No. 85 (80-23)

A STATICAL METHOD OF THE QUALITATIVE TREND CURVE ANALYSIS

By

Atsuyuki Okabe

September 1980

Institute of Socio-Economic Planning
University of Tsukuba
Sakura, Niihari, Ibaraki 305
Japan

ABSTRACT

This paper proposes a statistical method to analyze qualitative characteristics of a distribution, (represented by a curve, called a trend curve), of an attribute value with respect to a single parameter, (such as a time or a distance from a CBD). First trend curve is decomposed into local trends and these local trends are classified according to three categories: a "peak", a "bottom" and a "slope". Second a statistical method is formulated to test randomness in the appearance of "peaks" in the local trends. Third randomness in the arrangement of "peaks" is statistically tested. Fourth a qualitative correlation coefficient is formulated to examine the similarity between two trend curves. Last these statistical methods are applied to the study of the distribution of establishments on the Chuo Railway Line Area.

(1) INTRODUCTION

A problem of analyzing a trend of an attribute value y with respect to a single parameter x often arises in the context of urban and regional analysis. For example y is retail store density at location x along a street; y is population density at distance x from a CBD; y is a regional income at time x. In Sections 1, 2 and 3, y is, as an illustrative example, given by a ratio of the number of establishments belonging to the secondary sector at distance x from Shinjuku on the Chuo Railway Line Area. (See a map in Figure 4. The data are shown in Figure 1 by a continuous line. A detail explanation will be provided in Section 5 where an empirical study is pursued. For brevity y will be called the secondary establishment ratio hereafter.) To analyze a trend of

Figure 1. A ratio of the number of establishments belonging to the secondary sector at distance x from Shinjuku on the Chuo Railway Line Area, 1970.

y over $x \in L=\{x:0 \leq x \leq \ell\}$, the conventional method, which may be called the trend curve analysis, first assumes a certain functional form $y=f(x)$, such as an exponential function, (recall the study of urban population density), a polynomial function or the Fourier series function. Second the function $f(x)$ is estimated by data. Last a trend of the curve is stated in terms of the estimated function $\hat{y}=\hat{f}(x)$. In the example of Figure 1, the 5th order polynomial curve

$$y = a_0 + a_1x_1 + a_2x_2^2 + a_3x^3 + a_4x^4 + a_5x^5 \tag{1}$$

may fit. By use of the least square method, the estimated function is obtained as

$$\hat{y} = 1.61 + 8.90x - 2.17x^2 + .22x^3 - .10x^4 + .0002x^5 , \tag{2}$$

which is indicated by a broken line in Figure 1. A trend of the secondary establishment ratio over the Chuo Railway Line Area is thus characterized by the estimated coefficients $\hat{a}_0=1.61$, $\hat{a}_1=8.90$, ..., $\hat{a}_5=.0002$. This characterization, however, does not immediately give a clear implication of the trend. Alternatively a spectrum density function obtained from fitting the Fourier series function is sometimes employed. This attempt, however, is not always successful because the implications of the spectrum, such as "power" and "low frequency curves", are umbiguous in an urban and regional context. In urban and regional analysis, probably qualitative characteristics are of primary importance rather than quantitative characteristics. For example in Figure 1, the trend curve of the secondary establishment ratio is primarily characterized by two "peaks" (or two "bottoms" or five "slopes") and by these "peaks" being evenly spaced on the Chuo Railway Line Area. The gradient or curvature of the trend curve might be secondary. Generally the trend of y over L may qualitatively be characterized by the number of "peaks" (or "bottoms", "slopes") in L and by spacing of these "peaks" on L. To distinguish from the conventional trend curve analysis, the trend curve analysis focusing on the above qualitative characteristics will be called the qualitative trend curve analysis. The aim of this paper is to develop a statistical method for the qualitative trend curve analysis.

The qualitative trend curve analysis consists of the local trend curve analysis, global trend curve analysis and correlation analysis, which are respectively discussed in Sections 2, 3 and 4. The local trend curve analysis deals with a local trend of $y=f(x)$ in the neighbourhood of $x=x_0, x_1, \ldots, x_{n+2} \in L$. These local trends are classified according to a "peak", "bottom" or "slope" and a statistical test is formulated with respect to randomness in the appearance of "peaks" in L. The global trend curve analysis deals with the spacing of "peaks" on L. A statistical test is formulated with respect to randomness in the arrangement of "peaks" on L. The correlation analysis deals with a correlation between two trend curves $y_1=f(x)$ and $y_2=f(x)$, $x \in L$. A qualitative correlation coefficient is defined in terms of "peaks", "bottoms" and "slopes", and similarity (or dissimilarity) between two trend curves is examined by this coefficient. These qualitative trend curve analyses are employed in Section 5 to examine the trend curves of several kinds of establishment ratios on the Chuo Railway Line Area. The paper ends in Section 6 remarking a note for a further research.

Concerning the literature, the trend curve analysis is explicitly discussed in Ishimizu (1976, pp. 77-79) and it is implicitly delt with in the time-series analysis developed by many. (For instance, see Bennett (1979)). Also the literature of the trend surface analysis, (for example, King (1969, pp. 152-153), Cliff et al. (1975, Chapter 2) and others), should be noted because the trend curve analysis might be regarded as a specific case of the trend surface analysis.

(2) A STATISTICAL METHOD OF THE LOCAL TREND CURVE ANALYSIS

Consider an attribute value y being a function of x, $y=f(x)$, and let $L=\{x:0\leq x\leq \ell\}$ be the domain of $f(x)$. This domain is evenly devided into n+2 mutually exclusive and collectively exhausitive subdomains $L_i=\{x:i\leq x<i+1\}$, $i=0, 1, \ldots, n+1$, where $n+2=\ell$. (Note that the same size assumption is made for convenience.) The attribute value in each subdomain L_i is aggregated by

$$y_i = \int_i^{i+1} f(x)dx , \qquad i=0, 1, \ldots, n+1 , \tag{3}$$

and y_i is regarded as a function of a representative point $x=i$ of L_i, i.e., $y_i=f(i)$. The trend of y over L is hence described by a sequence of the aggregated attribute values $y_0, y_1, \ldots, y_{n+1}$ at $x=0, 1, \ldots, n+1$. (In practice data are usually given by y_i, $i=0, 1, \ldots, n+1$ instead of $y=f(x)$ $x\in L$).

To examine a qualitative local trend, a local area around $x=i$ is first defined: a <u>local area</u> around $x=i$ is an area consisting of L_i and subdomains directly adjacent to L_i, i.e., L_{i-1} and L_{i+1}. A <u>local trend</u> t_i of y around $x=i$ is then defined by a trend of y in the local area around $x=i$, that is, a trend described by a sequence of the attribute values y_{i-1}, y_i, y_{i+1}. (Note that the local area around $x=i$ and that around $x=j$ are overlapped if $j=i-2, i-1, i+1, i+2$.) Each local trend t_i is classified according to the following four categories: t_i is called a <u>peak</u> (denoted by $t_i=\Lambda$) if $y_{i-1}<y_i$ and $y_i>y_{i+1}$; t_i is called a <u>bottom</u> (denoted by $t_i=V$) if $y_{i-1}>y_i$ and $y_i<y_{i+1}$; t_i is called a <u>right downward slope</u> (denoted by $t_i=\backslash$) if $y_{i-1}>y_i>y_{i+1}$; t_i is called a <u>left downward slope</u> (denoted by $t_i=/$) if $y_{i-1}<y_i<y_{i+1}$. A local trend described by these categories is called a <u>qualitative local trend</u>. For notational convenience $t_i=T_i$ is used indicating that t_i takes a specific category Λ, V, $\backslash$ or /. With this notation the trend of y over the whole domain L is written as a sequence T of qualitative local trends $T=(t_1=T_1, t_2=T_2, \ldots, t_n=T_n)$ or briefly $T=(T_1, T_2, \ldots, T_n)$. In the example of Figure 1, T is written as

$$T = (\diagdown\diagdown V \,/\, \Lambda\, V \,/\,/\, \Lambda \diagdown) \ . \tag{4}$$

As is discussed in the introduction, a characteristic of the qualitative local trends T may be described by the numbers $N(\Lambda)$ of Λ, $N(V)$, $N(\diagdown)$ and $N(/)$ in T. Concerning these numbers the equation

$$N(\Lambda) + N(V) + N(\diagdown) + N(/) = n \tag{5}$$

holds, for there are n local trends in T. Moreover the equation

$$N(V) = N(\Lambda) - 1 \text{ or } N(\Lambda) \text{ or } N(\Lambda) + 1 \tag{6}$$

holds since one bottom always exists between two peaks. (Note that, this equation depends upon the boundary conditions $t_0=T_0$ and $t_n=T_n$.) From equations (5) and (6) it follows that

$$N(\diagdown) + N(/) = n - N(\Lambda) - 1 \text{ or } n - N(\Lambda) \text{ or } n - N(\Lambda) + 1 \ . \tag{7}$$

Equations (6) and (7) show that the number of bottoms and that of slopes can be written as the number of peaks. Therefore a characteristic of the qualitative local trends T can be described by the number of peaks alone.

In the example of Figure 1 it is observed that $N(\Lambda)=2$. Having noticed this, one may intuitively consider that a strong regularity exists in the qualitative local trends of the secondary establishment ratio because two peaks out of ten is few. However, is this conclusion theoretically appropriate? Is the same conclusion holds if the number of peaks is three? Or can it be concluded that the qualitative local trends are random if the number of peaks is three? How about four peaks? To answer these questions, let us now introduce a probabilistic model in which regularity in the qualitative local trends can statistically be tested.

Consider a probabilistic model in which attribute values y_i, $i=0, 1, \ldots, n+1$ are independently distributed according to the same probability density function $g(y)$. When the probability of inequation (such as $P\{y_{i-1}<y_i<y_{i+1}\}$), is of concern, $g(y)$ can be assumed as the uniform distribution,

$$g(y) = \begin{cases} 1 & \text{if } 0 \leq y \leq 1 , \\ 0 & \text{if otherwise} . \end{cases} \tag{8}$$

This assumption can be made without loss of generality because any probability density function of y' can be transformed into the uniform distribution by the probability integral transformation,

$$y = \int_{-\infty}^{y'} g(z)dz = G(y') \quad , \tag{9}$$

and the probability of inequation in y' is preserved in y by this transformation in the sense that

$$\begin{aligned} P\{y' \geq \bar{y}'\} &= \int_{-\infty}^{\bar{y}'} g(z)dz = G(\bar{y}') \\ &= P\{y \geq \bar{y} = G(\bar{y}')\} \quad . \end{aligned} \tag{10}$$

In the above probabilistic context a qualitative local trend t_i (which is determined by the random variables y_{i-1}, y_i, y_{i+1}) is a random variable and the number of peaks in $t=(t_1, t_2, \ldots, t_n)$ is also a random variable. Now suppose that an observed number $N(\Lambda)$ of peaks in $T=(t_1=T_1, t_2=T_2, \ldots, t_n=T_n)$ is a random realization of the random variable $n(\Lambda)$. Then randomness with respect to the number of peaks in T will statistically be tested when a certain significance level α is given. Stated explicitly if the inequation

$$P\{n(\Lambda) \geq N(\Lambda)\} = \sum_{m \leq N(\Lambda)} P\{n(\Lambda) = m\} \leq \alpha \quad , \tag{11}$$

or

$$P\{n(\Lambda) \leq N(\Lambda)\} = \sum_{m \geq N(\Lambda)} P\{n(\Lambda) = m\} \leq \alpha \tag{12}$$

holds, (where $P\{n(\Lambda)=m\}$ is the probability of the number of peaks being equal to m), then it is statistically rejected with significance level α that the appearance of peaks in the qualitative local trends T is random. In particular if equation (11) holds, the number of peaks is significantly many, and if equation (12) holds, the number of peaks is significantly few. It may hence be appropriate to say that the former case is peakish and the latter case is monotonous. To measure how peakish the qualitative local trends T are, it will be convenient to define the peakiness ρ by equation (12) or

$$\rho = P\{n(\Lambda) \leq N(\Lambda)\} . \tag{13}$$

Obviously the range of ρ is $0 \leq \rho \leq 1$ and the peakiness ρ increases as the number of peaks increases.

To employ the above statistical test, the probability function $P\{n(\Lambda)=m\}$ should be obtained. The calculation of this probability is not straightforward because of statistical dependence between t_i and t_j, $j=i-2, i-1, i+1, i+2$. (Observed that the event that $t_i=\Lambda$ and $t_{i+1}=\Lambda$ cannot occur, for if so, then $y_{i-1}<y_i$, $y_i>y_{i+1}$ and $y_i<y_{i+1}$, $y_{i+1}>y_{i+2}$, a contradiction.) It is noted, however, that t_i and t_j are statistically independent if

$i<j-2$ or $j>i+2$, and hence $t_1, t_4, \ldots, t_{3h}$, $(3h+1 \leq n)$ are statistically independent each other. With this property in mind, let us calculate $P\{n(\Lambda)=m\}$. The calculation is two-hold: First the conditional probability $P\{t_{3j-1}=T_{3j-1}, t_{3j}=T_{3j} | t_{3j-2}=T_{3j-2}, t_{3j+1}=T_{3j+1}\}$, (abbreviated to $P\{T_{3j-1}, T_{3j} | T_{3j-2}, T_{3j+1}\}$), is calculated; second by this conditional probability, the probability $P\{n_j=m\}$ of the number of peaks in $(t_1, t_2, \ldots, t_{3j-2}, t_{3j-1}, t_{3j})$ being m is expressed in terms of $P\{n_{j-1}=m\}$, and using this relation iteratively, $P\{n(\Lambda)=m\}$ is calculated.

Before calculating the conditional prbability $P\{T_2, T_3 | T_1, T_4\}$, (assuming j=0 without loss of generality), the following group-theoretic property should be noted to unburden the computational load. Let E, R_f and R_v be the functions of (T_1, T_2, T_3, T_4) defined by

$$E(T_1, T_2, T_3, T_4) = (T_1, T_2, T_3, T_4) \ , \tag{14}$$

$$R_f(T_1, T_2, T_3, T_4) = (sT_4, sT_3, sT_2, sT_1) \tag{15}$$

$$R_v(T_1, T_2, T_3, T_4) = (rT_1, rT_2, rT_3, rT_4) \ , \tag{16}$$

where operators s and r on T_i, i=1, 2, 3, 4, are defined by

$$s\Lambda = \Lambda, \ sV = V, \ s\diagdown = /, \ s/ = \diagdown \ ; \tag{17}$$

$$r\Lambda = V,\ rV = \Lambda,\ r\diagdown = /,\ r/ = \diagdown \ . \tag{18}$$

Second let $\bigcirc$ be the composition operator of functions, and R be the set of elements generated by the operation $\bigcirc$ on E, R_f and R_v. It can then be shown that the set R and the operator $\bigcirc$ form a group $<R, \bigcirc>$. To see the number of distinct elements in R, note that

$$R_v \bigcirc R_v = E,\ R_f \bigcirc R_f = E,\ R_f \bigcirc R_v = R_v \bigcirc R_f \ . \tag{19}$$

From this equation it is understood that the set R consists of four distinct elements, R_1, R_2, R_3 and R_4 that are respectively given by

$$R = \{E, R_v, R_f, R_v \bigcirc R_f\} \ . \tag{20}$$

Now we shall show that the probability $P\{T_1, T_2, T_3, T_4\}$ is invariant for $R_i \in R$. To show this intuitively let us look at Figure 1. In the figure, $(T_1, T_2, T_3, T_4) = (\diagdown \diagdown V /)$. If this figure is turned upside down, these local trends are written as $R_v(T_1, T_2, T_3, T_4) = (/ \Lambda \diagdown \diagdown)$. If Figure 1 is turned over, those local trends are written as $R_f(T_1, T_2, T_3, T_4) = (\diagdown V / /)$. If Figure 1 is turned over and upside down, those local trends are written as $R_v\ R_f(T_1, T_2, T_3, T_4) = (/ \Lambda \diagdown \diagdown)$. Since these four expressions describe the same event, it is understood that

$$P\{T_1, T_2, T_3, T_4\} = P\{R_i(T_1, T_2, T_3, T_4)\},\ R_i \in R\ . \tag{21}$$

Thus the probability $P\{R_i(T_1, T_2, T_3, T_4)\}$ is readily obtained once the probability $P\{T_1, T_2, T_3, T_4\}$ is calculated.

Keeping the above property in mind let us now calculate the conditional probability $P\{T_2, T_3 | T_1, T_4\}$. Since t_2 and t_3 are statistically dependent, (t_2, t_3) will have only two possible outcomes according to $y_2 > y_3$ or $y_2 < y_3$. For example, in the case of Figure 2 where $t_1 = \backslash$ and $t_4 = \Lambda$, the possible outcomes are: $(t_2, t_3) = (\backslash, V)$ (See Panel a)); $(t_2, t_3) = (V, /)$ (See Panel b)). The possible outcomes of the other cases are listed in Table 1.

Figure 2. An illustrative example of dependency between local trends t_2 and t_3 provided that $t_1 = \backslash$ and $t_4 = \Lambda$.

The probability of those outcomes is obtained from the probability density function $f(y_2 | T_1)$ and $f(y_3 | T_4)$. In the case of $f(y_2 | \Lambda)$, (as is shown in the example of Figure 1), $f(y_2 | \Lambda)$ is written as

$$f(y_2 | \Lambda) = f(y_2 | y_0 < y_1,\ y_1 > y_2)\ /\ P\{t_1 = \Lambda\}\ . \tag{22}$$

Since

$$P\{t_1=\Lambda\} = P\{t_1=V\} = \int_0^1 \int_0^{y_1} \int_0^{y_1} dy_0\, dy_1\, dy_2 \tag{23}$$

$$= 1/3 \ ,$$

$$P\{t_1=/\} = P\{t_1=\backslash\} = \int_0^1 \int_0^{y_1} \int_{y_1}^1 dy_0\, dy_1\, dy_2 \tag{24}$$

$$= 1/6 \ ,$$

equation (22) is written as

$$f(y_2|\Lambda) = 3 \int_{y_2}^1 \int_0^{y_1} dy_0\, dy_1 = 3(1 - y_2^{\,2})/2 \ . \tag{25}$$

Replacing y_2 by $1-y_2$ gives

$$f(y_2|V) = 3\, y_2\, (2 - y_2) \,/\, 2 \ . \tag{26}$$

In the similar way it is obtained that

$$f(y_2|\backslash) = 3y_2^{\,2} \ , \tag{27}$$

$$f(y_2|/) = 3\, (1 - y_2)^2 \ . \tag{28}$$

Concerning $f(y_3|T_4)$, the probability density function is readily obtained from the equation

$$f(y_3|T_4) = f(y_2|sT_1) \quad , \tag{29}$$

since equation (21) holds. The conditional probability $P\{T_2, T_3|T_1, T_4\}$ is therefore obtained from

$$P\{y_2>y_3|T_1, T_4\} = \int_0^1 \int_0^{y_2} f(y_2|T_1) f(y_3|T_4) \, dy_2 \, dy_3 \quad , \tag{30}$$

or

$$P\{y_2<y_3|T_1, T_4\} = 1 - P\{y_2>y_3|T_1, T_4\} \quad , \tag{31}$$

where $f(y_2|T_1)$ and $f(y_2|T_4)$ are given by equations (25), (26), (27) and (28). For example, in the case of Figure 2,

$$\begin{aligned} P\,\{\diagdown, V|\diagdown, \Lambda\} &= \int_0^1 \int_0^y 9\, y_2^2\, (1 - y_3^2)\, dy_2\, dy_3 \,/\, 2 \\ &= 13/20 \quad , \end{aligned} \tag{32}$$

$$P\,\{V, \diagup|\diagdown, \Lambda\} = 7/20 \quad . \tag{33}$$

The conditional probabilities of the other cases are tabulated in Table 1 where figures such as $\diagdown$ V (1/2) show that $t_2=\diagdown$, $t_3=V$ and $P\{\diagdown, V|\Lambda, \Lambda\}=1/2$. From this table the probability of the number of peaks in (t_1, t_2, t_3) provided that $t_1=T_1$, $t_4=T_4$ is calculated and tabulated in Table 2 where figures such as 1 (19/80) indicate the probability of one peak being 19/80.

Table 1. The conditional probability of $t_2=T_2$ and $t_3=T_3$ provided that $t_1=T_1$ and $t_4=T_4$.

Table 2. The conditional probability of the number of peaks in t_2 and t_3 provided that $t_1=T_1$ and $t_4=T_4$.

Having finished the first step, let us next get into the second step which is to obtain the probability $P\{n_j=m|t_{3j+1}=T_{3j+1}\}$ of the number of peaks in $(t_1, t_2, \ldots, t_{3j-2}, t_{3j-1}, t_{3j})$ being m provided that $t_{3j+1}=T_{3j+1}$. This probability $P\{n_j=m|t_{3j+1}=T_{3j+1}\}$ will be obtained from $P\{n_{j-1}=m'|t_{3j-2}=T_{3j-2}\}$ by iterative processes, j=1, 2, ... For illustrative purposes the case of $t_{3j+1}=\Lambda$ is first considered. According to $t_{3j-2}=T_{3j-2}$, the number of peaks changes from $n_j=m-2$, m-1 or m to $n_j=m$ with a certain probability. For example, if $t_{3j-2}=\Lambda$, then $n_{j-1}=m-1$ becomes $n_j=m$ with probability 1. (See Table 2.) The other cases are tabulated in Table 3. From this table and equations (23) and (24), the conditional probability $P\{n_j=m|t_{3j+1}=\Lambda\}$ is written as

Table 3. The conditional probability that n_{j-1} becomes $n_j=m$ provided that $t_{3j-2}=T_{3j-2}$.

$$\begin{aligned} &P\{n_j = m \mid t_{3j+1} = \Lambda\} \\ &= P\{n_{j-1}=m-1 \mid t_{3j-2}=\Lambda\} / 3 + 61P\{n_{j-1}=m-1 \mid V\} / 240 \\ &+ 19P\{n_{j-1}=m \mid V\} / 240 + P\{n_{j-1}=m \mid \diagdown\} / 6 \\ &+ 7P\{n_{j-1}=m-1 \mid /\} / 48 + P\{n_{j-1}=m \mid /\} / 48 . \end{aligned} \tag{34}$$

(Note that for bevity t_{3j-2} is omitted from the second term of the right hand side.) In the same way it is obtained that:

$$\begin{aligned} &P\{n_j = m \mid t_{3j+1} = V\} \\ &= 61P\{n_{j-1}=m-2 \mid t_{3j-2}=\Lambda\} / 240 + 19P\{n_{j-1}=m-1 \mid \Lambda\} / 240 \\ &+ P\{n_{j-1}=m-1 \mid V\} / 3 + 7P\{n_{j-1}=m-1 \mid \diagdown\} / 48 \\ &+ P\{n_{j-1}=m \mid \diagdown\} / 48 + P\{n_{j-1}=m-1 \mid /\} / 6 , \end{aligned} \tag{35}$$

$$\begin{aligned} &P\{n_j = m \mid t_{3j+1} = \diagdown\} \\ &= 7P\{n_{j-1}=m-2 \mid t_{3j-2}=\Lambda\} / 24 + P\{n_{j-1}=m-1 \mid \Lambda\} / 24 \\ &+ P\{n_{j-1}=m-1 \mid V\} / 3 + 19P\{n_{j-1}=m-1 \mid \diagdown\} / 120 \\ &+ P\{n_{j-1}=m \mid \diagdown\} / 120 + P\{n_{j-1}=m-1 \mid /\} / 6 , \end{aligned} \tag{36}$$

$$\begin{aligned} &P\{n_j = m \mid t_{3j+1} = /\} \\ &= P\{n_{j-1}=m-1 \mid t_{3j-2}=\Lambda\} / 3 + 7P\{n_{j-1}=m-1 \mid V\} / 24 \\ &+ P\{n_{j-1}=m \mid V\} / 24 + P\{n_{j-1}=m \mid \diagdown\} / 6 \\ &+ 19P\{n_{j-1}=m-1 \mid /\} / 120 + P\{n_{j-1}=m \mid /\} / 120 . \end{aligned} \tag{37}$$

Since the initial value j=0 is obtained from Table 2 as Table 4, the probability $P\{n_j=m|t_{3j+1}=T_{3j+1}\}$ can be obtained from equations (34), (35), (36) and (37) by iterative processes j=1, 2, ... Therefore the probability of the number of peaks in (t_1, t_2, ..., $t_{n=3h+1}$) being m is written as

$$P\{n(\Lambda) = m\} = P\{n_h = m - 1|t_{3h+1} = \Lambda\} / 3$$
$$+ P\{n_h = m|t_{3h+1} = V\} / 3 + P\{n_h = m|t_{3h+1} = \diagdown\} / 6 \qquad (38)$$
$$+ P\{n_h = m|t_{3h+1} = /\} / 6 .$$

Table 4. The conditional probability of the number n, of peaks in t_1, t_2, t_3 being m provided that t_4=T_4.

With the above statistical method, let us now test whether the qualitative local trends of the secondary establishment ratio on the Chuo Railway Line Area are peakish or monotonous. From equations (34), (35), (36), (37) and the initial values given by Table 4, the probability $P\{n_2=m|t_7=T_7\}$ is calculated and listed in Table 5, from which Table 6 is obtained where the probability $P\{n_3=m|t_{10}=T_{10}\}$ is shown. Substitution of these values into equation (38) gives the probability function $P\{n(\Lambda)=m\}$ of the number of peaks in t_1, t_2, ..., t_{10} which is shown in Table 7. In observing Figure 1 or equation (4), it is noticed that two peaks exist, i.e., $N(\Lambda)=2$. Since the probabilities $P\{n(\Lambda)\geq 2\}$

and $P\{n(\Lambda)\leq 2\}$ are respectively calculated from Table 7 as .1213 and .9957, both of which are larger than .05, it is concluded with .95 confidence level that the qualitative local trends of the secondary establishment ratio on the Chuo Railway Line Area are neither peakish nor monotonous. Stated differently it cannot be rejected with significance level α=.05 that the number of peaks in the local trends appears randomly according to the probabilistic model stated in the above.

Table 5. The conditional probability of the number n_2 of peaks in t_1, t_2, ..., t_6 being m provided that $t_7=T_7$.

Table 6. The conditional probability of the number n_3 of peaks in t_1, t_2, ..., t_9 being m provided that $t_9=T_9$.

Table 7. The probability that the number of peaks in t_1, t_2, ..., t_{10} is m.

Before closing this section, let us refer to the case of a large number n of local trends. In this case the calculation of $P\{n(\Lambda)=m\}$ becomes laborious. To avoid this difficulty the following less powerful but more practical method can be used.

Recalling that the random variables in $t_a=(t_1, t_4, \ldots, t_{3h-2})$, $t_b=(t_2, t_5, \ldots, t_{3h-1})$ or $t_c=(t_3, t_6, \ldots, t_{3h})$, (assuming n=3h) are statistically independent, it is readily understood that the number $n_k(\Lambda)$ of peaks in t_k, k=a, b, c, being m is given by the binomial distribution

$$P\,\{n_k(\Lambda) = m\} = \binom{h}{m}\, P\,\{t = \Lambda\}^m (1 - P\,\{t = \Lambda\})^{h-m}$$

$$= \binom{h}{m} (1/3)^m\, (2/3)^{h-m}\,. \tag{39}$$

It is well known that if h is large, as is supposed in this case, the binomial distribution is approximated by the normal distribution. Therefore if the inequation

$$F(N_k'(\Lambda)) \le \alpha \quad \text{or} \quad 1 - F(N_k'(\Lambda)) \le \alpha\,, \quad k=a,\, b,\, c\,, \tag{40}$$

holds, where $N_k'(\Lambda)=(N_k(\Lambda) - h/3) / \sqrt{2h/9}$ and F is the cumulative distribution function of the standard normal distribution, then it is statistically rejected with significance level α that appearance of peaks is random. Finally it should be noted that compared with the exact test given by equation (38), this alternative test is less powerful because this test is based upon the information T_k which is one-third of T.

(3) A STATISTICAL METHOD OF THE GLOBAL TREND CURVE ANALYSIS

As is seen in the above section the local trend curve analysis does not pay attention to the arrangement of peaks on L. Hence the conclusion that the appearance of peaks in the local trends is random does not imply that the arrangement of peaks is random. To consider this problem, the global trend curve analysis is to be developed here.

Randomness in the arrangement of peaks may be tested by the run test. If should be noted, however, that a slight modification is necessary because of a specific nature of a run of peaks, that is, the length of a peak run is always one since $t_i=\Lambda$ and $t_{i+1}=\Lambda$ cannot occur in $T=(T_1, T_2, \ldots, T_n)$. To employ the ordinary run test, an alternative definition is necessary: a <u>run of peaks</u> is a sequence of peaks given by $t_i=\Lambda$, $t_{i+2}=\Lambda$, ..., $t_{i+2m}=\Lambda$. This sequence is written as $t'_i=\Lambda$, $t'_{i+1}=\Lambda$, ..., $t'_{i+m}=\Lambda$, if t_{i+2j} and t_{i+2j+1} are grouped in one t'_{i+j} or t_{i+2j+1} is discarded. By this modification, T is transformed into $T'=(t'_1=T'_1, t'_2=T'_2, \ldots, t'_{n'}=T'_{n'})$ in which t'_i is labelled a peak Λ or a non-peak $\bar{\Lambda}$. For example, in Figure 1, since t_2 and t_3 are grouped in t'_2, and t_6 and t_7 are grouped in t'_5, T' is written as

$$T' = (\bar{\Lambda}\ \Lambda\ \bar{\Lambda}\ \bar{\Lambda}\ \Lambda\ \bar{\Lambda}\ \bar{\Lambda}\ \bar{\Lambda}) \qquad (41)$$

Once the above modification is made, the ordinary run test can be applied to T'. Since the number of peaks in T' is given by $N(\Lambda)$ and that of non-peaks is given by $n'-N(\Lambda)$, the probability $P\{z=k\}$ of the number z of runs of either kind being exactly k is given by

$$P\{z=k\} = \begin{cases} 2\binom{N(\Lambda)-1}{u-1}\binom{n'-N(\Lambda)-1}{u-1} \Big/ \binom{n'}{N(\Lambda)} & \text{if } k=2u \\ \left\{\binom{N(\Lambda)-1}{u}\binom{n'-N(\Lambda)-1}{u-1} + \binom{N(\Lambda)-1}{u-1}\binom{n'-N(\Lambda)-1}{u}\right\} & \\ \Big/ \binom{n'}{N(\Lambda)}, & \text{if } k=2u+1 \end{cases} \tag{42}$$

(See Feller (1950), p. 62). If the inequation

$$\sum_{k\geq\bar{z}} P\{z=k\} \leq \alpha \tag{43}$$

or

$$\sum_{k\leq\bar{z}} P\{z=k\} \leq \alpha \tag{44}$$

holds, (where $\bar{z}$ is the observed number of runs and α is a given significance level), then randomness in the arrangement T' of peaks is rejected with significance level α. In particular, if inequation (43) holds, the arrangement T' of peaks is significantly concentrated; if inequation (44) holds, the arrangement T' of peaks is significantly evenly spaced. To see the magnitude of concentration of peaks, it will be useful to define the peak concentration γ by the left hand side of inequation (43), i.e.,

$$\gamma = \sum_{k \geq \bar{z}} P \{ z = k \} . \tag{45}$$

Obviously the inequation $0 \leq \gamma \leq 1$ holds and the arrangement of peaks in T' is concentrated if $\gamma \leq \alpha$ and T' is evenly spaced if $1 - \gamma + P\{z=\bar{z}\} \leq \alpha$.

Let us now apply the above global trend curve analysis to the example in Figure 1. As is seen in equation (41), the observed number $\bar{z}$ of runs is five. Substitution of $N(\Lambda)=2$, $n'=8$, $\bar{z}=5$ into equation (42) gives the probability function of the number of runs, from which the peak concentration γ is obtained as $\gamma=.3571$. If the significance level α is fixed .05, it is statistically concluded with .95 confidence that the arrangement of peaks in Figure 1 is neither concentrated nor evenly spaced.

Finally it should be noted that if the number of peaks and that of non-peaks are both 10 or more, the distribution of a random variable z' defined by

$$z' = \sqrt{\frac{n' - 1}{(2N(\Lambda)n' - 2N(\Lambda)^2)\ (2N(\Lambda)n' - 2N(\Lambda)^2 - n')}} \times (n'z - 2N(\Lambda)n' + 2N(\Lambda)^2 + n') \tag{46}$$

can be approximated by the standard normal distribution. (See Brunk (1965), p. 357.) Hence inequations (43) and (44) can be replaced by

$$\gamma = F(\bar{z}') \leq \alpha \qquad \text{or} \qquad 1 - F(\bar{z}') \leq \alpha\ , \qquad (43)',\ (44)'$$

where $\bar{z}'$ is given by equation (46) with $z=\bar{z}$, (i.e., the observed number of runs).

(4) CORRELATION ANALYSIS OF TWO TREND CURVES

In the preceeding sections, the analysis was focussed upon only one trend curve. In the trend curve analysis, however, the comparative analysis of two trend curves, $y_{1i}=f(i)$ and $y_{2i}=f(i)$, $i=0, 1, \ldots, n+2$, is also important. For example it is of interest to examine similarity (or dissimilarity) between the trend curve of the construction establishment ratio (Figure 5a) and that of the manufacturing establishment ratio (Figure 5b). Since relative behavior of two trend curves can be described by a set $\{(y_{1i}, y_{2i})\}$ of coordinates (y_{1i}, y_{2i}), $i=0, 1, \ldots, n+2$, the correlation coefficient

$$R = \frac{\Sigma\ y_{1i}\ y_{2i}}{\sqrt{\Sigma\ y_{1i}{}^2\ \Sigma\ y_{2i}{}^2}} \tag{47}$$

can be used as one of the simplest measures of similarity between two trend curves. (Not that $\Sigma y_{1i}=\Sigma y_{2i}=0$ is assumed in equation (47).)

Alternatively if attention is paid to a qualitative nature of local trends, relative behavior of two trend curves can be described by a set $\{(t_{1i},\ t_{2i})\}$ of coordinates $(t_{1i},\ t_{2i})$, $i=1,\ 2,\ \ldots,\ n$, where t_{ji} is a qualitative local trend of y_j at $x=i$, $j=1,\ 2$. Then, like the case of $\{(y_{1i},\ y_{2i})\}$, a "qualitative" correlation coefficient defined by $r = \Sigma t_{1i} t_{2i}\ /\ \sqrt{\Sigma t_{1i}{}^2\ \Sigma t_{2i}{}^2}$ may be useful to see qualitative similarity between two trend curves. It should be noted, however, that this correlation coefficient does not make sense if the ordinary algebraic operation is assumed. To draw a meaningful implication from r, the multiplication operator on t_{1i} and t_{2i} should newly be defined. To do so let us first classify a relation between t_{1i} and t_{2i} according to the "same" relation, the "reverse" relation or the "independent" relation. Obviously the <u>same relation</u> implies $t_{1i}=t_{2i}$, for example, $t_{1i}=t_{2i}=\Lambda$. The <u>reverse relation</u> is naturally defined by t_{1i} and $t_{2i}=rt_{1i}$ where the operator r is defined in Section 2. Namely the reverse relation of Λ is V and that of $\setminus$ is /, and vice versa. The independent relation is defined by a relation in which either $(y_{1i-1}-y_{1i})$ and $(y_{2i-1}-y_{2i})$, or $(y_{1i}-y_{1i+1})$ and $(y_{2i}-y_{2i+1})$ has the same sign and the other has the oposit sign. For example $t_{1i}=\Lambda$ and $t_{2i}=/$, where $y_{1i-1}<y_{1i}$, $y_{2i-1}<y_{2i}$ but $y_{1i}>y_{1i+1}$, $y_{2i}<y_{2i+1}$. Since the same and reverse relations both exist in Λ and $/$, it appears to be natural say that Λ and $/$ are independent. With those relations let us next define the multiplication operator $\otimes$ on t_{1i} and t_{2i}. If a relation between t_{1i} and t_{2i} is the same, a positive value, say 1, is given to $t_{1i}\otimes t_{2i}$:

$$\Lambda \otimes \Lambda = V \otimes V = \diagdown \otimes \diagdown = / \otimes / = 1 . \tag{48}$$

If the relation is reverse, a negative value -1 is given:

$$\Lambda \otimes V = V \otimes \Lambda = \diagdown \otimes / = / \otimes \diagdown = -1 . \tag{49}$$

If the relation is independent, $t_{1i}\otimes t_{2i}=0$:

$$\begin{aligned} \Lambda \otimes \diagdown &= \Lambda \otimes / = V \otimes \diagdown = V \otimes / \\ &= \diagdown \otimes \Lambda = / \otimes \Lambda = \diagdown \otimes V = / \otimes V = 0 \end{aligned} \tag{50}$$

The multiplication operation defined above now makes the qualitative correlation coefficient r meaningful. To be explicit, the <u>qualitative correlation coefficient</u> r is defined by

$$r = \frac{\sum_{i=1}^{n} t_{1i} \otimes t_{2i}}{\sqrt{(\sum_{i=1}^{n} t_{1i} \otimes t_{1i})\ (\sum_{i=1}^{n} t_{2i} \otimes t_{2i})}} = \frac{\sum_{i=1}^{n} t_{1i} \otimes t_{2i}}{n} . \qquad (51)$$

In the example referred to in the above, (Figures 5a and 5b), the qualitative local trends of the construction establishment ratio and those of the manufacturing establishment ratio are listed in the first and second row of Table 8, from which $\Sigma t_{1i} \otimes t_{2i}$ is obtained as 3. Therefore the qualitative correlation coefficient r_{12} between those two trend curves is given by $r_{12}=.3$. This implies that the local trends of the construction establishment ratio and those of the manufacturing establishment ratio are weakly similar.

It should be noted that the ordinary correlation coefficient R and the qualitative correlation coefficient r are complementary for the trend curve analysis. To see this clearly, Figure 3 is depicted, from which R and r are respectively calculated as -.83 and 1. At first glace these results appear to be contradictory but they are not so. From Figure 3 it is noticed that the qualitative correlation coefficient r focuses on each local trends, whereas the ordinary correlation coefficient R focuses on a global trend. Actually in each local trend the reverse relation is observed; at the same time it is observed that one trend curve is globally decreasing, while the other is globally increasing. Therefore in the trend curve analysis, the qualitative correlation coefficient r can be used to see similarity of local trends and the ordinary correlation coefficient can be used to see similarity of global trends.

Figure 3. An example of two trend curves whose correlation is globally positive but locally negative.

(5) THE QUALITATIVE TREND CURVE ANALYSIS OF THE ESTABLISHMENT RATIOS ON THE CHUO RAILWAY LINE AREA

Having formulated a statistical method of the qualitative trend curve analysis, let us now apply it to the study of a spatial distribution of establishments on the Chuo Railway Line Area. The study area L is a 2 km x 24 km strip area from Shinjuku (one of the subcenters in Tokyo) to Tachikawa (one of the suburban centers in Tokyo) where a streight railway line, called the Chuo Railway Line, goes through. (See a map in Figure 4.) This area is divided into twelve 2 km x 2 km quadrates, L_1, L_2, ..., L_{12}, which are used as data units. Six kinds of attribute values are used for the analysis: y_{1i} = a ratio of the number of construction establishments to the total number of establishments in the i^{th} quadrat; y_{2i} = that of manufacturing establishments; y_{3i} = that of wholesale and retail establishments; y_{4i} = that of finance and insurance establishments; y_{5i} = that of real estate establishments; y_{6i} = that of service establishments. (For brevity y_{1i} is called

Figure 4. The Chuo Railway Line Area and the 60 km x 60 km square region centered at the Tokyo Station.

the construction establishment ratio, y_{2i} is called the manufacturing establishment ratio, and so forth.) The trend curve y_{j0}, y_{j1}, ..., y_{j11} of the j^{th} establishment ratio, j=1, 2, ..., 6, is depicted in Figures 5a-5f, (Japanese Bureau of Statistics (1975)), from which the qualitative local trends (T_{j1}, T_{j2}, ..., T_{j10}) are obtained as the j^{th} row of Table 8.

Figure 5a. The construction establishment ratio on the Chuo Railway Line Area, 1970.

" 5b. The manufacturing establishment ratio.

" 5c. The wholesale and retail establishment ratio.

" 5d. The finance and insurance establishment ratio.

Figure 5e. The real estate establishment ratio.

" 5f. The service establishment ratio.

Table 8. The sequences qualitative local trends of the establishment ratios on the Chuo Railway Line Area.

By consulting Table 7, the peakiness ρ_j of T_j is calculated and listed in the last column of Table 8. It can first be read from this table that the trend curve of the wholesale and retail establishment ratio has the highest peakiness, $\rho_3=.9540$; that of manufacturing establishment ratio shows the lowest peakiness, $\rho_2=.1213$; the peakiness of the other establishment ratio is the same. This fact may imply that the wholesale and retail establishments are more likely to have subcenters than the manufacturing establishments on the Chuo Railway Line Area. Second, since $P\{n(\Lambda)\leq 2\}=.1213$ and $P\{n(\Lambda)\geq 4\}=.4132$, (equations (11) and (12)), it is statistically concluded with .95 confidence level that the number of peaks observed in each trend curve is neither significantly many nor few. Stated differently all the trend curves are neither peakish nor monotonous.

Next to examine a global trend of the establishment ratios, a sequence T_j of qualitative local trends is transformed into T_j' by the procedure stated in Section 3. Those modified sequences T_j' are shown in Table 9, from which the number $\bar{z}$ of runs is counted. Substitution of this number into equation (45) gives the peak concentration γ_j of each trend curve.

Table 9. The modified sequences of the qualitative local trends of Table 8.

It can first be read from the last two columns of Table 9 that the trend curve of the construction establishment ratio has the highest peak concentration, ρ_1=1.0000 and that of the finance and insurance establishment ratio has the lowest peak concentration, ρ_4=.2000; the peak concentration is fairly high in the trend curve of the real estate establishment ratio, ρ_5=.8571, and it is fairly low in that of the service establishment ratio, ρ_6=.2857. Second it cannot be rejected with significance level α=.06 that the peaks in the trend curve of the construction establishment ratio are randomly arranged. Stated differently it is concluded with .94 confidence level that the arrangement of peaks in the trend curve of the construction establishment ratio is concentrated. This may imply that the subcenters of the construction establishments are gathered on the Chuo Railway Line Area. Last it can be said with .94 confidence level that peaks in all the trend curves except the trend curve of the construction establishment ratio are neither concentrated nor evenly spaced. This may imply that a distinctive regularity does not exist in the configuration of subcenters of those kinds of establishments.

Concerning the correlation analysis, the correlation coefficient R_{ij} and the qualitative correlation coefficient r_{ij} are calculated and tabulated in Table 10. It is noticed from this table that the qualitative local trends of the wholesale and retail establishment ratio are reversely related to those of the other establishment ratios ($r_{3j}<0$, j=1, 2, 4, 5, 6). In particular a high reverse relation is observed between the qualitative local trends of the wholesale and retail establishment ratio and those of the service establishment ratio ($r_{36}=-.7$). This may imply that subcenters of the wholesale and retail establishments and those of the service establishments appear alternately. The highest positive qualitative correlation is seen between the qualitative local trends of the construction establishment ratio and those of the manufacturing establishment ratio although the absolute value $r_{12}=.4$ is not very high. Since almost all qualitative correlation coefficients are within $-.5\leq r_{ij}\leq .5$ except $r_{36}=-.7$, it may be said that neither strongly similar local trends nor strongly reverse local trends do not exist between any two trend curves except between the trend curve of the wholesale and retail establishment ratio and that of the service establishment ratio.

Concerning R_{ij}, a fairly high positive correlation is seen between the trend curve of the construction establishment ratio and that of the wholesale and retail establishment ratio (R_{12}=.67). This may imply that those two trend curves are globally similar to each other. Like the case of r_{ij}, almost all correlation coefficients are within $-.5 \leq R_{ij} \leq .5$ except R_{12}=.67. Hence a strong global similarity does not exist between any two trend curves except R_{12}.

Table 10. The correlation coefficients (upper figures) and the qualitative correlation coefficients (lower figures) between two trend curves.

Last to see the relation between the local correlation and global correlation, Figure 6 is depicted, in which coordinates (r_{ij}, R_{ij}), $i<j$, are indicated by points labelled "i-j". In this figure points 1-2, 3-6, 2-3 and 1-4 appear distinctive. The point 1-2 shows that the trend curve of the construction establishment ratio is not only locally but also globally similar to that of the manufacturing establishment ratio, ($R_{12}>0$, $r_{12}>0$). The point 2-3 is completely opposite to the point 1-2. The reverse correlation is found between the trend curve of the manufacturing establishment ratio and that of the wholesale and retail establishment ratio not only in the local trends but also in the global trends, ($R_{23}<0$, $r_{23}<0$). The points 3-6 and 1-4 show a difference between the local correlation and the global correlation. The point 3-6 shows that the local trends of the wholesale and retail establishment ratio are reversely correlated to those of the service establishment ratio but the global trend of the former is positively correlated to that of the latter, ($R_{36}>0$, $r_{36}<0$). The opposite phenomena are observed between the trend curve of the construction establishment ratio and that of the service establishment ratio, ($R_{16}<0$, $r_{16}>0$).

Figure 6. Correlation coefficients R_{i}, and r_{ij} between the i^{th} establishment ratio and the j^{th} establishment ratio.

The trend curve analysis is originally developed for the analysis of a trend with respect to one parameter, $y_i=f(i)$. Because of this restriction one might consider that the trend curve analysis cannot be applied to the analysis of a trend surface, $y_{ij}=f(i, j)$. It should be noted, however, that as a simple method, the trend curve analysis can be used for the trend surface analysis by cutting a surface into thin slices (see Figure 7) and joining these slices into one curve. That is, a surface y_{ij}, $i=1, 2, \ldots, n_1$, $j=1, 2, \ldots, n_2$, is transformed into a curve by

$$y_{11},\ y_{12},\ \ldots,\ y_{1n_2},\ y_{2n_2},\ y_{2n_2-1},\ \ldots,\ y_{21},\ y_{31},\ \ldots,\ y_{n_1n_2}\ , \quad (52)$$

or

$$y_{11},\ y_{21},\ \ldots,\ y_{n_11},\ y_{n_12},\ y_{n_1-12},\ \ldots,\ y_{12},\ y_{13},\ \ldots,\ y_{n_1n_2}\ . \quad (53)$$

If a trend surface is randomly constructed, (that is, variables y_{ij} are randomly distributed according to the same uniform distribution), the trend curve given by equation (52) or (53) behaves randomly. Hence randomness in the trend surface can be examined by the test developed in the trend curve analysis. That is, if randomness in the trend curve $y_{11}, y_{12}, \ldots, y_{n_1n_2}$ is rejected, randomness in the trend surface y_{ij}, $i=1, 2, \ldots, n_1$, $j=1, 2, \ldots, n_2$ is rejected.

As an empirical example let us consider Figure 7 which shows the construction establishment ratio in the 60 km x 60 km square region centered at the Tokyo Station. Since the number of quardrates is large, the peakiness is examined by equation (40). The number of peaks is observed as $N_a(\Lambda)=59$, $N_b(\Lambda)=50$, $N_c(\Lambda)=51$, from which $N'_a(\Lambda)=.1581$, $N'_b(\Lambda)=1.5811$, $N'_c(\Lambda)=1.4230$. Hence the peakiness is obtained as $\rho_a=.5636$, $\rho_b=.9429$, $\rho_c=.9222$. Since any of ρ_k, $k=a, b, c$, is larger than .95, it cannot be rejected with significance level $\alpha=.05$ that the appearance of peaks is random, although the peakiness is fairly high. Concerning the global trend analysis, the number of runs is calculated as z=266. Upon substituting $n'=379$ and $N(\Lambda)=160$ into equation (46), z' is obtained as 8.45, from which the peak concentration γ is calculated as 1.00. It is hence concluded with confidence level .95 that the arrangement of peaks is not random.

Figure 7. The construction establishment ratio in the 60 km x 60 km square region centered at the Tokyo Station, 1970.

(6) A CONCLUDING REMARK

To close this paper let us briefly remark a note for a further development. In Section 5 it is shown that the qualitative trend curve analysis can be applied to the trend surface analysis by the transformation given by equation (53). It should be noted, however, that this application has some limitations although it provides a convenient method for testing randomness in a trend surface. Obviously a peak in a surface y_{ij}, $i=1, 2, \ldots, n_1$, $j=1, 2, \ldots, n_2$, is a peak in the curve $y_{11}, y_{12}, \ldots, y_{n_1 n_2}$, but it is not always true that a peak in the curve is a peak in the surface. Hence the peakiness of $y_{11}, y_{12}, \ldots, y_{n_1 n_2}$ becomes ambiguous in the context of y_{ij}, $i=1, 2, \ldots, n_1$, $j=1, 2, \ldots, n_2$. To clarify this ambiguity, a qualitative local trend should be defined by $y_{i-1j}, y_{ij}, y_{i+1j}, y_{ij-1}, y_{ij+1}$.

Then paralleling the qualitative trend curve analysis developed in this paper, the qualitative trend surface analysis can be developed. It is hoped that this subject will be taken up in a subsequent paper.

REFERENCES

Bennett, R. J., (1979), Spatial Time Series, Academic Press (New York).

Brunk, H. D., (1965), An Introduction to Mathematical Statistics, Blaisdell (New York).

Cliff, A. D., Haggett, P., Ord, J. K., Bassett, K., Davies, R., (1975), Elements of Spatial Structure, Cambridge University Press (Cambridge).

Feller, W., (1950), An Introduction to Probability Theory and Its Applications, Volume 1, (Third Edition), Wiley (New York).

Ishimizu, T., (1976), An Outline of Quantitative Geography, Kokon-Shoin (Tokyo).

Japanese Bureau of Statistics, (1970), Establishment Census, (Regional Mesh Data Statistics), Office of the Prime Minister (Tokyo).

King, L. J., (1969), Statistical Analysis in Geography, Prentice-Hall (Englewood Cliffs, N. J.).

t_1 \ t_4	Λ	V	\	/
Λ	\ V (1/ 2)	V Λ (61/80)	V Λ (7/ 8)	\ / (13/20)
	V / (1/ 2)	\ \ (19/80)	\ \ (1/ 8)	V / (7/20)
V	Λ V (61/80)	/ Λ (1/ 2)	/ Λ (13/20)	Λ V (7/ 8)
	/ / (19/80)	Λ \ (1/ 2)	Λ \ (7/20)	/ / (1/ 8)
\	V / (13/20)	V Λ (7/ 8)	V Λ (19/20)	V / (1/ 2)
	\ V (7/20)	\ \ (1/ 8)	\ \ (1/20)	\ V (1/ 2)
/	Λ V (7/ 8)	Λ \ (13/20)	Λ \ (1/ 2)	Λ V (19/20)
	/ / (1/ 8)	/ Λ (7/20)	/ Λ (1/ 2)	/ / (1/20)

Table 1. The conditional probability of $t_2=T_2$ and $t_3=T_3$ provided that $t_1=T_1$ and $t_4=T_2$.

t_1 \ t_4	Λ	V	\	/
Λ	1 (1)	2 (61/80) 1 (19/80)	2 (7/ 8) 1 (1/ 8)	1 (1)
V	1 (61/80) 0 (19/80)	1 (1)	1 (1)	1 (7/ 8) 0 (1/ 8)
\	0 (1)	1 (7/ 8) 0 (1/ 8)	1 (19/20) 0 (1/20)	0 (1)
/	1 (7/ 8) 0 (1/ 8)	1 (1)	1 (1)	1 (19/20) 0 (1/20)

Table 2. The conditional probability of the number of peaks in t_2 and t_3 provided that $t_1=T_1$ and $t_4=T_4$.

t_{3j-2}	n_{j-1}	n_j	Probability
Λ	m − 1	m	1
V	m − 1	m	61/80
	m	m	19/80
╲	m	m	1
/	m − 1	m	7/ 8
	m	m	1/ 8

Table 3. The conditional probability that n_{j-1} becomes n_j=m provided that $t_{3j-2}=T_{3j-2}$, (n_j = the number of peaks in $t_1, t_2, \ldots, t_{3j}$).

n_1 = m	t_4			
	Λ	V	╲	/
0	.2667	.0208	.0083	.2167
1	.7333	.7250	.7000	.7833
2		.2542	.2917	

Table 4. The conditional probability of the number of peaks in t_1, t_2, t_3 being m provided that $t_4=T_4$.

n_2 = m	t_7			
	Λ	V	╲	/
0	.0075	.0002	.0001	.0041
1	.3162	.0800	.0613	.2827
2	.6117	.6062	.5938	.6391
3	.0646	.3136	.3448	.0741

Table 5. The conditional probability of the number n_2 of peaks in $t_1, t_2, \ldots, t_6$ being m provided that $t_7=T_7$.

$n_3 = m$	t_{10}			
	Λ	V	\	/
0	.0001	.0000	.0000	.0001
1	.0256	.0026	.0016	.0191
2	.3272	.1220	.1038	.3030
3	.5350	.5312	.5232	.5531
4	.1121	.3278	.3526	.1247
5		.0164	.0188	

Table 6. The conditional probability of the number n_3 of peaks in $t_1, t_2, \ldots, t_a$ being m provided that $t_9 = T_7$.

$n = m$	$P\{n = m\}$	$\sum_{i \leq m} P\{n = j\}$
0	.0000	.0000
1	.0043	.0043
2	.1170	.1213
3	.4655	.5868
4	.3672	.9540
5	.0460	1.0000

Table 7. The probability that the number n of peaks in $t_1, t_2, \ldots, t_{10}$ is m.

Establishments	1	2	3	4	5	6	7	8	9	10	Number of Peaks	Peakiness ρ_j
Construction T_{1i}	$\backslash$	$\vee$	$/$	$/$	Λ	$\vee$	Λ	$\vee$	Λ	$\backslash$	3	.5868
Manufacturing T_{2i}	$\backslash$	$\backslash$	$\vee$	$/$	Λ	$\vee$	$/$	$/$	$/$	Λ	2	.1213
Wholesale & Retail T_{3i}	Λ	$\vee$	Λ	$\backslash$	$\vee$	Λ	$\vee$	$/$	Λ	$\backslash$	4	.9540
Finance & Insurance T_{4i}	Λ	$\vee$	$/$	Λ	$\backslash$	$\vee$	Λ	$\backslash$	$\vee$	$/$	3	.5868
Real Estate T_{5i}	$\vee$	$/$	Λ	$\vee$	Λ	$\backslash$	$\backslash$	$\backslash$	$\vee$	Λ	3	.5868
Service T_{6i}	$\vee$	Λ	$\vee$	$/$	$/$	$/$	Λ	$\backslash$	$\vee$	Λ	3	.5868

Table 8. The sequences of qualitative local trends of the establishment ratios on the Chuo Railway Line Area.

Establishments	1	2	3	4	5	6	7	8	9	Number or Runs	Peak Concentration γ	Equation (44)
Construction T'_{1i}	$\bar{\Lambda}$	$\bar{\Lambda}$	$\bar{\Lambda}$	$\bar{\Lambda}$	Λ	Λ	Λ			2	1.0000	.0571
Manufacturing T'_{2i}	$\bar{\Lambda}$	$\bar{\Lambda}$	$\bar{\Lambda}$	$\bar{\Lambda}$	Λ	$\bar{\Lambda}$	$\bar{\Lambda}$	$\bar{\Lambda}$	Λ	4	.6000	.5833
Wholesale & Retail T'_{3i}	Λ	Λ	$\bar{\Lambda}$	Λ	$\bar{\Lambda}$	Λ				5	.5714	1.0000
Finance & Insurance T'_{4i}	Λ	$\bar{\Lambda}$	Λ	$\bar{\Lambda}$	Λ	$\bar{\Lambda}$	$\bar{\Lambda}$			6	.2000	.9715
Real Estate T'_{5i}	$\bar{\Lambda}$	$\bar{\Lambda}$	Λ	Λ	$\bar{\Lambda}$	$\bar{\Lambda}$	$\bar{\Lambda}$	Λ		4	.8571	.4285
Service T'_{6i}	$\bar{\Lambda}$	Λ	$\bar{\Lambda}$	$\bar{\Lambda}$	$\bar{\Lambda}$	Λ	$\bar{\Lambda}$	Λ		6	.2857	.9285

Table 9. The modified sequences of the qualitative local trends of Table 8.

R_{ij} r_{ij}	Construction	Manufacturing	Wholesale & Retail	Finance & Insurance	Real Estate	Service
Construction	1.0	.6715 .4	-.1463 -.1	-.4062 .2	-.0049 .0	.0303 .0
Manufacturing		1.0	-.2808 -.3	-.1357 .0	.2008 -.2	-.0912 .2
Wholesale & Retail			1.0	.0087 -.3	-.0093 -.3	.1118 -.7
Finance & Insurance				1.0	.3003 .0	-.3583 .0
Real Estate					1.0	.3335 .2
Service						1.0

Table 10. The correlation coefficients R_{ij} (upper figures) and the qualitative correlation coefficients r_{ij} (lower figures) between two trend curves.

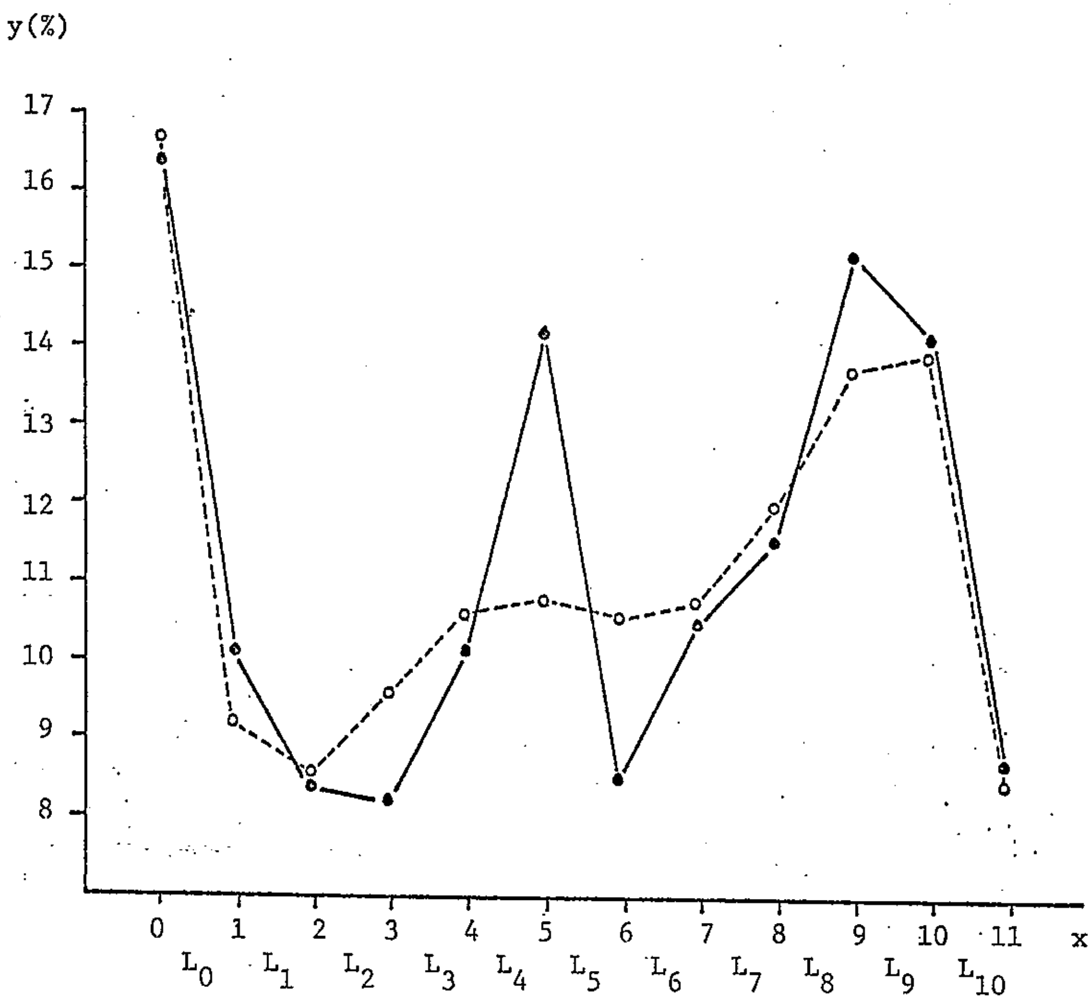


Figure 1 A ratio of the number of establishments belonging to the secondary sector at distance x from Shinjuku on the Chuo Railway Line Area, 1970

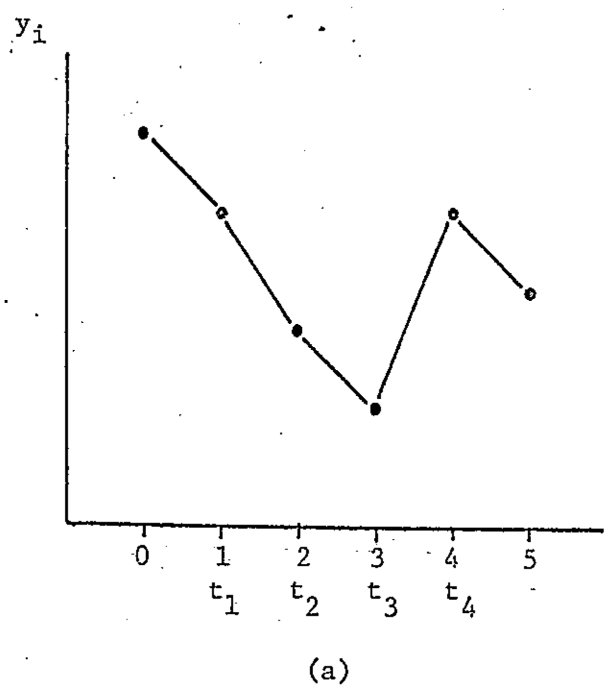


Figure 2 An illustrative example of dependence between local trends t_2 and t_3 provided that t_1= and t_4=

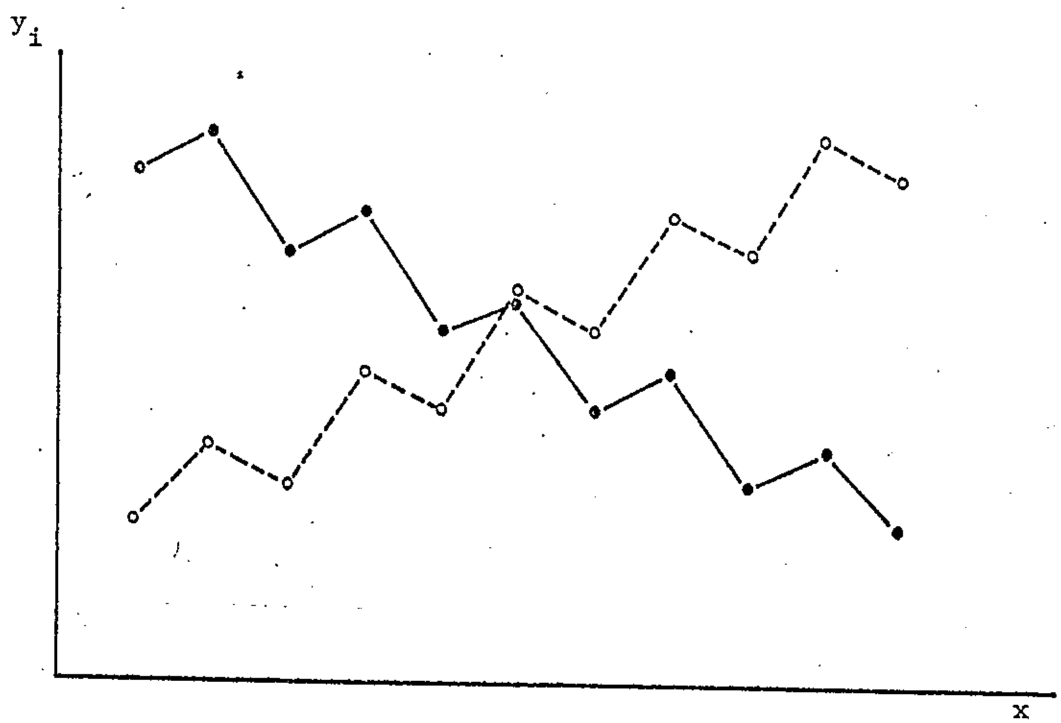


Figure 3 An example of two trend curves whose correlation is globally positive but locally negative

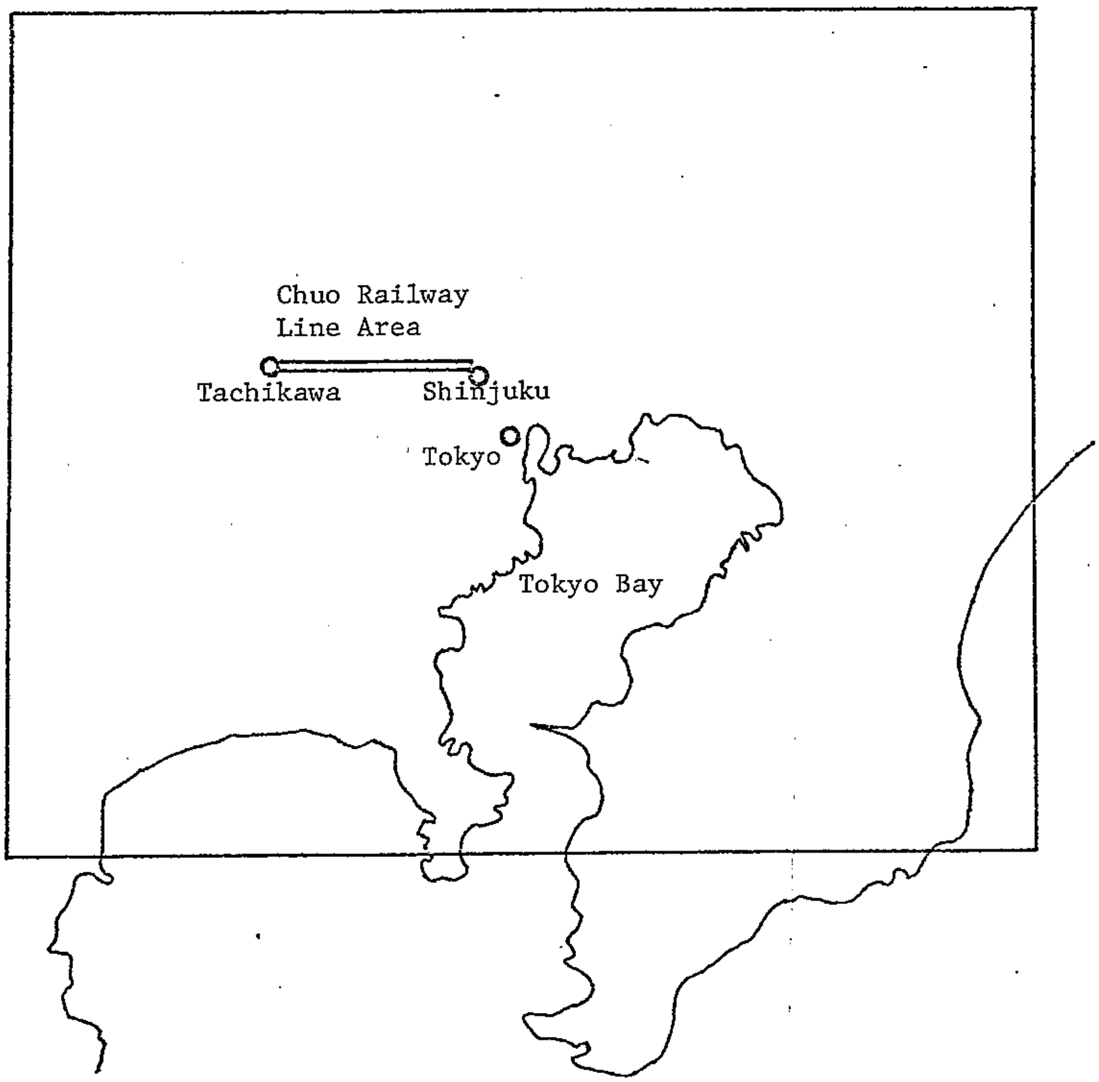
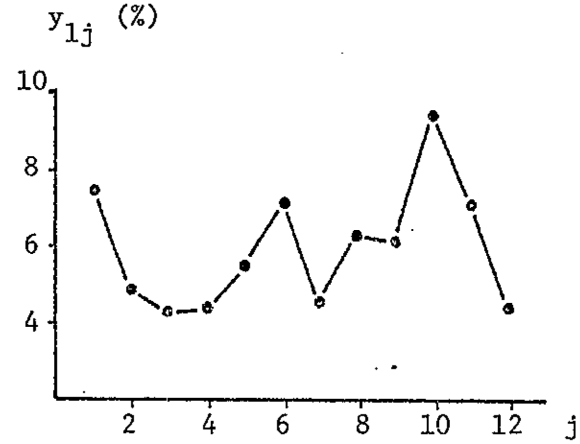
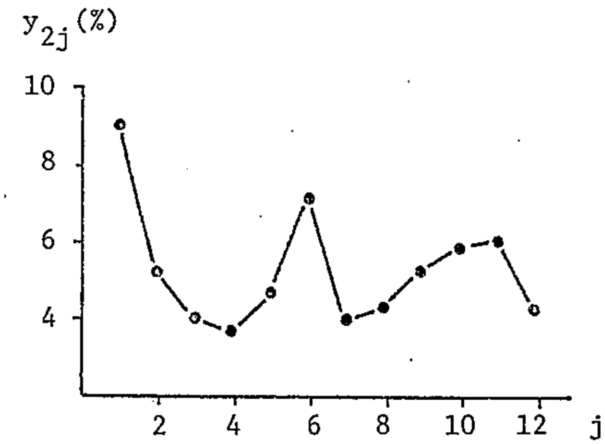


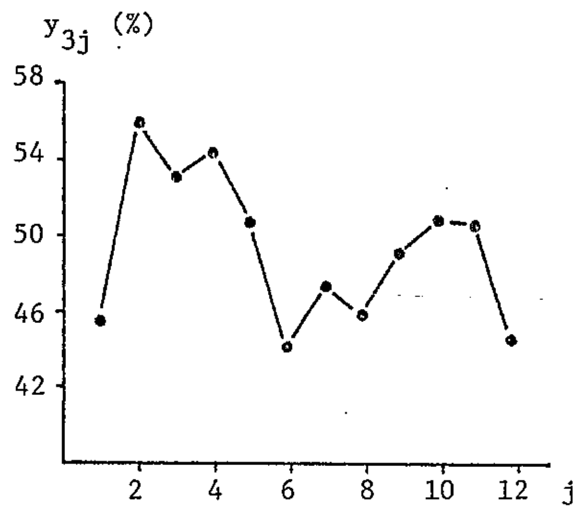
Figure 4 The Chuo Railway Line Area and the 60km x 60km square Region centered at the Tokyo Station



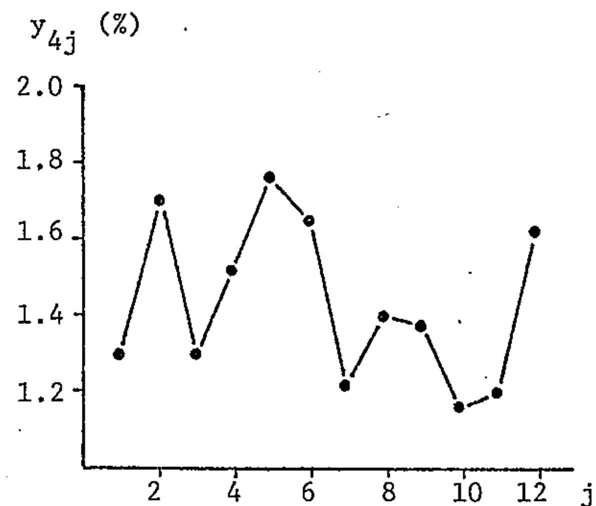
(a) The constraction establishment ratio



(b) The manufacturing establishment ratio

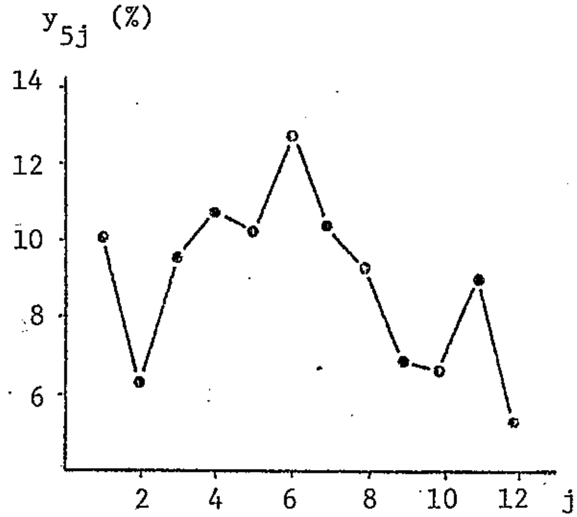


(c) The wholesale and retail establishment ratio

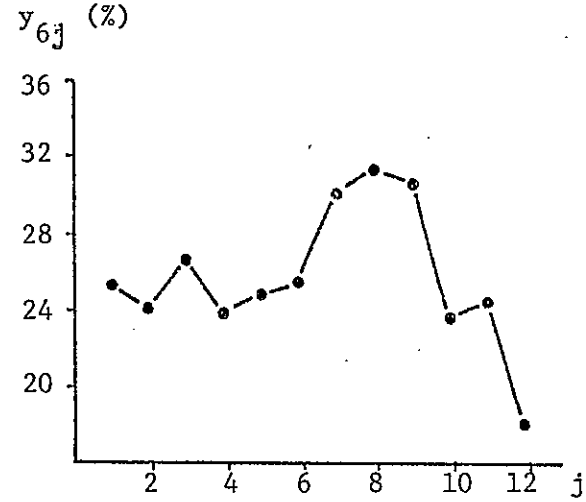


(d) The finance and insurance establishment ratio

(Figure 5)



(e) The real estate establishment ratio



(f) The service establishment ratio

Figure 5 The establishment ratio on the Chuo Railway Line Area, 1970

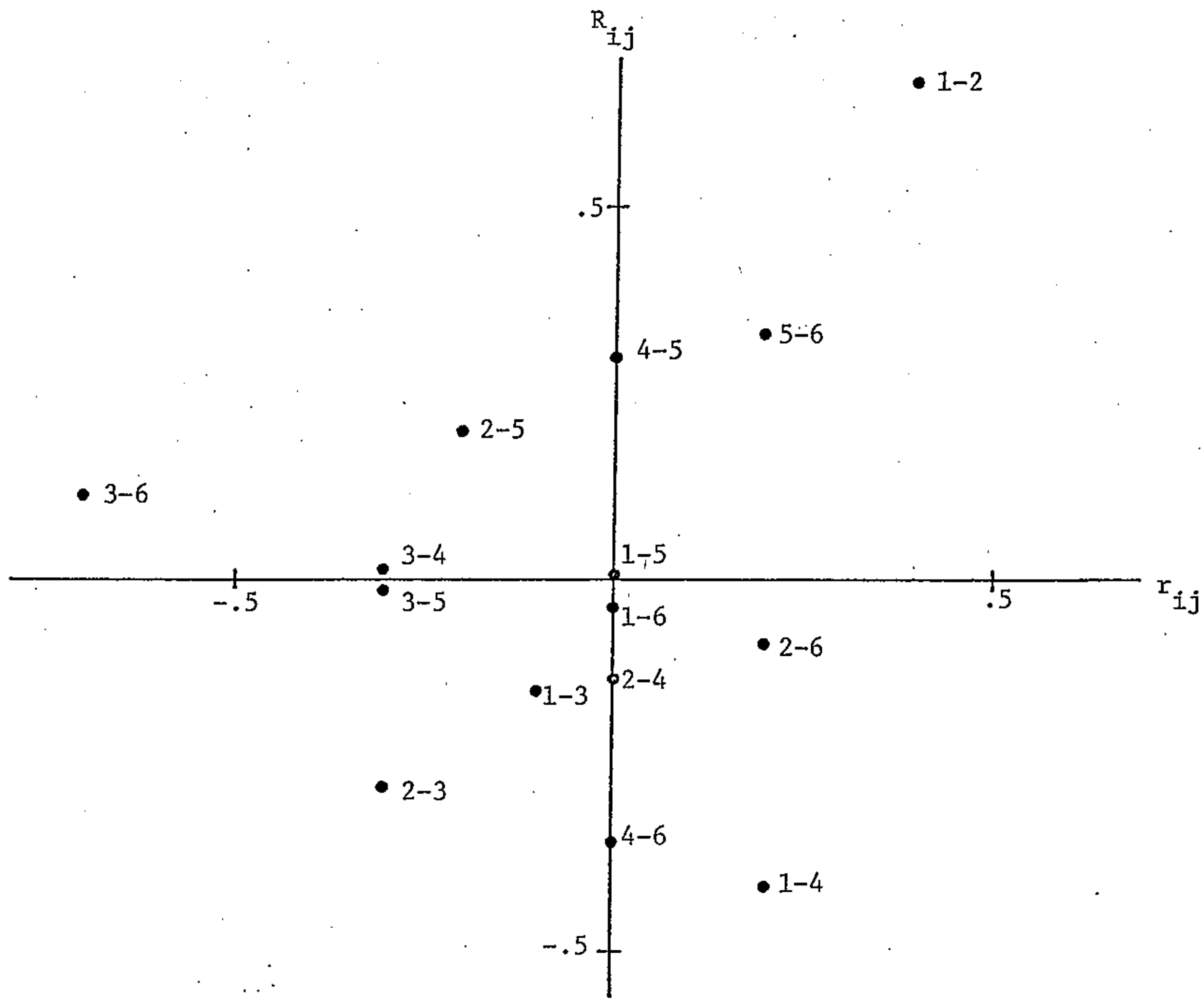


Figure 6 Correlation coefficients R_{ij} and r_{ij} between the i th establishment ratio and the j th establishment ratio

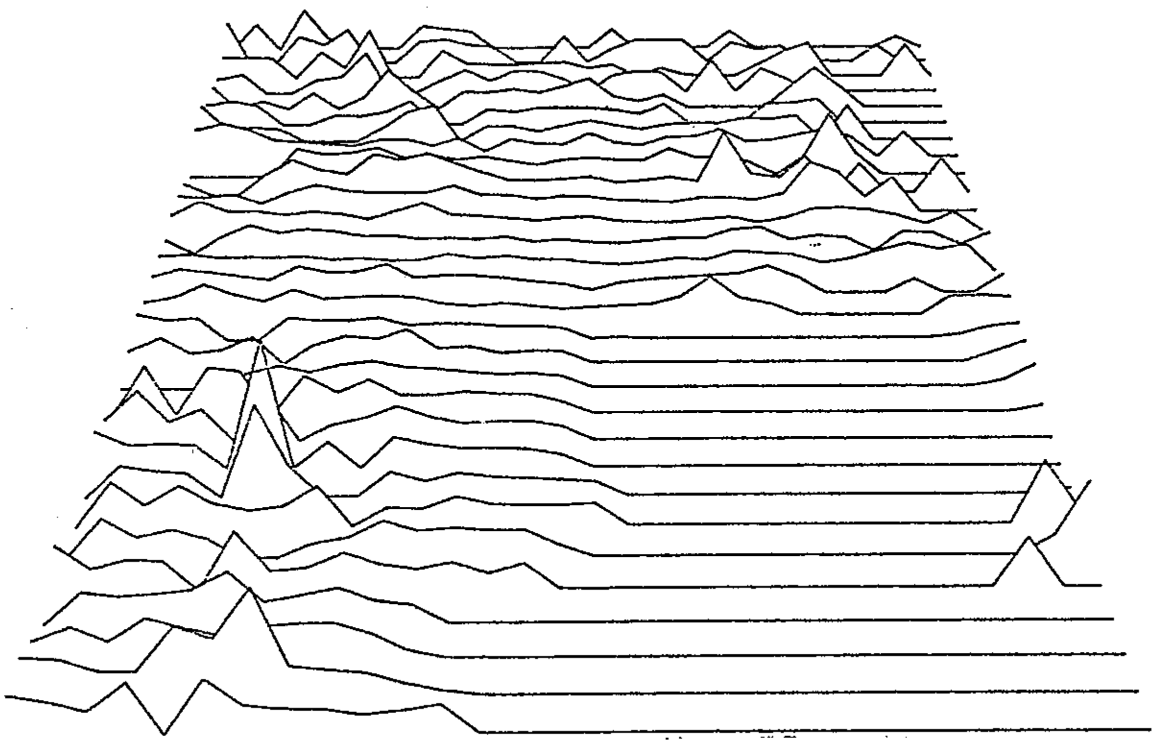


Figure 7 The Constraction Establishment Ratio in the 60km x 60km Square Region centered at the Tokyo Station, 1970