasonality. However, processing the data may change their dynamics to some extent. The changes to occur in the individual variables are considered to reflect their own dynamics. Although it is difficult to reveal their dynamics in an explicit form, the changes observed in the analyses of the processed data are expected to provide some useful information to characterize the economic variables with respect to their dynamics.

In the next section the correlation integral method is explained briefly. The correlation dimensions are calculated for trended/detrended and unadjusted/seasonally-adjusted/moving-averaged data for the six kinds of economic variables of Japan in section 3. From a comparison of the correlation dimensions calculated, it is shown that the six kinds of economic variables can be classified into three classes. Finally, the implication of the results is discussed in section 4.

2. Method

2.1 Correlation integral method

In this study we use the correlation integral method proposed by Grassberger and Procaccia (1983a, b) (referred to as the G-P method hereinafter) to analyze economic time

series data. Here, we summarize the method.

Let $\{x_0, x_1, \dots, x_n\}$ be a one-dimensional time series. From this primary one-dimensional data, we define a d -dimensional vector ζ_i as

$$\zeta_i = \{x_i, x_{i+\tau}, x_{i+2\tau}, \dots, x_{i+(d-1)\tau}\} \quad (2.1)$$

d is called embedding dimension. The time delay τ is set to one throughout this study. Then we consider a d -dimensional vector time series

$$\{\zeta_0, \zeta_1, \dots, \zeta_m\} \quad (2.2)$$

where $m = n - d + 1$.

The first step of the G-P method is to calculate a correlation integral $C(\varepsilon)$ for the data embedded in d dimensions by counting the number of pairs of elements of the time series (2.2) which are separated by a distance smaller than a given distance ε ; i.e.,

$$C(\varepsilon) = \frac{2}{m(m-1)} \sum_{i < j} \Theta(\varepsilon - |\zeta_i - \zeta_j|) \quad (2.3)$$

where Θ and $|\zeta_i - \zeta_j|$ are a Heaviside step function (i.e., $\Theta(y) = 0$ if $y < 0$, and 1 otherwise) and a distance between ζ_i and ζ_j , respectively. The number of pairs separated by a distance smaller than ε is normalized by the total number of pairs in eq. (2.3).

If there exists a chaotic attractor, the correlation integral $C(\varepsilon)$ increases at a rate of ε^{D_2} for a given embedding dimension d over the range of smaller ε (To the contrary, in the range of ε larger than a maximum distance, $C(\varepsilon)$ is equal to one) (Schuster, 1988). This is ideally expressed in the following scaling relation:

$$D_2 = \lim_{\varepsilon \rightarrow 0} \log C(\varepsilon) / \log \varepsilon \quad (2.4)$$

Equation (2.4) states that D_2 can be obtained as a linear slope of $\log C(\varepsilon)$ versus $\log \varepsilon$ plot (the Grassberger-Procaccia plot, abbreviated as the G-P plot hereinafter), if linear regions exist in the plot. The existence of the linear regions implies the fractal structure of the chaotic attractor. D_2 is a function of d . If D_2 converges to some finite values D_2^* as the embedding dimension d increases, D_2^* is called as correlation dimension of the system.

In practice, it is not always easy to define the region to measure slopes in the G-P plot and to judge convergence of D_2 to D_2^* as d increases, when we treat real economic data. The correlation dimension of a system is determined according to the following criteria in this study.

First of all, essentially straight regions have to be detected in the G-P plot: The straightness of individual regions is assessed by the correlation coefficients of the regression lines. Usually the two extreme regions of ε are omitted because of errors arising from smallness of available data. If there is more than one linear region in between and then more than one correlation dimension is obtained, one of them is selected.

Before explaining the selection rule, let us describe how to obtain a correlation dimension for a given region of ε . The slopes D_2 are plotted against embedding dimension d (referred to as a D_2 - d plot hereinafter) for the given region of ε . The feature of the D_2 - d plot depends on the region where D_2 is measured. Either of the two cases occurs in general. In the

first case the slope becomes almost constant for d larger than a certain embedding dimension d_1 . In some plots the slope that saturates at d_1 resumes increasing at a certain dimension d_2 ($> d_1$). Such a case is also assumed to be included in this case. A correlation dimension D_2^* is defined for the first case as a mean value of D_2 over the plateau region in the D_2 - d plot. In the second case the slope D_2 diverges without any plateau region as d increases. In some D_2 - d plots D_2 increases so slowly that the difference from a plateau is subtle. Including the slowly-increasing case, we regard that the system has no correlation dimension for the second case.

If more than one correlation dimension is obtained from different regions of ε in a G-P plot, we give the first priority to the D_2 - d plot in which D_2 is almost constant for d larger than a certain embedding dimension. If there still remains more than one correlation dimension, the one measured in the region of the smallest distance ε (lower left region in the G-P plot; see Fig. 2) is taken in principle. This criterion reflects the fact that the correlation dimension is defined as the limit of small ε in a strict sense as shown in eq. (2.4). However, we face a dilemma, because the region of very small ε may have errors arising from shortage of data, and because we cannot reject a possibility that a multiple correlation dimension is an essential nature of some economic systems. Consequently, it is hard to determine proper linear regions in the G-P plot thoroughly automatically. A visual inspection and subjective judgement are inevitable in some cases. It means that there still remains ambiguity in the determination of the correlation dimension. Accordingly, we will confine ourselves mainly to qualitative discussion below.

2.2 Economic variables examined

The six kinds of economic variables, GNP, private equipment investment (PEI), private housing investment (PHI), private final consumption expenditure (PFCE), exports (EXP), and imports (IMP) in Japan, are examined in this study.

The quarterly time series data from the first quarter of 1955 to the first quarter of 1989 (137 data points each) were obtained from "Annual Report on National Accounts" by the Economic Agency, Government of Japan.

For a comparative study, two more types of processed data were generated besides the original trended data. They were obtained by seasonally adjusting by Census X-11 and by taking a moving average over six quarters. In addition percentage changes from a previous quarter were calculated for each of the three types of data. As a result we have the six types of data for each economic variable, which are referred to as (tr, un), (tr, X11), (tr, mv), (% un), (% X11) and (% mv) below, where tr and % denote trended and percentage-change (detrended) data, respectively, and un, X11 and mv denote unadjusted, seasonally-adjusted by Census X-11, and moving-averaged data, respectively.

3. Results

At first the procedure we followed in this study is illustrated in Figs. 1–3 with the quart-

erly time series of Japanese exports (EXP).

Figures 1(a) and (a') show the time series of the unadjusted (original) data and its percentage change from the previous quarter, respectively. They are referred to as (tr, un) and (% , un), respectively. Figures 1(b) and (b') show the seasonally-adjusted time series by the Census X-11 and its percentage change, referred to as (tr, X11) and (% , X11), respectively. Figures 1(c) and (c') show the moving average of the original data over six quarters and its percentage change, referred to as (tr, mv) and (% , mv), respectively.

Figures 1(a), (b) and (c) reveal an exponential growth. The curves for the processed data (b) and (c) are smoother than the unadjusted data (a). On the other hand, Figs. 1(a'), (b') and (c') are bounded. While Fig. (a') looks more periodic, Figs. 1(b') and (c') more erratic.

Figures 2(a) to (c') show the G-P plots of EXP, which are calculated from Figs. 1(a) to (c'), respectively. Correlation integrals are calculated for various embedding dimensions ($d=1, 2, \dots, 15$). As shown in Fig. 2 the correlation integrals are not necessarily straight in the whole region of ϵ . Since a correlation dimension is defined from linear slopes of the plots, we must be careful to choose the linear regions in the G-P plot, as described in section 2.

In Fig. 3 the slope D_2 is plotted against the embedding dimension d . The correlation dimension D_2^* is defined as the saturated value of D_2 . In this case D_2^* 's are (a) 1.56 ($4 \leq d \leq 9$), (a') 2.88 ($5 \leq d \leq 9$), (b) 0.94 ($2 \leq d \leq 7$), (b') diverging, (c) 1.13 ($2 \leq d \leq 9$), and (c') diverging for (a) (tr, un), (a') (% , un), (b) (tr, X11), (b') (% , X11), (c) (tr, mv), and (c') (% , mv), respectively.

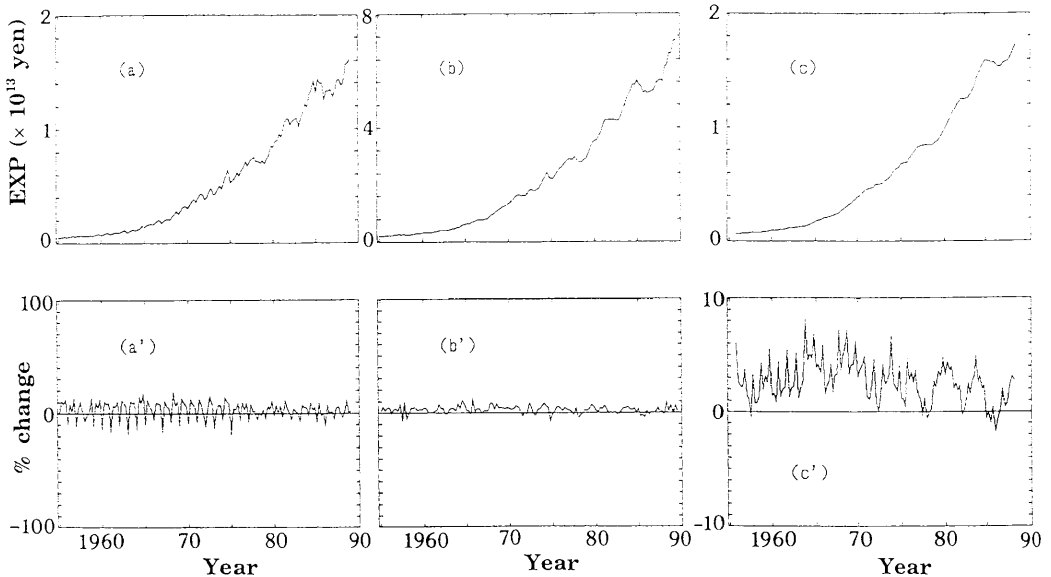


Figure 1 Time series of EXP (exports); (a) seasonally-unadjusted, (b) seasonally-adjusted, and (c) moving-averaged data; (a') to (c'), percentage change from the previous quarter for (a) to (c), respectively.

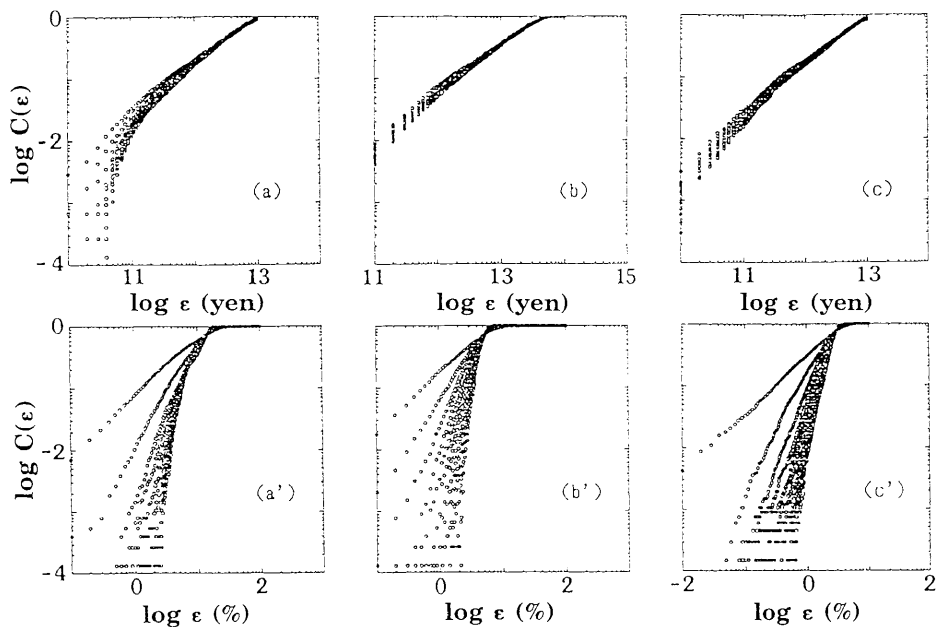


Figure 2 Correlation integrals (G-P plots) for the time series (a) to (c') of EXP shown in Fig. 1. The embedding dimension d is taken as a parameter.

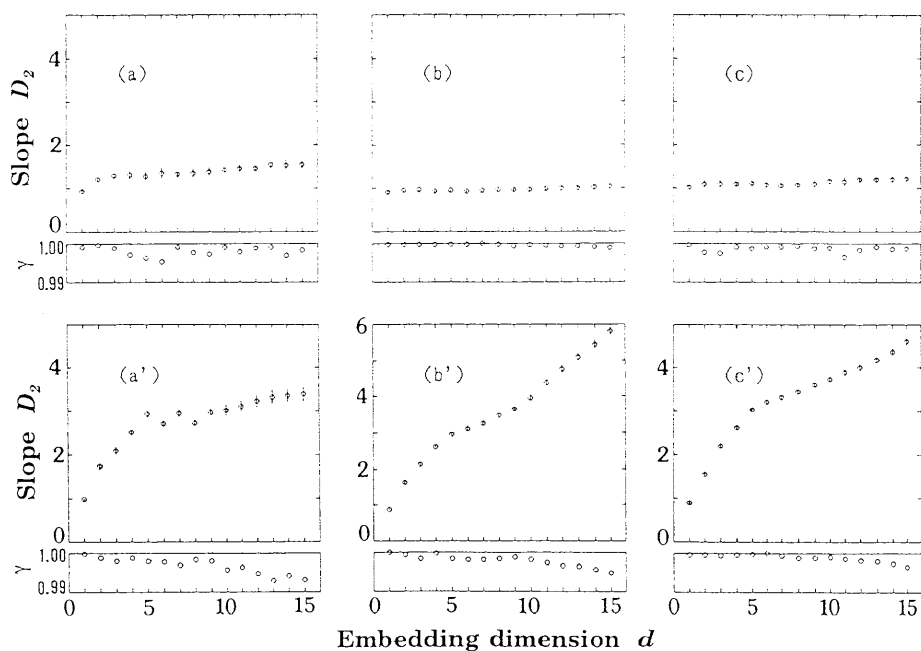


Figure 3 Slopes D_2 of the G-P plots (a) to (c') of Fig. 2, plotted against the embedding dimension d . The correlation coefficient γ of the regression line used to calculate the slope is also shown.

In Table II the correlation dimensions obtained for six types of data of the six kinds of economic variables are summarized. In general either of the two cases occurs; (a) a finite correlation dimension D_2^* is obtained, and (b) not obtained. In the case (a), D_2^* is defined as the mean D_2 over the range of correlation dimension d indicated in the parenthesis. In some cases (e.g., GNP (tr, mv) and PEI (% , mv)) more than one D_2^* value is given for different regions of ε . As described in section 2 it is not easy to apply the selection rule to them strictly to define a unique correlation dimension. In other cases (e.g., PHI (tr, un) and PFCE (% , un)) more than one D_2^* value is given for different regions of d in the same D_2 - d plot, because the plot has more than one plateau region. In the case (b) there are two cases; D_2 diverges obviously and increases slowly, as d increases. The term “slowly increasing” implies that the difference from a plateau is subtle.

The correlation dimensions for the six types of data of a given economic variable differ. It means that processing the data changes the dynamics of the system. The changes are assumed to reflect the dynamics of the original system. In this context we may classify the six economic variables examined into three classes on the basis of the results in Table II; (a) GNP and PEI, (b) EXP, PHI and PFCE, and (c) IMP. We will discuss the implication of this classification in the next section.

4. Discussion

We investigated the time series of the six kinds of Japanese economic variables, GNP, PEI, EXP, PHI, PFCE, and IMP, using the correlation integral method. For each of the variables the six types of data, trended/detrended and unadjusted/seasonally-unadjusted/moving-averaged i.e., (tr, un), (tr, X11), (tr, mv), (% , un), (% , X11) and (% , mv), were prepared to explore the dynamical aspects of the economic variables. It is found that finite correlation dimensions exist for some types of the data, but do not for others, even for the same economic variable, as shown in Table II. This fact suggested that the dynamical structures of the economic variables are changed by processing the data (i.e., seasonal adjustment or moving average). The changes by the processing depend on the dynamics of the variables. Accordingly, the classification of the six economic variables into the three classes based on the results in Table II reflects their dynamics.

The first class includes GNP and PEI, whose correlation dimensions are obtained for five of the six types of data, but not for (tr, un). It should be emphasized that finite correlation dimensions for both (% , X11) and (% , mv) are obtained only for GNP and PEI. Roughly speaking, these results imply that these economic variables are confined in low dimensions with respect to the growth rate. In other words, the growth rate is a more essential variable to describe the dynamical behavior of GNP and PEI than the total amount.

The second class includes EXP and PHI, whose correlation dimensions are obtained for (tr, un), (tr, X11), (tr, mv), and (% , un), but not for (% , X11) and (% , mv). PFCE is likely to be classified into this class, although the correlation dimension for (tr, X11) is not

Table II Correlation dimensions^{a)}

Economic variable	Unadjusted		Seasonally-adjusted		Moving-averaged	
	trended (tr, un)	%-change (%, un)	trended (tr, X11)	%-change (%, X11)	trended (tr, mv)	%-change (%, mv)
GNP	— (slowly increasing)	0.75 ($7 \leq d \leq 13$)	1.94 ^{b)} ($7 \leq d \leq 9$) or — (diverging)	1.85 ($5 \leq d \leq 9$)	2.42 ^{b)} ($7 \leq d \leq 10$) 2.64 ($6 \leq d \leq 12$)	1.47 ($3 \leq d \leq 8$)
PEI	— (slowly increasing)	2.04 ($8 \leq d \leq 15$)	1.54 ($9 \leq d \leq 15$)	1.92 ($8 \leq d \leq 12$)	1.80 ($7 \leq d \leq 10$)	3.30 ^{b)} ($10 \leq d \leq 15$) or 3.99 ($7 \leq d \leq 14$)
EXP	1.56 ^{b)} ($4 \leq d \leq 9$) or — (slowly increasing)	2.88 ($5 \leq d \leq 9$)	0.94 ($2 \leq d \leq 7$)	— (diverging)	1.13 ($2 \leq d \leq 9$)	— (diverging)
PHI	1.15 ($2 \leq d \leq 7$)	1.40 ^{c)} ($12 \leq d \leq 15$)	2.33 ($4 \leq d \leq 8$)	1.23 ($3 \leq d \leq 9$)	— (diverging)	1.47 ($7 \leq d \leq 14$)
PFCE	1.61 ($4 \leq d \leq 13$)	1.43 ($4 \leq d \leq 7$)	1.36 ^{b,c)} ($9 \leq d \leq 13$) or 0.95 ($2 \leq d \leq 4$)	— (slowly increasing or diverging)	— ^{d)}	1.49 ^{b)} ($11 \leq d \leq 15$) or 2.21 ($4 \leq d \leq 15$)
IMP	— (slowly increasing)	— (diverging)	— (slowly increasing)	— (diverging)	1.44 ^{b)} ($7 \leq d \leq 12$) or 1.83 ($5 \leq d \leq 11$)	— (diverging)

^{a)} A correlation dimension is defined as the mean value of D_2 averaged over the range of embedding dimensions d specified in the parenthesis. — indicates that no correlation dimension is obtained, because D_2 diverges obviously or increases slowly as the embedding dimension d increases.

^{b)} Two correlation dimensions determined in the different regions of ε are given. The value in the upper row is obtained in the region of relatively larger ε than that in the lower row.

^{c)} Two correlation dimensions determined in the different regions of embedding dimension d are given. The range of d is indicated in the parenthesis.

^{d)} 5.60 ($6 \leq d \leq 11$) or (diverging). However, 5.60 is too large to accept as a correlation dimension for the case with small size of data (see Eckmann and Ruelle, 1992).

obtained. In contrast to the first class, finite correlation dimensions for (tr, un) are obtained only for EXP, PHI and PFCE. Therefore, we may say that the total amount is a more essential variable to describe their dynamics than the growth rate for EXP, PHI, and PFCE.

The third class includes IMP, whose correlation dimensions are obtained only for (tr, mv). The correlation dimensions for (tr, mv) of the other variables are also obtained. The simple moving average of the data with trend seem to quench random deviations from the average value or to remove random noise. In this sense the correlation dimensions for (tr, mv) may be exceptional. Accordingly, the IMP result suggest that its behavior is rather stochastic.

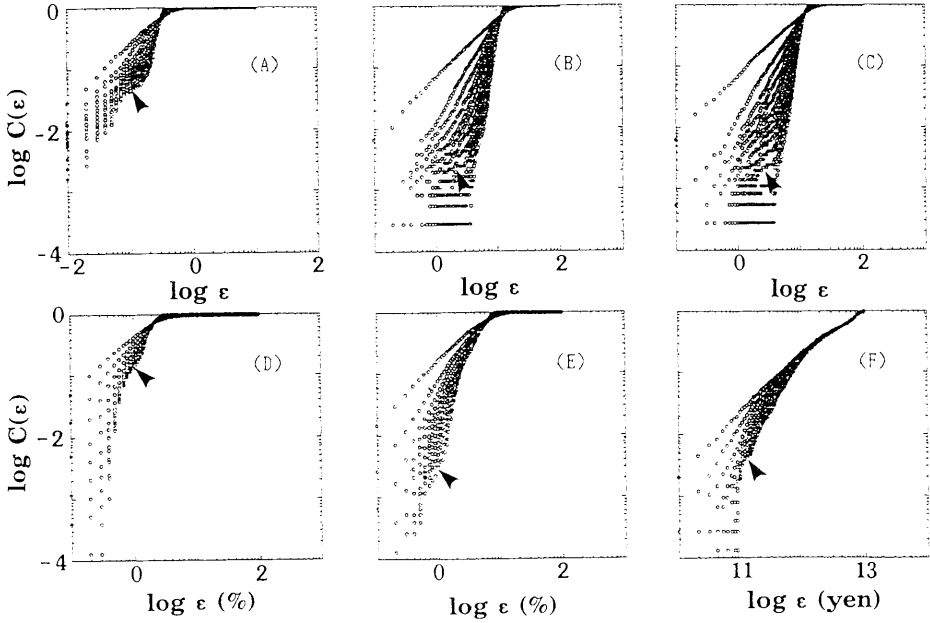


Figure 4 Correlation integrals for (A) Hénon map with 100 data points, (B) x-variable of Lorenz equation with 110 data points, (C) z-variable of Lorenz equation with 110 data points, (D) GNP (%), (E) PEI (%), and (F) IMP (tr, un). The arrows indicate the regions whose slopes are small locally as compared with the neighboring ones.

Finally, we comment on a phenomenon peculiar to small-sized data in determining correlation dimensions. Figs. 4(A) to (C) show the G-P plots for Hénon map (Hénon, 1976) and Lorenz equation (Lorenz, 1963) calculated only with a small number of data ($\sim 10^2$) comparable to the size of the economic data we used here. There appear the regions whose slopes D_2 are considerably small as compared with the neighboring regions (indicated by arrows in Fig. 4). The correlation dimensions calculated in the regions just on the right side of such regions are slightly higher than the exact ones. Since such regions are not observed in the G-P plots with enough data, they are obviously an artifact owing to the small-sized data. Similar phenomena were found in the economic data we analyzed. The examples are shown in Figs. 4(D) to (F). The slopes of such regions as indicated by arrows in Fig. 4(D) to (F) have unreasonably small values. Eventually we ignored these region in measuring the slope.

In conclusion, it is a challenging theme to reveal the dynamics of economic systems. Use of the correlation integral method is the one of the efficient approaches. Although small-sized data make it difficult to obtain the correlation dimensions exactly, the applications to various types of data of the same economic variable have provided a novel aspect in classification of the economic variables from the dynamical point of view.

References

- Barnett, W. A. and P. Chen, 1988, The aggregation-theoretic monetary aggregates are chaotic and have strange attractors, in: *Dynamic Econometric Modelling* (Barnett, W. A., E. Berndt and H. White, eds.), Cambridge University Press, 199–245.
- Barnett, W. A., A. R. Gallant, M. J. Hinich, and M. J. Jensen, 1992, Robustness of nonlinearity and chaos test to measurement error, inference method, and sample size, Working Paper #167, Washington University.
- Barnett, W. A., and M. J. Hinich, 1993, Has chaos been discovered with economic data? in: *Nonlinear Dynamics and Evolutionary Economics*, (Day, R. H., and P. Chen, eds.), Oxford University Press, 254–265.
- Benhabib, J. and R. H. Day, 1981, Rational choice and erratic behavior, *Review of Economic Studies* 48, 459–471.
- Brock, W. A., 1986, Distinguish random and deterministic systems, *Journal of Economics Theory* 40, 168–195.
- Brock, W. A., 1988, Nonlinearity and complex dynamics in economics and finance, in: *The Economy as an Evolving Complex System* (Anderson, P. W, K. J. Arrows and D. Pines, eds.), Addison-Wesley, 77–97.
- Brock, W. A. and C. L. Sayers, 1988, Is the business cycle characterized by deterministic chaos? *Journal of Monetary Economics* 22, 71–90.
- Chen, P., 1988, Empirical and theoretical evidence of economic chaos, *System Dynamics Review* 4, 81–108.
- Day, R. H., 1982, Irregular growth cycles, *American Economic Review* 72, 406–414.
- Eckmann, J. P. and D. Ruelle, 1992, Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems, *Physica D* 56, 185–187.
- Frank, M., R. Gencay and T. Stengos, 1988, International chaos? *European Economic Review* 32, 1569–1584.
- Frank, M. and T. Stengos, 1988, Some evidence concerning macroeconomic chaos, *Journal of Monetary Economics* 22, 423–438.
- Frank, M. and T. Stengos, 1989, Measuring the strangeness of gold and silver rates of return, *Review of Economics studies* 56, 553–567.
- Grandmont, J. M., 1985, On endogenous competitive business cycles, *Econometrica* 53, 995–1045.
- Grassberger, P. and I. Procaccia, 1983a, Characterization of strange attractors, *Physical Review Letters* 50, 346–349.
- Grassberger, P. and I. Procaccia, 1983b, Measuring the strangeness of strange attractors, *Physica*, 9D, 189–208.
- Hénon, M., 1976, A two-dimensional mapping with a strange attractor, *Communications in Mathematical Physics*, 50, 69–77.
- Lorenz, E. N., 1963, Deterministic Nonperiodic Flow, *Journal of Atmospheric Sciences*, 20, 130–141.
- Owase, T., 1991, Nonlinear dynamical systems and economic fluctuations: A brief historical survey, *IEICE Trans. E74*, 1393–1400.
- Peters, E. E., 1991, A chaotic attractor for the S & P 500, *Financial Analysis Journal*, 55–62.
- Scheinkman, J. A. and B. LeBaron, 1989, Nonlinear dynamics and stock returns, *Journal of Business* 62, 311–337.
- Schuster, H. G., 1988, *Deterministic Chaos—An Introduction*, VHS, Weinheim.
- Tanaka, T., 1993, Is business cycle interpreted as a deterministic chaos? Working Paper GLOCOM 24, International University of Tokyo (in Japanese).