Paper

Application of Correlation Integral Method to Economic Data

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

Abstract: The correlation integral method is applied to seasonally adjusted and non-adjusted economic data. From the analysis of the correlation dimensions of several kinds of economic data, which can be derived from their correlation integrals, it is found that the economic data can be classified into three categories from a viewpoint of a dynamical system: (a) no correlation dimension exists for the adjusted data, while a finite correlation dimension is obtained for the non-adjusted data (e.g., private housing investment), (b) no correlation dimension exists for either data, and (c) finite correlation dimensions are obtained for both data (e.g., private equipment investment).

1. Introduction

Recently economy has been investigated from the view-points of deterministic chaos to explore some dynamical aspects in economic system [Benhabib and Day (1981), Day (1982), Grandmont (1985), Brock (1986, 1988), Barnett and Chen (1988), Chen (1988), Hiromatsu and Tanaka (1988), Peters (1991)]. One of their interests is to seek the determinisity in economic data with erratic behavior. For example, GNP, private equipment investment, and private consumption expenditure have been analyzed by the correlation integral method proposed by Grassberger and Procaccia (1983a, b) with such a motivation.

However, it should be noticed that in those analyses only seasonally adjusted data were considered. As a matter of fact, it has been pointed out that one should be much careful in using seasonally adjusted economic data to investigâte dynamic properties of economic variables, since there exists a possibility to cause some artifact in the seasonally adjusted data by taking a moving average of original data. For example, an artificial fluctuation, known as Yule-Slutky effect, produced by inappropriate moving average of original data, and an artificial periodicity resulting from averaging random variables (Gaussian random numbers) are well known. In addition to these effects, when we study economic data as a dynamical system as described in this paper, we are afraid that averaging procedure of original data may introduce a kind of stochasticity into resultant averaged data, although original data has a deterministic structure in itself. Since there is a possibility at any rate for averaging original data to change underlying dynamical structures in original economic data, it is necessary to examine results derived from seasonally adjusted data, being compared with those from non-adjusted original data.

In this paper, we take an approach to such a problem by means of the correlation integral method, which is used to assess whether or not a dynamical system has deterministic structure. Differences between correlation integrals calculated for several kinds of

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

adjusted and non-adjusted economic data are examined. In the next section the correlation integral method is explained briefly. Since a size of economic is data considered too small to apply the correlation integral method, applicability of the method to small sized data, such as economic data, is also discussed, based on the previous work where well-known physical systems (Lorenz equation and Hénon map) are analyzed in this point of view [Nagai, Inaba, and Wako (1993)]. In section 3 the correlation integral method is applied to the economic variables (private equipment investment, private housing investment), and the results are discussed in section 4.

2. Method

In this study we utilize 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. In the following the G-P method is summarized briefly and its applicability to small-sized data, such as economic data, is examined, because it is usually considered that the G-P method can be applied to a time series with more than 10^4 data points, whereas the sizes of most of economic data are in the order of 10^2–10^3.

2.1 G-P method

Takens (1983) showed that higher-dimensional manifold can be reconstructed from any low dimensional sequential data by the method of embedding if the manifold is smooth, and that some topological properties of the manifold are conserved. Following such nature of embedding of sequential data, Grassberger and Procaccia designed a method to estimate a fractal dimension of the reconstructed manifold by calculating correlation integrals of the sequential data. The fractal dimension provides a lower bound to the degrees of freedom for the system. First we summarize the method briefly.

Let $\{x_0, x_1,\ldots, x_n\}$ be a one-dimensional time series. Making use of this primary one-dimensional data, we define a d-dimensional vector $\vec{\zeta}_i$ as

$$\vec{\zeta}_i=\{x_i, x_{i+\tau}, x_{i+2\tau},\ldots, x_{i+(d-1)\tau}\} \tag{2.1}$$

with time delay τ (in practice τ is set to one throughout this study), where d is called an embedding dimension. Then we consider the d-dimensional vector time series

$$\{\vec{\zeta}_0, \vec{\zeta}_1, \ldots, \vec{\zeta}_m\}, \tag{2.2}$$

where $m=n-d+1$.

The first step of the G-P method is to calculate the 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) separated by a distance smaller than a given distance ε, i.e.,

$$C(\varepsilon)=\sum_{i=1}^{n}\sum_{j=i+1}^{n}\Theta(\varepsilon-|\vec{\zeta}_i-\vec{\zeta}_j|)/m(m-1) \tag{2.3}$$

where Θ and $|\vec{\zeta}_i-\vec{\zeta}_j|$ are a Heaviside step function (i.e., $\Theta(y)=0$ if $y<0$ and 1 otherwise) and a distance between $\vec{\zeta}_i$ and $\vec{\zeta}_j$, respectively, and the number of pairs separated by ε is normalized by the total number of pairs $m(m-1)$.

For a chaotic attractor the correlation integral $C(\varepsilon)$ increases at a rate of ε^D for a given embedding dimension d over the range of smaller ε (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

$$D_2 = \lim_{\varepsilon \to 0} \log C(\varepsilon) / \log \varepsilon. \tag{2.4}$$

Equation (2.4) indicates that D_2 can be obtained as a linear slope of log C(ε) versus log ε plots (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. If D_2 converges to some finite value as the embedding dimension d increases, a converged value of D_2 is called a correlation dimension. It is known that the correlation dimension is less than or equal to the fractal dimension (box counting dimension) of the chaotic attractor. However, in practice, estimation of the correlation dimension for an actual system is not straightforward because of a shortage of data points available. This point will be discussed in the section 3 below.

2.2 Applicability of the G-P method to small-sized data

The G-P method is usually applied to a time series consisting of a large number of data points, say, greater than the order of 10^4. Even though the G-P method has been utilized to analyze economic data from the viewpoint of deterministic chaos, much attention has not been paid to applicability of the G-P method to such small-sized data of the order of 10^2 [Chen (1988)]. Therefore, before applying the G-P method to economic data, we discuss its applicability here and examine it in the previous work [Nagai, Inaba, and Wako (1993)].

In the previous work, we take the Lorenz equation and Hénon map as examples, because the correlation dimensions for these models have been already known [Grassberger and Procaccia (1983)] According to the results of these analyses [Nagi, Inaba, and Wako (1993)], although the correlation dimension of the system determined from a small number of data points is not exactly the same as the true one, the deviations of the calculated correlation dimensions are less than ten percent. It is also found that, in the G-P plot, the region of ε in which a slope converges as an embedding dimension increases shifts to a larger ε region as the number of data points decreases [Nagai, Inaba, and Wako (1993)]. In other words, although it is better that the slope of the G-P method is measured in the region of smaller ε [see eq. (2.4)], one should be careful that the error becomes dominant in the region of smaller ε for small-sized data. As a result the correlation integral can be applied to small-sized data, as far as we confine ourselves to the qualitative nature of the system, but we should keep in mind that there remains uncertainty in the results calculated from the small-sized data.

3. Results

The results obtained for seasonally adjusted data are compared with those for seasonal-

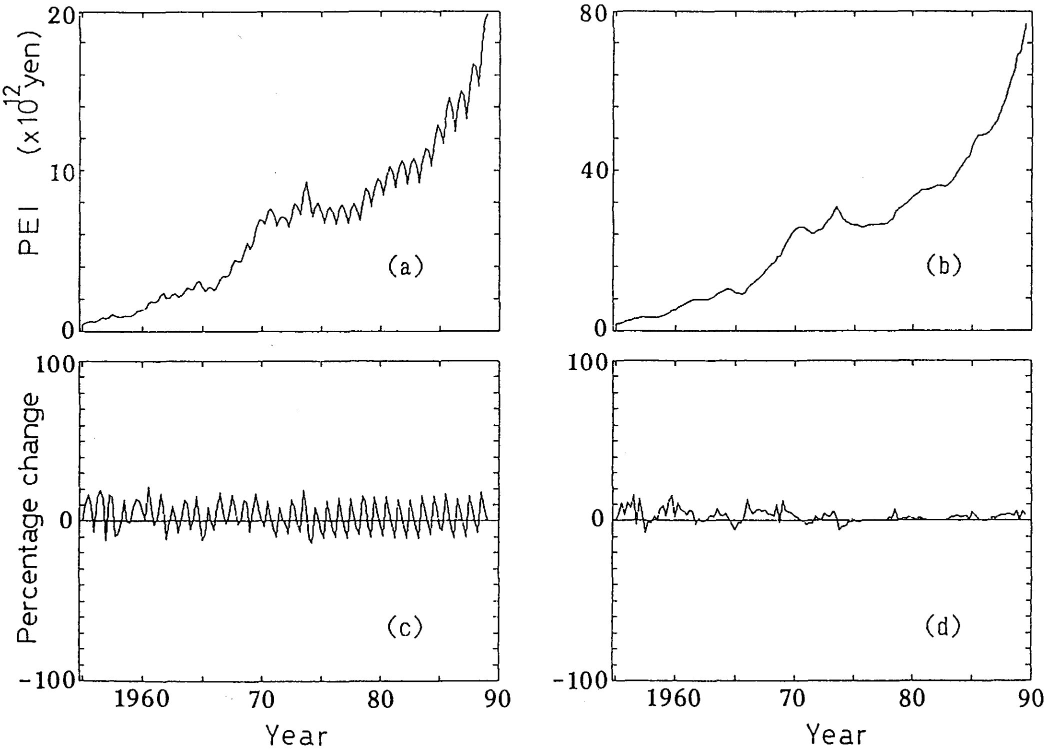

Figure 1 Time series of PEI (private equipment investiment); (a) non-adjusted date and (b) seasonally adjusted data; (c) percentage change from previous quarter for non-adjusted data and (d) percentage change from previous quarter for seasonally adjusted data.

ly non-adjusted data. Out of several economic variables studied, private equipment investment (PEI) and private housing investment (PHI) are explored in detail. In this study we used the data of these economic variables in the "Annual Report on National Accounts" by the Economic Planning Agency, Government of Japan.

3.1 Private equipment investment (PEI)

Figure 1 shows the quarterly time series of Japanese PEI at constant price from the first quarter of 1955 to the first quarter of 1989 (137 data points). Figures 1a and c show the time series of the seasonally non-adjusted (original) data and its percentage change from previous quarter, respectively. The latter is referred to as detrended data for convenience sake in this paper, since the percentage change is widely used for detrending (however, in actual fact, choice of a detrending method for economic time series is not a very simple problem [Chen (1988)]). Therefore, it is not sure that the percentage change is the best detrending method for the PEI time series. The same is true for other economic variables discussed below.

Figures 1b and d, which correspond to Fig. 1a and c, respectively, show the seasonally adjusted time series of PEI calculated in annual rate. The seasonal adjustment was perform-

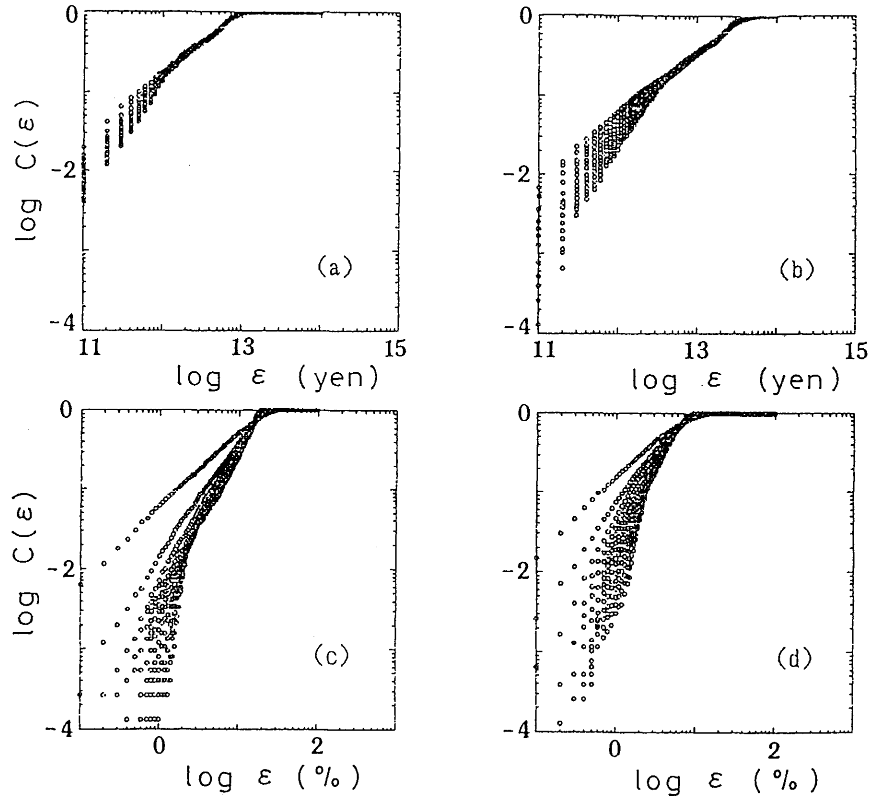

Figure 2 Correlation integrals (G-P plots) for the time series, (a)–(d) of PEI given in Fig. 1. The embedding dimension $d=1, 2, ..., 15$ is taken as a parameter.

ed by the Census Method 11 (X-11).

Figures 1a and b, the non-adjusted and adjusted data involving trend, reveal exponential growth. Naturally the adjusted data (Fig. 1b) is more smooth than the non-adjusted one (Fig. 1a). On the other hand, the detrended non-adjusted time series (Fig. 1c) looks more periodic (or less erratic) than the detrended adjusted one (Fig. 1d).

Figure 2 shows the G-P plots of PEI data given in Fig. 1. Each correlation integral is calculated for various embedding dimensions ($d=1, 2, ..., 15$). As shown in Fig. 2, the correlation integrals are not necessarily straight in the whole region of ε, although the correlation dimension is defined from slopes of the plots. Therefore we must carefully choose the linear regions in the G-P plot.

In order to illustrate how the correlation dimension is determined from the G-P plot, the analysis of the G-P plots for the detrended non-adjusted data (Fig. 2c) is shown in Fig. 3 as an example. In the analysis intermediate region of ε is divided into four regions. Since $C(\varepsilon)$ becomes saturated at one for too large ε, and since the noise of the data resulting from the shortage of the pairs separated by small ε dominates for too small ε, the region of very large ε and that of very small ε are not taken into account.

Figures 3b-e reveal the dependence of the slope D_2 on the region in which it is measured

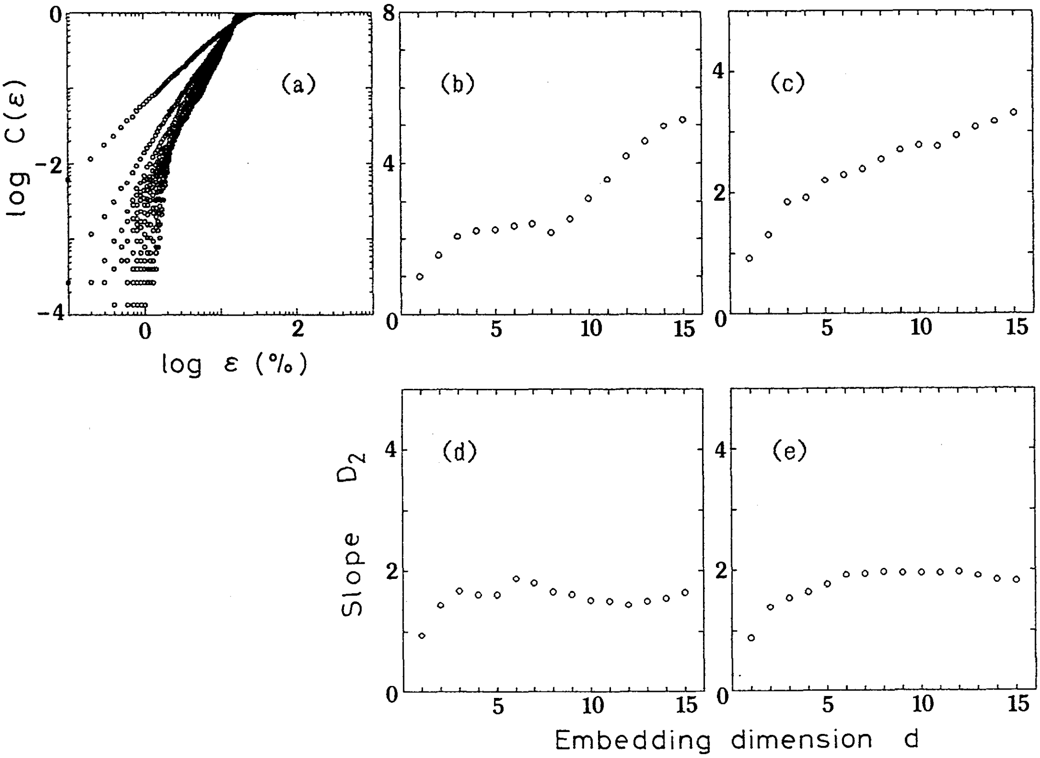

Figure 3 Dependence of a slope of the G-P plot on a region of ε. The slopes of the G-P plots given in Fig. 2(c) are measired at four regions of ε [(b) to (e)], namely, (b) $0.2<\varepsilon<0.3$ (c) $0.3<\varepsilon<0.48$ (d) $0.6<\varepsilon<0.77$ (e) $0.8<\varepsilon<1.12$.

(the slope is calculated by the linear regression method in each region). Clearly the slope diverges in the region c as the embedding dimension d increases. On the other hand, the slopes in regions d and e have plateau regions for relatively heigher embedding dimensions. The correlation dimension is defined as the saturated value of D_2. In this sense, although the correlation dimension can be definitely determined in the region e ($D_2=1.95$, if averaged over the range $6 \leqq d \leqq 12$), it is relatively ambiguous in the region d because of the fluctuation of the slopes ($D_2=1.61$ if averaged over the range $3 \leqq d \leqq 15$). In the region b the slope saturates in the range $4 \leqq d \leqq 8$ (mean D_2 value is 2.28 over this region), but increases beyond $d=8$. Since it is known that the G-P plots for the embedding dimensions considerably larger than the actual correlation dimension sometimes give rise to an unexpected increase of the slope, the increase of the slopes beyond $d=8$ in Fig. 3b is considered to be artifact. Consequently the plateau at $d=4$ through 8 is significant. Based on such a consideration, it is necessary to define some criterion to determine the correlation dimension from the slopes of the G-P plot.

The correlation dimension of a system is determined according to the following criterion in this study. First of all, the essentially straight regions in the G-P plot are detected and a slope of the region is measured. The straightness of the G-P plot in each region is

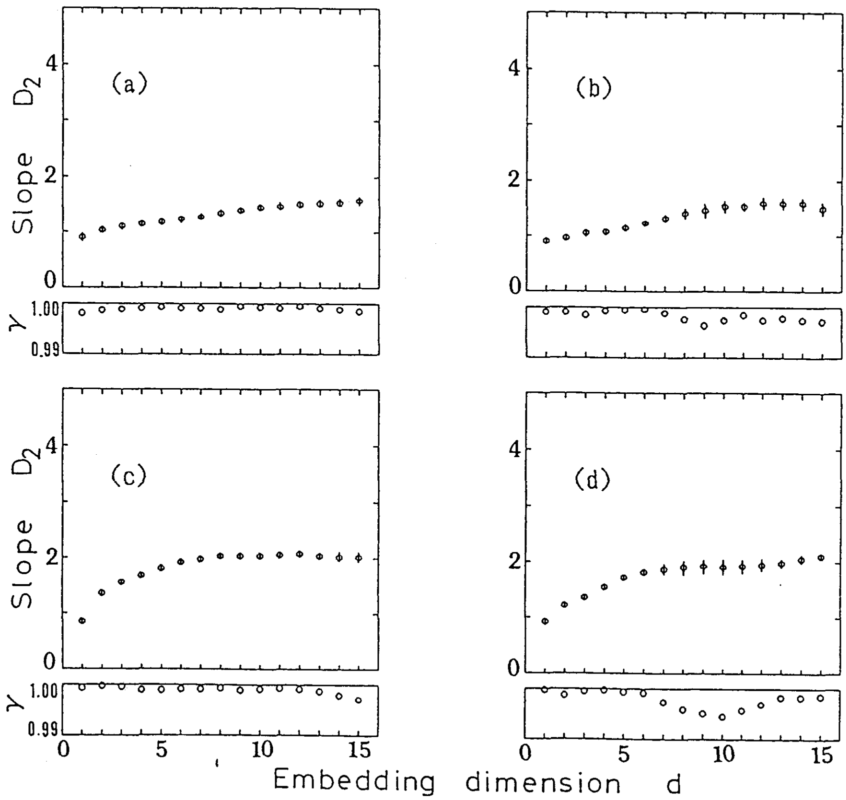

Figure 4 Slopes of the G-P plots (a)-(d) given in Fig. 2.

assessed by the correlation coefficient of the regression line. Then the slope is plotted against embedding dimension d. There are two cases of in the plot of slope for correlation integral versus embedding dimension. The first case is that the slope becomes almost constant for d larger than a certain embedding dimension d_1. The case that the slope saturates at d_1, but resumes increasing at a certain dimension d_2 larger than d_1(e.g., Fig. 3b) is also included in this case. However, if the former type plot is obtained besides the latter type one, the latter type plot is discarded. If the G-P plot have several linear regions and consequently more than one slopes are obtained, the slopes measured in the region of the smallest distance ε (lower left region in the G-P plot) is taken. 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, as a matter of fact, it is very hard to determine proper linear region in the G-P plot thoroughly automatically. As a result there still remains ambiguity in the determination of the correlation dimension. The second case is that the slopes diverge without any plateau regions as d increases. For this case we regard that the system has no correlation dimension.

Figure 4 shows the slopes calculated from Figs. 2a-d, together with correlation coefficients of regression lines determined for the slope. The very high correlation coefficients of greater than 0.98 throughout this study indicate that the G-P plots in the

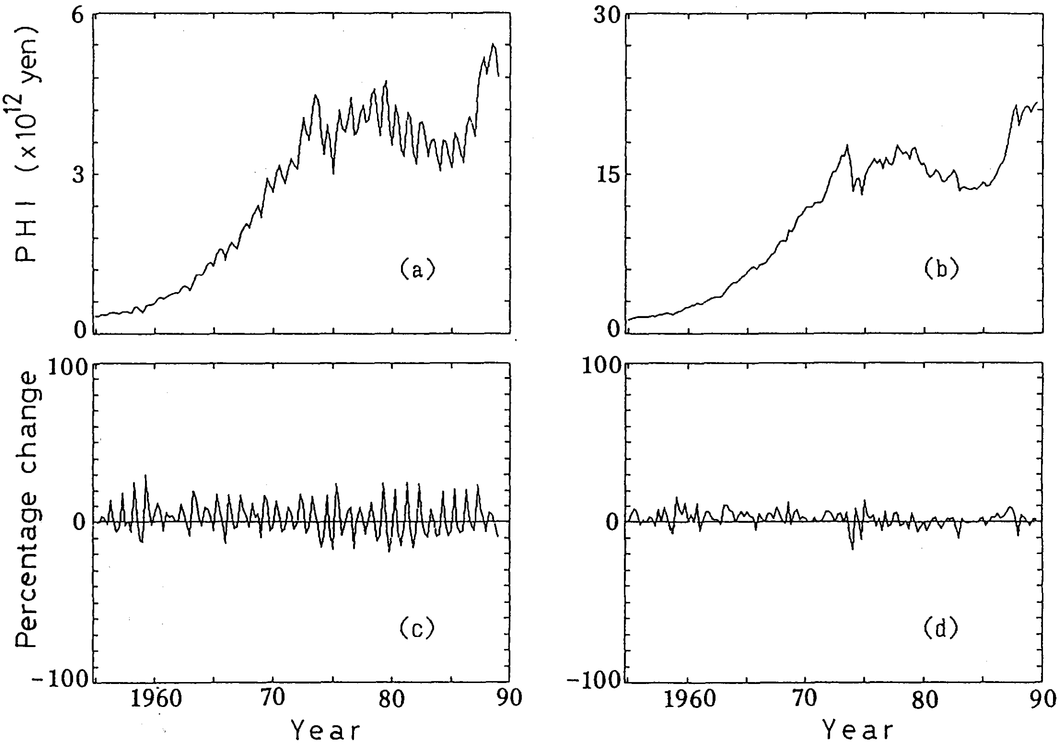

Figure 5 The same as Fig. 1, but for PHI (private housing investiment).

region where slopes are calculated are almost on a straight line. The correlation dimensions of detrended non-adjusted and detrended adjusted data (Figs. 4c and d, respectively) are near 2 from the range $8 \leqq d \leqq 15$, respectively. On the other hand, since the slope of the non-adjusted data including trend (Fig. 4a) gradually increases as the embedding dimension increases, we conclude that the correlation integral has no correlation dimension. Although the slope for the adjusted data including trend (Fig. 4b) seems to converge for relatively larger embedding dimensions, it does not have a plateau region for relatively smaller embedding dimension (i.e., $d < 10$). It is hard to determine the correlation dimension definitely for this case.

3.2 Private housing investment (PHI)

Next, we investigate the properties of the Japanese PHI data at constant price for the same period as the PEI data above. The time series data are shown in Fig. 5 in the same way as Fig. 1; Figures 5a and c are the time series of non-adjusted (original) data and its percentage change from previous quarter, respectively, and Figs. 5b and d the time series of the seasonally adjusted data including trend and its percentage change from previous quarter, respectively. As seen in Figs. 5a and c PHI increased exponentially before the first oil shock (in 1974), but became stagnant after that (even declined in the early 1980s), and again turned to increase in the late 1980s. On the other hand, detrended non-adjusted data (Fig. 5c) looks

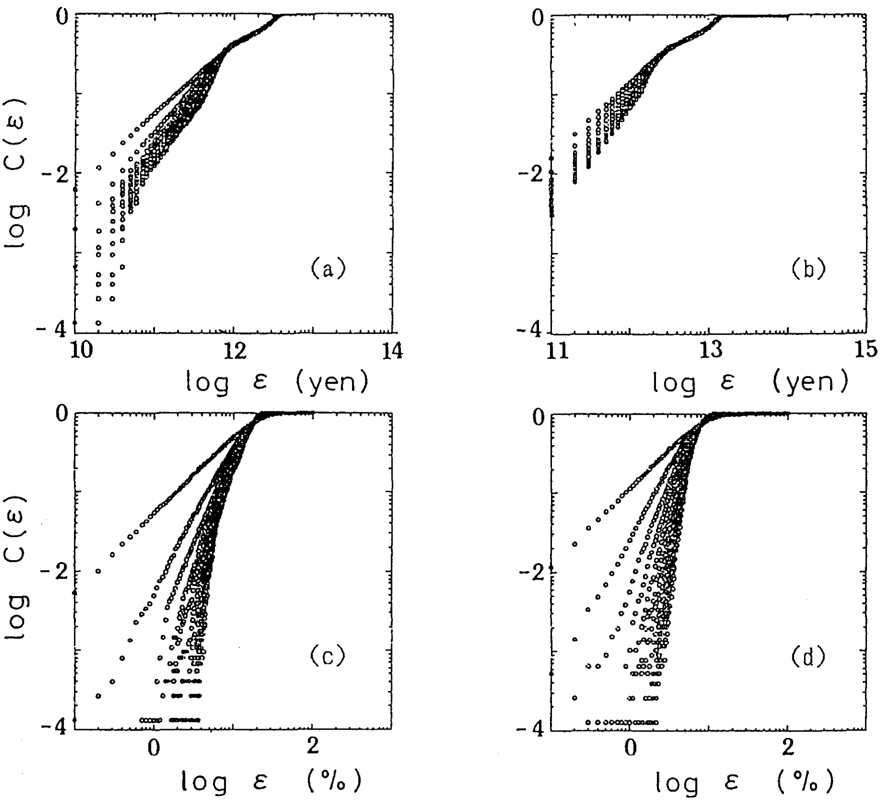

Figure 6 The same as Fig. 2, but for PHI.

more periodic than detrended adjusted one (Fig. 5d).

Figure 6 shows the correlation integrals for the PHI time series. They are also calculated for embedding dimensions $d=1$ to 15.

Then the correlation dimensions are determined. The slopes versus embedding dimension plots utilized to determine the correlation dimensions are shown in Fig. 7. As for the correlation dimensions for the data not detrended (Figs. 7a and b), it is well defined for Fig. 7b, while there are two plateaus for Fig. 7a. The correlation dimension of detrended non-adjusted data (Fig. 7c) is obtainable from the region $4 \leqq d \leqq 8$. On the other hand, no correlation dimension is defined in the detrended adjusted data (Fig. 7d).

4. Discussion

In the previous section it is found that the seasonal adjustment (or a certain averaging process) of economic data, which is most conventional approach, sometimes changes the underlying dynamical structures which the original data has. The changes in the structure may give rise to the change in the correlation dimension. Three cases of changes (including the case in which no change occurs) are found in the study: (a) no correlation dimension exists for the adjusted data, while a finite correlation dimension is obtained for the non-

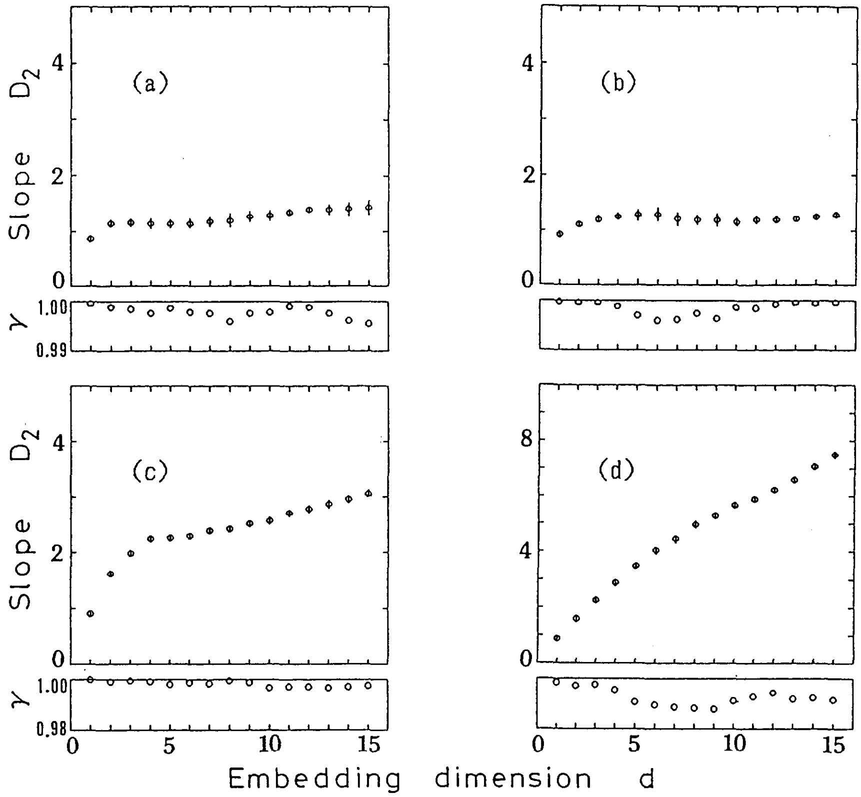

Figure 7 The same as Fig. 4, but for PHI.

adjusted data, (b) no correlation dimension exists for either data, and (c) finite correlation dimensions are obtained for both data.

As for the case (a), the reason why the economic structure change in such an averaging operation is considered as follows. Since for the non-adjusted data a finite correlation dimension is obtained, the time series in such a system is likely to be generated in a certain kind of deterministic mechanism. However, the averaging of the data makes it approach to random numbers (i.e., it obeys the law of large numbers). As a result of such randomization the correlation dimension cannot be determined from the correlation integral for the adjusted data in some cases (e.g., for Gaussian noise, the slope diverges as the embedding dimension d increases [e.g. papers in the book edited by Mayer-Kress (1986)]). It is the case of PHI detrended data (Figs. 7c and d). Hénon map and z variable in the Lorenz equation also correspond to this case [Nagai, Inaba, and Wako, (1993)]. For the data categorized into this case one should be careful that the seasonally adjusted data looses some kind of information they have in their structure.

In the case (b) there has no correlation dimension for either the non-adjusted or adjusted data. This case is very difficult to find the determinisity in the system generating such data.

In the case (c) the finite correlation dimension is obtained even for adjusted data, in spite of averaging. PEI belongs to this class. In PEI the correlation dimension of the seasonally adjusted data gets lower than that of the non-adjusted data. Altho is not sure whether or not the amount of decrease is significant, if so, it is interesting in the following sense. As shown in the previous work [Nagai, Inaba, and Wako (1993)] the correlation dimension of the averaged data becomes lower than the correlation dimension of the original data in x variable of Lorenz equation. Therefore, if a correlation dimension for the adjusted data smaller than that for the non-adjusted data is obtained, it implies that the economic data is generated from the similar mechanism to Lorenz system in some sense. In the Lorenz system the correlation dimension becomes smaller by averaging operation of x variable, since it compresses butterfly-shaped Lorenz attractor into caterpillar-shaped attractor lying between two unstable fixed points. In this case the procedure to adjust the original data provides useful information about the system analyzed.

At present there is no reliable approach to identifying deterministic chaos in economic data, overcoming the shortage of the data points. Furthermore, knowledge of the exact correlation dimension does not necessarily mean that the dynamical structure of the economic variables investigated can be enough understood. However, the analysis described here provide very suggestive information about the dynamical structure, and it is emphasized in this paper that comparative study between seasonally adjusted and non-adjusted data can add more information to it.

References

Barnett, W. A. and P. Chen, 1988, The aggregation-theoretic monetary aggregates are chaotic and have strange attractors, in Dynamic Econometric Modeling (by Barnett, W. A., E. Berndt and H. White), Cambridge University Press, 199–246.

Benhabib, J. and R. H. Day, 1981, Rational choice and erratic behavior, Review of Economic Studies **48**, 459–471.

Brock, W. A., 1986, Distinguishing random and deterministic systems, Journal of Economic Theory **40**, 168–195.

Brock, W. A., 1988, Nonlinearity and complex dynamics in economics and finance, in The Economy as an Evolving Complex system (by Anderson P. W., K. J. Arrows and D. Pines), Addison-Wesley, 77–97.

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.

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.

Hiromatsu, T., and T. Tanaka, 1988, Another interpretation of economic time series; An application of deterministic chaos equations to economic data analysis, Working Paper, No. 5, University of Tokyo, Komaba (in Japanese).

Nagai, Y., T. Inaba, and H. Wako, 1993, Decreasing or increasing of correlation dimension in the attractor reconstructed by the moving-averaged time series, **26**, Journal of the Faculty of General Education (Azabu University), 89–100.

Mayer-Kress, G. (ed.), 1986, Dimensions and Entropies in Chaotic Systems, Springer, Berlin-Heidelberg-New York-Tokyo.

Peters, E. E., 1991, A chaotic attractor for the S&P 500, Financial Analysis Journal, 55–62.

Schuster, H. G., 1988, Deterministic Chaos – An Introduction, VHS, Weinheim.

Takens, F., 1984, Detecting strange attractors in turbulence, Lecture Notes in Mathematics, **898**, 366–381.