Profile Analysis of 24-Hours Measurements of Blood Pressure

Chihiro Hirotsu,[1,*] Eri Ohta,[1] Nobuyoshi Hirose,[2] and Kenichiro Shimizu[2]

[1]Faculty of Science and Technology, Meisei University, Tokyo
[2]Department of Geriatrics, Keio University School of Medicine, Tokyo
*email: hirotsu@ge.meisei-u.ac.jp

SUMMARY. A method is proposed for classifying subjects according to their convex, flat, or concave change patterns of 24-hours blood pressure measurements. To obtain such a classification is useful for detecting subjects who show abnormal change patterns and giving them appropriate medical treatments. Therefore, an appropriate statistic is proposed for detecting a systematic change along the time axis, as well as a statistic with its inverse characteristic appropriate for evaluating the noise variation. The method is based on the ratio of those two types of statistics; it is verified to work well on real data, giving a classification of subjects into four types of subgroups: extreme dipper, dipper, nondipper, and inverted dipper. It also suggests that there might be an ultra-extreme dipper subgroup.

KEY WORDS: Analysis of interaction; Classification; Convex and concave patterns; Multiple comparisons; Order-restricted inference.

1. Introduction

There are 48 observations of the systolic blood pressure measured every 30 minutes for 24 hours. Those measurements are known normally to go down slightly in the night; a large depression is abnormal, as is a flat or a tendency towards elevation. In the following, we call these down- and upward tendencies in the night "convex and concave patterns," respectively, since we take the starting point at 3:00 p.m. We are, therefore, interested in classifying the 24-hours profiles of blood pressure based on those convex and concave features. To obtain such a classification is useful for detecting those subjects suffering from abnormal profile patterns and giving them appropriate medical treatments. The classification is expected further to give important information by a covariate analysis.

As background, it was first pointed out by Millar-Craig, Bishop, and Raftery (1978) that the systolic and diastolic blood pressures follow a circadian pattern in both normotensive and hypertensive individuals, with values being lower at night than during the day. Then, an inverse pattern has been reported in subjects with autonomic failure by Mann et al. (1983), and Kobrin et al. (1984) reported the elderly, who have a higher incidence of cardiovascular disease, may lose this diurnal variation. Following these reports, O'Brien, Sheridan, and O'Malley (1988) analyzed 24-hours ambulatory blood pressure measurements from 123 hypertensive consecutive patients; they classified those patients into dippers (102 patients, 82.9%) or nondippers (21,17.1%), according to whether the difference between mean daytime blood pressure and mean nighttime blood pressure showed the normal circadian variation (dipper) or not (nondipper). Thus, they first concluded that there is a group of patients whose 24-hours blood pressure does not follow the normal circadian pattern and who thus may be at higher risk of cerebrovascular complications. Yamamoto et al. (1995) supported this idea by showing that the diminished nocturnal blood pressure decline is associated with the specific location of cerebrovascular lesion site, as well as with the extent of the vascular disease. Ohkubo et al. (1997) further defined the four classes of the extreme dipper (ED), dipper (D), nondipper (ND), and the inverted dipper (ID), finding that the mortality rate is highest in the inverted dipper subgroup. Another example of analysis of 24-hours blood pressure measurements is given by Degaute et al. (1994), who found, by a twin study, the simultaneous genetic effects on the circadian blood pressure and the heart rate rhythms. In these works, however, the circadian rhythm has been characterized by several parameters, such as the 24-hours mean of measurements, daytime mean, sleep mean, evening acrophase, nocturnal nadir, and the amplitude. This last thing is defined as 50% of the difference between the maximum and minimum of the fitted curve. Separate analyses are performed for those parameters. Although each analysis is useful and informative, the simultaneous inferences suffer from the well-known multiplicity problem. Another process is required to combine those separate analyses. Also, some parameters are known to be sensitive to the outliers. This, together with the correlations among parameters, makes the simultaneous inferences very difficult. Generally speaking, the clinical data are essentially multivariate, and to have a good summary measure is rather a large problem.

In the present article, we therefore intend to characterize the 24-hour profile of blood pressure measurements on the whole, with particular interest in their convex and concave patterns. We succeeded with real data to obtain four classes that are very interestingly interpreted as ED, D, ND, and ID. Our analysis also suggests that the ED group might be

divided up more minutely into two classes, isolating the ultra extreme dipper subgroup.

There have been several attempts at classifying subjects based on their time series profiles, such as blood pressure or total cholesterol amount. Among them, a naive method used by clinical doctors is to classify subjects according to several prototype patterns. It has, however, been pointed out that there are various difficulties in classifying a 100 subjects by a sight. Some statistical approaches have also been proposed. Hirotsu (1991), among others, proposed a nonparametric approach, assuming only a monotone change of the responses along the time axis. The method was successfully applied to the cholesterol data to obtain improved, deteriorated, and unchanged subgroups after a medical treatment. The 24-hours profile of the blood pressure measurements is, however, characterized by coming back to approximately the initial value after 24 hours, whereas the previous approach assumes the monotone tendency and should not be successful. Also, in this case, we cannot assume the independence of measurements, since they are taken every 30 minutes, whereas the previous example of cholesterol measurements are taken every 6 months. This allows the independence assumption between successive measurements. In Sections 3 to 7 of the present article, we therefore propose a modification of the previous method, accommodating to the situation here. The proposed method is based on the squared distance between two profiles, defined so as to be sensitive to the difference of convex and concave patterns. As shown in (A.1), the squared distance is a weighted sum of squares representing mainly quadratic and cubic variations in time. The evaluation of the method is given in Section 8 and an application to real data is given in Section 9.

2. The Data Set and the Classical Approach

We first give the summary data of 24-hours blood pressure measurements of 203 subjects in Table 1, which will be analyzed later in Section 9. In the table, we've included the average of change rate of blood pressure defined by

Δ = (Average in the daytime − Average in the nighttime)

$\div$ (Overall average),

where the day- and the nighttime are tentatively defined as 10 a.m.~8 p.m. and 0 a.m.~6 a.m., respectively. A subject is classified as depression, flat, or elevation type according to $\Delta \geq 0.1$, $-0.1 < \Delta < 0.1$ or $\Delta \leq -0.1$, respectively. The flat or elevation type should be treated (see Shimizu, 1994). The classification was, however, rather arbitrary. As stated in the

Table 1
Summary of blood pressure data (n = 203)

Data profile	Mean (mmHg)	Standard deviation (mmHg)
Average of 24-hours measurements	137.44	17.47
Minimum of 24-hours measurements	104.07	15.47
Maximum of 24-hours measurements	173.28	22.85
Average of change rate of blood pressure	0.074	0.098

introduction, it has only recently been proposed to define the four classes ED, D, ND, and ID. We therefore propose a more scientific approach to obtaining subgroups, significant both statistically and clinically, without prespecifying the number of subgroups. It should be noted that our approach is to compare the shape of response curves on the whole, rather than based on somewhat arbitrary amount Δ, which should depend on the definition of the day- and nighttime.

3. Statistical Model

We denote by y_{ij} the jth observation of the ith subject and assume a model

$$y_{ij} = \mu_{ij} + \varepsilon_{ij}, \quad i = 1, \ldots, n; \ j = 1, \ldots, t. \qquad (1)$$

We do not employ any particular parametric model, but assume some smooth and systematic changes in μ_{ij} with respect to the passage of time. We tentatively assume the observational errors ε_{ij} are independently distributed as $N(0, \sigma^2)$. (We discuss later the procedure for correlated errors.) This means that the high correlation between the two close measurements is modeled by a smooth systematic change in the mean level and that the casual measurement errors are assumed to be mutually independent. It should be noted that we now have two components, μ_{ij} and ε_{ij}, the former changing slowly and the latter moving rather quickly and independently. Along this line, Hirotsu (1991) proposed a method for comparing profiles, which tries to detect a systematic departure from the null model

$$H_0 : \mu_{ij} - \mu_{i'j} = \mu_{ij+1} - \mu_{i'j+1},$$
$$i, i' = 1, \ldots, n(i \neq i'); \ j = 1, \ldots, t - 1$$

in the direction $H_1 : \mu_{i1} - \mu_{i'1} \leq \cdots \leq \mu_{it} - \mu_{i't}$. The hypothesis H_1 asserts that the difference of two profiles of the subjects i and i' is monotone increasing in time, while the difference is a constant in H_0. The method is, therefore, useful for detecting an up- or downward tendency of measurements after a treatment, whereas the data here are characterized by returning to approximately the same level after 24 hours. We also want a procedure for distinguishing the convex, flat or concave subgroup. For accommodating to the pattern of interest here, we assume a systematic departure from a null model

$$H'_0 : \mu_{ij} - \mu_{ij+1} - (\mu_{i'j} - \mu_{i'j+1})$$
$$= \mu_{ij+1} - \mu_{ij+2} - (\mu_{i'j+1} - \mu_{i'j+2}),$$

in the direction

$$H_2 : \mu_{ij} - \mu_{ij+1} - (\mu_{i'j} - \mu_{i'j+1})$$
$$\leq \mu_{ij+1} - \mu_{ij+2} - (\mu_{i'j+1} - \mu_{i'j+2}),$$
$$i, i' = 1, \ldots, n(i \neq i'); j = 1, \ldots, t - 2.$$

The null model H'_0 asserts the parallelism of slope changes between the profiles of subjects i and i'. Rewriting H_2 as $H_2 : \mu_{ij} - 2\mu_{ij+1} + \mu_{ij+2} \leq \mu_{i'j} - 2\mu_{i'j+1} + \mu_{i'j+2}$, the resulting test is expected to be useful for detecting a difference among subjects in the slope change patterns, namely, in the convex, flat, or concave tendency. As the result, the present method is useful for obtaining homogeneous subgroups, in the sense of a convex or concave pattern.

In the next section, we derive a systematic statistic for the specified alternative H_2. In Section 5, we introduce a statistic for evaluating the error variance σ^2 which has an inverse characteristic to the systematic statistic. Then, the distribution of the ratio of those two types of statistics is derived in Section 6. Finally, in Section 7, a classification procedure is proposed based on the distribution theory.

4. Test Statistics and the Distribution Theory

4.1 *Squared Distance between the Profiles of Subjects i and i'*

The alternative hypothesis H_2 can be rewritten in matrix notation as

$$H_2 : \boldsymbol{L}_t'(\boldsymbol{\mu}_{i'} - \boldsymbol{\mu}_i) \geq \boldsymbol{0}, \tag{2}$$

where $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{it})'$, $\boldsymbol{L}_t'$ a second order difference matrix defined by

$$\boldsymbol{L}_t' = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \multicolumn{8}{c}{\dotfill} \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{pmatrix}_{t-2 \times t} .$$

and the inequality is elementwise, with at least one inequality strong. It should be noted that H_2 is a two-sided alternative with respect to i and i', since we assume no natural ordering in subjects; so i and i' are permutable. By virtue of a complete class lemma given in Hirotsu (1982), a quadratic statistic appropriate for detecting a difference of profiles between the subjects i and i', with particular interest in the systematic change denoted by equation (2), is formed by

$$\chi^{\dagger 2}(i; i') = 2^{-1} \times (\boldsymbol{y}_i - \boldsymbol{y}_{i'})'\boldsymbol{Q}_t\boldsymbol{Q}_t'(\boldsymbol{y}_i - \boldsymbol{y}_{i'}), \tag{3}$$

where $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{it})'$ is the observation vector for the ith subject. The explicit form of the key matrix $\boldsymbol{Q}_t$, formed from $\boldsymbol{L}_t$, as well as its motivation, are given in Appendix A. In short, the τth component of $\boldsymbol{Q}_t'(\boldsymbol{y}_i - \boldsymbol{y}_{i'})$ denotes a difference in the amount of slope changes between the two profiles of i and i' in the $(\tau + 1)$th period, $\tau = 1, \ldots, t - 2$, see also Hirotsu and Marumo (2002). The statistic $\chi^{\dagger 2}(i; i')$ is well characterized by the expansion given in (A.1), and may be called a squared distance between subjects i and i'. Next, we proceed to defining a squared distance between two subgroups of subjects.

4.2 *Squared Distance between Two Subgroups of Subjects G_1 and G_2*

We need to define the distance between two subgroups of subjects consistently with the definition for two subjects. Without any loss of generality, let G_1 and G_2 be composed of subjects $(1, \ldots, q_1)$ and $(q_1 + 1, \ldots, q_1 + q_2)$, respectively. Then, a squared distance between G_1 and G_2 is defined by

$$\chi^{\dagger 2}(G_1; G_2) = \left(q_1^{-1} + q_2^{-1}\right)^{-1}\left(\bar{\boldsymbol{y}}_{G_1} - \bar{\boldsymbol{y}}_{G_2}\right)'\boldsymbol{Q}_t\boldsymbol{Q}_t'\left(\bar{\boldsymbol{y}}_{G_1} - \bar{\boldsymbol{y}}_{G_2}\right),$$

where $\bar{\boldsymbol{y}}_{G_l}$ is the average of $\boldsymbol{y}_i$ over the subjects belonging to the subgroup G_l. If G_1 and G_2 are composed of only one element i and i', say, then $\chi^{\dagger 2}(G_1; G_2)$ reduces to $\chi^{\dagger 2}(i; i')$ of (3). The statistic $\chi^{\dagger 2}(G_1; G_2)$ can be expanded in exactly the same way as in (A.1).

4.3 *Generalized Squared Distance among Subgroups $G_1, \ldots, G_m$*

Finally, we need to define the generalized distance among subgroups $G_1, \ldots, G_m$ consistent with the definition for $m = 2$. In particular, let G_l, $l = 1, \ldots, m$, be a partition of n subjects with size q_l, $q_1 + \cdots + q_m = n$. Then, a generalized squared distance $\chi^{\dagger 2}(G_1; \cdots; G_m)$ among those subgroups is defined by

$$\max_{\lambda_1, \ldots, \lambda_m} \left\{\boldsymbol{y}'(G_1; G_2; \cdots; G_m)\boldsymbol{Q}_t\boldsymbol{Q}_t'\boldsymbol{y}(G_1; G_2; \cdots; G_m)\right\},$$

where the generalized difference $\boldsymbol{y}(G_1; G_2; \cdots; G_m)$ among subgroups $G_1, \ldots, G_m$ is defined as the orthonormalized contrast of $\boldsymbol{y}_i'$s, $\gamma_1\boldsymbol{y}_1 + \cdots + \gamma_n\boldsymbol{y}_n$, with the coefficient γ_i equal to λ_l if the ith subject belongs to the lth subgroup G_l, so that $\sum q_l\lambda_l = 0, \sum q_l\lambda_l^2 = 1$. The value of the statistic $\chi^{\dagger 2}(G_1; \cdots; G_m)$ can be obtained as the maximal latent root of the $(t - 2) \times (t - 2)$ matrix $\sum_l^m q_l\boldsymbol{Q}_t'(\bar{\boldsymbol{y}}_{G_l} - \bar{\boldsymbol{y}})(\bar{\boldsymbol{y}}_{G_l} - \bar{\boldsymbol{y}})'\boldsymbol{Q}_t$, where $\bar{\boldsymbol{y}} = (\bar{y}_{.1}, \ldots, \bar{y}_{.t})$ is the average of $\boldsymbol{y}_i$'s. The proof of this statement is similar to the proof of Lemma 2.1 of Hirotsu (1991) and is omitted here. If $m = 2$, then

$$\boldsymbol{y}(G_1; G_2) = \left(q_1^{-1} + q_2^{-1}\right)^{-1/2}\left(\bar{\boldsymbol{y}}_{G_1} - \bar{\boldsymbol{y}}_{G_2}\right)$$

and the generalized squared distance is consistent with the squared distance between two subgroups as defined in Section 4.2.

All the squared distances introduced in Sections 4.1 to 4.3 are obviously bounded above by a statistic

$$\chi_{\max}^{\dagger 2} = \max\left\{(\gamma_1\boldsymbol{y}_1 + \cdots \gamma_n\boldsymbol{y}_n)'\boldsymbol{Q}_t\boldsymbol{Q}_t'(\gamma_1\boldsymbol{y}_1 + \cdots \gamma_n\boldsymbol{y}_n)\right\},$$

where the maximum is taken over all $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)'$, subject to $\sum \gamma_i = 0, \sum \gamma_i^2 = 1$. It is easy to verify that all the squared distances are obtained by specifying $\boldsymbol{\gamma}$ appropriately. We can therefore apply the null distribution of $\chi_{\max}^{\dagger 2}$ to evaluating those squared distances conservatively. When $n \geq t$, and under the null hypothesis H_0', the statistic $\chi_{\max}^{\dagger 2}$ is shown to be distributed as the largest root of a Wishart matrix $\sigma^2 W(\boldsymbol{Q}_t'\boldsymbol{Q}_t, n - 1)$. The proof of this statement follows exactly that of Lemma 2.1 in Hirotsu (1991) and so is omitted here. Then, by a similar argument in Hirotsu (1991), the asymptotic null distribution is very well approximated by $\sigma^2 \gamma_{(2)}\chi_{(2)}^2 (n - 1)$, with $\gamma_{(2)}$, the largest root of $\boldsymbol{Q}_t'\boldsymbol{Q}_t$, given by (A.2), and $\chi_{(2)}^2 (n - 1)$ a chi-squared variable with $n - 1$ df, detecting a departure in the direction of the second-order polynomial in time. It is well known that the approximation is good if the largest root of $\boldsymbol{Q}_t'\boldsymbol{Q}_t$ is large compared to the remaining roots. In this case it is, indeed, since the ratio of the largest root $\gamma_{(2)}$ to the total of all the roots of $\boldsymbol{Q}_t'\boldsymbol{Q}_t$ is

$$\gamma_{(2)}\left/\sum_{k=2}^{t-1} \gamma_{(k)}\right. = \frac{3}{4}\frac{t(t^2 - 1)}{(t - 2)(t^2 + 2t + 3)} > \frac{3}{4}.$$

Thus, $\gamma_{(2)}$ dominates other roots $\gamma_{(3)}, \ldots, \gamma_{(t-1)}$. See Section 8 for the accuracy of the chi-squared approximation.

5. Evaluating the Error Variance σ^2

The null distribution of $\chi_{\max}^{\dagger 2}$ involves a nuisance parameter σ^2, so we need to evaluate it. There is, however, no degree of freedom left in model (1) for estimating σ^2, since the two-way data are without replication, and we are not assuming

any parametric model for μ_{ij}. A possible approach would be to cancel σ^2 by dividing $\chi^{\dagger 2}_{\max}$ by an appropriate quadratic form less affected by the systematic change in μ_{ij}. What we are doing here is detecting dominant interaction effects between subject and time in two-way data without replication. So, the approach taken here is regarded as an extension of Johnson and Graybill (1972) and Hirotsu (1983), taking into consideration the natural ordering in time. The usual sum of squares for interaction used there is, however, inappropriate for our denominator statistic, as it is easily affected by a systematic component. We therefore introduce $S^- = (S^-_1 + \cdots + S^-_n)$ by inverting a quadratic form for detecting a systematic departure from the previous section, where the explicit form of S^-_i and its distributional characteristic are given in Appendix B. Similar considerations are taken by Sen and Srivastava (1975), who introduced a sum of squares for evaluating σ^2 less affected by a change in the systematic component in the one-way change-point analysis. As stated in Appendix B, our S^- is based on the weighted sum of squares of the second order differences in the subsequent measurements, whereas their statistic is the sum of squares of the subsequent differences.

6. Distribution of the Ratio of $\chi^{\dagger 2}_{\max}$ to S^-

By virtue of the expansion of S^- given in Appendix B, the null distribution of $\chi^{\dagger 2}_{\max}/S^-$ is the distribution of the largest root of $W(\boldsymbol{Q}'_t\boldsymbol{Q}_t, n-1)$ divided by $\{\gamma_{(2)}^{-1}\chi^2_{(2)}(n-1) + \cdots + \gamma_{(t-1)}^{-1}\chi^2_{(t-1)}(n-1)\}$, where the $\chi^2_{(k)}(n-1)$ are the independent chi-squared variables with df $n-1$, for detecting departure in the direction of the kth-order polynomial in time. Since the number n of subjects is sufficiently large in our application, we have a very nice asymptotic formula for evaluating the p-value of the observed maximum s_0 of $\chi^{\dagger 2}_{\max}/S^-$. First, by virtue of the chi-squared approximation for the largest root, we have an approximation,

$$p = \Pr\left\{\chi^{\dagger 2}_{\max}/S^- \geq s_0\right\}$$

$$= \Pr\left\{\frac{\gamma_{(2)}\chi^2_{(2)}(n-1)}{\gamma_{(2)}^{-1}\chi^2_{(2)}(n-1) + \cdots + \gamma_{(t-1)}^{-1}\chi^2_{(t-1)}(n-1)} \geq s_0\right\}$$

$$= \Pr\left\{\frac{\left(\gamma_{(2)} - s_0\gamma_{(2)}^{-1}\right)\chi^2_{(2)}(n-1)}{\gamma_{(3)}^{-1}\chi^2_{(3)}(n-1) + \cdots + \gamma_{(t-1)}^{-1}\chi^2_{(t-1)}(n-1)} \geq s_0\right\}.$$

$$\tag{4}$$

Now, the denominator in (4) is a positive linear combination of the independent chi-squared variables, and is well approximated by a constant times chi-squared variable $d\chi^2_f$, with constants obtained by equating the first two moments,

$$df = (n-1)\left(\gamma_{(3)}^{-1} + \cdots + \gamma_{(t-1)}^{-1}\right),$$

$$d^2 f = (n-1)\left(\gamma_{(3)}^{-2} + \cdots + \gamma_{(t-1)}^{-2}\right).$$

Since the numerator and the denominator are mutually independent, we finally obtain an asymptotic formula

$$p = \Pr\left\{F_{n-1,f} \geq s_0\left(\gamma_{(3)}^{-1} + \cdots + \gamma_{(t-1)}^{-1}\right)\Big/\left(\gamma_{(2)} - s_0\gamma_{(2)}^{-1}\right)\right\}, \quad (5)$$

where $F_{n-1,f}$ is an F variable with df $(n-1, f)$.

For comparisons with S^-, we introduce the usual unweighted sum of squares for interaction $S_{R\times C} =$ $\sum_i^n \sum_j^t (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{\bar{y}}_{..})^2$ which has been used by previous authors. Also, we introduce $S^{-l}_{R\times C} = S_{R\times C} - \chi^2_l$, subtracting a linear component $\chi^2_l = \sum_i^n \left[\sum_j^t \{l_j (y_{ij} - \bar{y}_{.j})\}\right]^2$, where the coefficients l_j are obtained by standardizing a vector $(1, 2, \ldots, t)$ so they have the summation 0 and the sum of squares unity. χ^2_l is distributed as $\sigma^2\chi^2_{(1)}(n-1)$, where $\chi^2_{(1)}$ is a chi-squared variable with df $n-1$ and independent of the other component chi-squared of S^-. To the contrary, we may add the linear component to S^- using the same weight as $\chi^2_{(t-1)}(n-1)$ in expansion (B.1), say, since in our case, we hardly expect a linear component to exist as a systematic component. We denote the statistic by S^{+l}. The modifications of the formula (5) for evaluating the p-values for newly introduced error sum of squares are straightforward. Comparisons of these four methods are given in Section 8.

7. Method for Classifying Subjects Based on the Distribution Theory of Section 6

We first apply the following procedures to obtain a classification of subjects into a given number of subgroups so that the squared distances between subjects are small if they belong to the same group, and large otherwise, and then search for the appropriate number of subgroups.

1. Specify K the number of subgroups.
2. Let $G_1, \ldots, G_{n-k+1}$ be the partition of subjects at the kth stage. Find two subgroups G_i and $G_{i'}$ that give the smallest squared distance $\chi^{\dagger 2}(G_i; G_{i'})$ from among all the possible combinations of $G_1, \ldots, G_{n-k+1}$. Then, combine those two subgroups to form $(n-k)$ subgroups at the next $(k+1)$th stage. Note that, at the first stage, each of $G_1, \ldots, G_n$ is composed of only one subject.
3. Continue the procedure 2 until the number of subgroups becomes the prespecified number K. The resulting partition is denoted by $G_1, \ldots, G_K$.
4. Adjust the classification by the following algorithms. First, for $i = 1$, calculate the squared distance between the subject 1 and subgroups $G_1(1), \ldots, G_K(1)$, where $G_k(1)$ denotes that the subject 1 is eliminated from G_k. Then, classify the subject 1 into the subgroup that gives the smallest squared distance $\chi^{\dagger 2}(1; G_k(1))$ among $k = 1, \ldots, K$. Do the same thing between the subject 2 and the renewed subgroups, with subject 2 eliminated. Continue the process repeatedly until no reduction in the generalized squared distance $\chi^{\dagger 2}(G_1; \cdots; G_K)$ is obtained. We may instead base our stopping rule on the reduction of the total within the sum of squares

$$\sum_{k=1}^{K}\sum_{i \in G_k} \left(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{G_k}\right)' \boldsymbol{Q}_t\boldsymbol{Q}'_t\left(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{G_k}\right)$$

$$= \sum_{i=1}^{n} \boldsymbol{y}'_i\boldsymbol{Q}_t\boldsymbol{Q}'_t\boldsymbol{y}_i - \sum_{k=1}^{K} q_k\bar{\boldsymbol{y}}'_{G_k}\boldsymbol{Q}_t\boldsymbol{Q}'_t\bar{\boldsymbol{y}}_{G_k}.$$

5. Evaluate the p-value of $\chi^{\dagger 2}(G_1; \cdots; G_K)/S$ using the method of Section 6, where S stands for S^-, S^{+l}, $S_{R\times C}$, or $S^{-l}_{R\times C}$.

It should be noted that $\chi^{\dagger 2}$ is monotone increasing in the number K of subgroups, so that if a significant result is obtained at $K = k$, then we have also significant result for every K larger than k. We therefore additionally recommend, in step 5, calculating the percent contribution of the resulting classification to the total sum of squares

$$\frac{\sum_{k=1}^{K} q_k \left(\bar{\boldsymbol{y}}_{G_k} - \bar{\boldsymbol{y}}\right)' \boldsymbol{Q}_t \boldsymbol{Q}'_t \left(\bar{\boldsymbol{y}}_{G_k} - \bar{\boldsymbol{y}}\right)}{\sum_{i=1}^{n} (\boldsymbol{y}_i - \bar{\boldsymbol{y}})' \boldsymbol{Q}_t \boldsymbol{Q}'_t (\boldsymbol{y}_i - \bar{\boldsymbol{y}})}$$

as useful information for a stopping rule.

8. Evaluation of the Procedure

8.1 *Accuracy of the Procedure Based on the Approximate Distribution*

The accuracy of the procedure is evaluated under the null model using simulation. According to the data set of Section 9, we set $n = 200$ and $t = 48$, and create 200 $\times$ 48 pseudo-independent normal variables repeatedly 1000 times. The proportion of the number of cases where the significant classification at significance level 0.05 was obtained at any number K of classes by the procedure explained in Section 7 is given in Table 2 ($\rho = 0$); it gives the type I error rate of the procedure. The meaning of each data set in the table is explained in the next section, where $t_{1(0)}$ is the original data set and others are created from it. Since the standard deviation of the estimated rate is approximately 0.007, we see that the approximate distribution theory is satisfactory for each method of evaluating errors.

8.2 *Effects of Correlation among Errors*

Actually, however, the serial correlation coefficient ρ as high as 0.45 is observed in the data set of Section 9. We therefore created correlated data with that amount of correlation coefficient between the two subsequent measurements; we give the simulated type I error rate in the first row of Table 2 ($\rho = 0.45$). Then, we notice that the effects of correlation is unexpectedly large and some protection against this is needed. This is because our numerator sum of squares is formed to also be sensitive to the systematic change in errors caused by high correlations among neighboring data. We therefore compare

the following six procedures to the naive procedure based on $t_{1(0)}$.

1. Employ 1 measurement at every 3-point interval to create the 200 $\times$ 12 data set $t_{1(3)}$.
2. Average 4 subsequent points to create the 200 $\times$ 12 data set $\bar{t}_{4(0)}$.
3. Average 2 subsequent points at every 2-point interval to create the 200 $\times$ 12 data set, $\bar{t}_{2(2)}$.
4. Employ 1 measurement at every 7-point interval to create the 200 $\times$ 6 data set $t_{1(7)}$.
5. Average subsequent points 4 at every 4-point interval to create the 200 $\times$ 6 data set $\bar{t}_{4(4)}$.
6. Average 8 subsequent points to create the 200 $\times$ 6 data set $\bar{t}_{8(0)}$.

Averaging several subsequent points is expected to have a smoothing effect, as well as reducing the correlation.

The result of simulation is given in the last 6 rows of Table 2 ($\rho = 0.45$). We see that averaging subsequent 8 measurements is not sufficient. The only acceptable procedure is analyzing $t_{1(7)}$ or $\bar{t}_{4(4)}$, of which the latter is expected to be desirable regarding the power. From the table, we also notice that the linear component in the error sum of squares is more easily affected by the correlation, showing upper bias for evaluation of errors that result in the reduced false positive for S^{+l} and $S_{R \times C}$, as compared with S^{-} and $S^{-l}_{R \times C}$, respectively. Adding a linear component will also be preferable regarding the power if the sample size is not large enough.

To confirm these assumptions, we created a nonnull model with 4 types of subgroups, using cosine curves with maximum deviations -2σ (highly convex), $-2/3\sigma$ (convex), 0 (flat), and $4/3\sigma$ (concave), respectively. This was suggested by the real data analyzed in Section 9. We give in Table 3 the proportion of significant results, at each number K, of classes from 2 to 7 for each of the testing procedures. We performed a simulation for various numbers of subjects, but only give here the case of $n = 30$, allocated by the proportion of 2:4:3:1 to the 4 types of subgroups representing actual situation.

For reference, we include the original data set $t_{1(0)}$ in the table, which cannot actually be used because of its extreme false positive. From Table 3, we see that there is a definite difference in the power between the data sets $t_{1(7)}$ and $\bar{t}_{4(4)}$, as expected. The power of the latter procedure is large enough compared to the original data set, even for the sample size as small as 30. This is partly because the averaging process reduces

Table 2

Estimated type I error rate for the procedure with significance level 0.05

t	data set	$\rho = 0$				$\rho = 0.45$			
		S^{-}	S^{+l}	$S_{R \times C}$	$S^{-l}_{R \times C}$	S^{-}	S^{+l}	$S_{R \times C}$	$S^{-l}_{R \times C}$
48	$t_{1(0)}$	0.058	0.057	0.058	0.057	1.000	1.000	1.000	1.000
12	$t_{1(3)}$	0.040	0.046	0.049	0.054	0.262	0.175	0.162	0.172
	$\bar{t}_{2(2)}$	0.060	0.074	0.069	0.064	0.494	0.288	0.262	0.311
	$\bar{t}_{4(0)}$	0.047	0.046	0.048	0.048	0.994	0.874	0.778	0.874
6	$t_{1(7)}$	0.057	0.058	0.063	0.062	0.055	0.050	0.050	0.049
	$\bar{t}_{4(4)}$	0.055	0.055	0.056	0.055	0.065	0.055	0.053	0.060
	$\bar{t}_{8(0)}$	0.062	0.056	0.064	0.062	0.324	0.140	0.155	0.255

Table 3
Power comparisons (n = 30)

Correlation	data set	K	S^-	S^{+l}	$S_{R\times C}$	$S_{R\times C}^{-l}$
$\rho = 0$	$t_{1(0)}$	2	1.00	1.00	1.00	1.00
		3	1.00	1.00	1.00	1.00
		4	1.00	1.00	1.00	1.00
		5	1.00	1.00	1.00	1.00
		6	1.00	1.00	1.00	1.00
		7	1.00	1.00	1.00	1.00
	$t_{1(7)}$	2	0.04	0.03	0.01	0.01
		3	0.18	0.15	0.11	0.11
		4	0.24	0.23	0.21	0.21
		5	0.27	0.25	0.25	0.26
		6	0.29	0.28	0.28	0.29
		7	0.30	0.29	0.31	0.31
	$\bar{t}_{4(4)}$	2	0.85	0.91	0.67	0.48
		3	0.99	1.00	0.99	0.98
		4	1.00	1.00	1.00	1.00
		5	1.00	1.00	1.00	1.00
		6	1.00	1.00	1.00	1.00
		7	1.00	1.00	1.00	1.00
$\rho = 0.45$	$t_{1(7)}$	2	0.04	0.03	0.01	0.01
		3	0.18	0.15	0.11	0.12
		4	0.24	0.21	0.20	0.22
		5	0.27	0.24	0.24	0.25
		6	0.29	0.25	0.27	0.28
		7	0.30	0.27	0.29	0.30
	$\bar{t}_{4(4)}$	2	0.38	0.41	0.18	0.11
		3	0.72	0.78	0.71	0.66
		4	0.81	0.86	0.83	0.80
		5	0.84	0.88	0.87	0.84
		6	0.85	0.89	0.89	0.86
		7	0.85	0.90	0.90	0.86

the noisy variation without losing the shape information, such as convex or concave. The number of the degrees of freedom for error is large enough, even after reducing the number of points to 6.

The somewhat reduced power of $\bar{t}_{4(4)}$ in the case of the correlated data is due to averaging four points with a positive correlation. This does not reduce the variance to 1/4, as in the independent case. It is, however, inevitable, and not a defect in the procedure, that there is some loss of information in the correlated data compared to independent data when the sample size is the same. Adding a linear component to the error sum of squares gains us some power, especially when it is not large enough. Also, in that case, S^{+l} seems more preferable to $S_{R\times C}$. We therefore employ the procedure based on $\bar{t}_{4(4)}$ and S^{+l} in the application to real data in Section 9.

One may still feel a bit of loss of power by not using all the data. However, for noisy and correlated data such as blood pressure measurements, it is often necessary to average or skip some of data. This is seen in Degaute et al. (1994), where, for example, all measurements representing an isolated increase of more than 50% over the previous measurement were deleted and replaced by linear interpolation between the previous and following measurements. Although our systematic statistic in the numerator is not as sensitive to the outliers compared to those in the approach taken by Degaute et al., it is still

worthwhile to consider a smoothing process that will not disturb the convex and concave patterns of interest here. So, the analyzing data set $\bar{t}_{4(4)}$ should be justified. Actually, a few outlying observations seen in the plots of raw data are not seen in the averaged data of Figure 1.

9. Application

We apply the procedure proposed in the previous section to the real data of $n = 203$ and $t = 48$. After that discussed in Section 8, we formed a 203×6 data set $\bar{t}_{4(4)}$ from the original data set by taking the averages of the 4 successive measurements at every 4-point interval. We show the obtained significance level and the percent contribution for each number of classifications in Table 4.

First, we note that all the classifications beyond $K = 2$ show highly significant results. As for the mean profiles of the obtained subgroups, however, we see the classification $K = 3$ fails to detect a flat pattern, and that classification $K = 5$ is very similar to $K = 4$. A major difference between $K = 5$ and $K = 4$ is that the highly convex subgroup of $K = 5$ is isolated from the moderately convex subgroup of $K = 4$. The classification of $K = 6$ gains us very little in the generalized distance over $K = 5$ and will not make sense. Noting also that the increase of the percent contribution is 6.9% from $K = 3$ to 4, whereas it is only 3.0% from $K = 4$ to $K = 5$, the first choice will be the classification into 4 subgroups. This also happens to correspond with the classification employed by Ohkubo et al. (1997), with the subgroups 1, 2, 3, and 4 corresponding to the extreme dipper, dipper, nondipper, and inverted dipper, respectively. The classification into the 5 subgroups is also very easy to interpret. It is isolating the ultra-extreme dipper (UED) group from the ED. Up to now, it has been pointed out clinically that the stroke rate is higher in the ND and ID subgroups (O'Brien et al., 1988) and that the mortality risk is higher in the ID subgroup. However, no significant difference is observed between the ED and D subgroups (Ohkubo et al., 1997). We, however, believe the UED subgroup might suggest a higher risk than the ED and D subgroups, and encourage a further clinical research.

For reference, we applied a procedure based on the cumulative chi-squared statistic χ^{*2} introduced in Hirotsu (1991) for distinguishing up- and downward tendencies. We give in Figure 1 the profiles of the obtained 5 classes based on $\chi^{\dagger 2}$ and χ^{*2}, respectively; this visualizes well the difference between the two methods. It is inevitable that the concave class looks a bit irregular, since it is the clinically most abnormal and unstable class.

10. Discussion

The proposed algorithm worked well for the data set $\bar{t}_{4(4)}$. This is because 6 points are enough to characterize the convex or concave pattern. The averaging process works well to reduce the noise without loosing the important shape information. Our method is based on the systematic statistic detecting those characteristics. Thus, the present algorithm gives a good objective procedure for classifying subjects, replacing rather subjective judgments by clinical doctors. The shape information obtained is very useful for giving so-called tailor-made medicine to individual hypertensive patients.
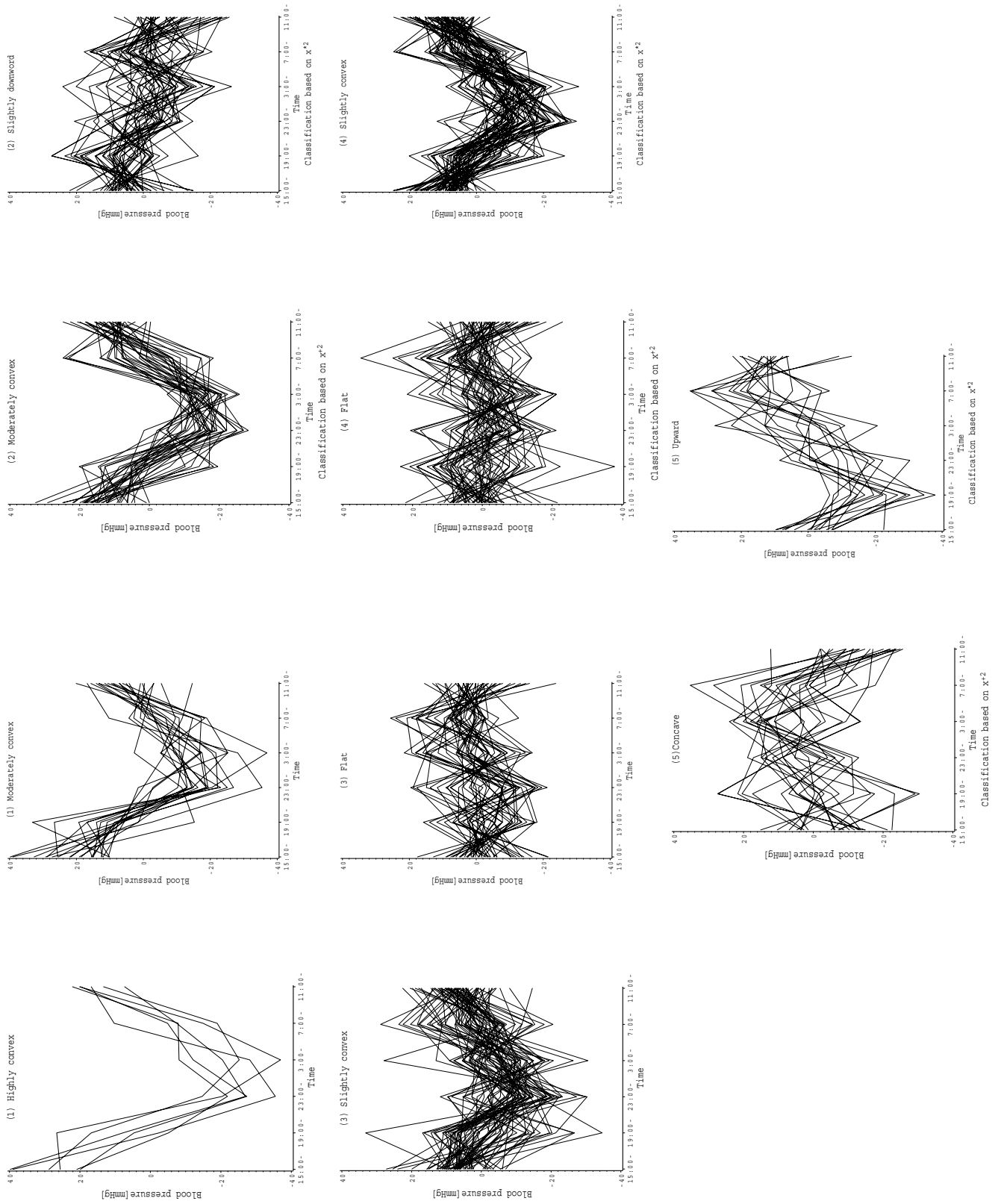
Figure 1. Profiles of the resulting five subgroups for the data set $\bar{t}_{4(4)}$.

Table 4
Classification of data set $\bar{t}_{4(4)}$ based on $\chi^{\dagger 2}$ 15:00–14:30

K	Significance level by S^{+l}	Mean profile						n	Generalized distance	Percent contribution
		15:00	19:00	23:00	3:00	7:00	11:00			
2	1.12E-1	148	138	126	128	139	147	117	9.43E4	50.3
		137	138	136	137	141	132	86		
3	8.99E-5	157	145	128	129	143	154	65	1.24E5	66.3
		137	132	126	129	136	135	102		
		137	143	146	145	146	131	36		
4	1.34E-6	161	146	126	128	144	157	39	1.37E5	73.2
		142	134	125	129	137	142	78		
		138	138	133	133	140	134	62		
		134	139	144	148	145	126	24		
5	1.92E-7	173	152	119	122	136	160	6	1.43E5	76.2
		158	145	129	128	145	155	39		
		141	134	125	129	136	141	74		
		137	137	133	133	141	134	60		
		134	139	144	148	145	126	24		
6	1.99E-7	173	152	119	122	136	160	6	1.43E5	79.3
		158	145	129	128	145	155	39		
		140	132	124	128	136	140	75		
		144	126	128	147	151	129	15		
		137	143	136	132	139	136	47		
		133	141	147	147	145	127	21		

Finally, we made a similar analysis of the heart rate data obtained simultaneously with the blood pressure measurements, using the polygraphic recording. We found the resulting two-way classification of blood pressure and heart rate shows a significantly high correlation. This observation supports the thesis of study by Degaute et al. (1994): that a genetic effect is observed simultaneously on the same profile parameters, such as the amplitude of the diastolic blood pressure and heart rate.

Résumé

Une méthode est proposée pour classifier des sujets selon l'évolution convexe, plate ou concave de leur pression sanguine au cours de 24 heures. L'obtention d'une telle classification est utile pour détecter les sujets qui montrent des évolutions incorrectes et pour fournir des traitements médicaux appropriés. Une statistique appropriée est proposée pour détecter un changement systématique au cours du temps ainsi qu'une statistique avec sa caractéristique inverse appropriée pour évaluer la variation résiduelle. La méthode est basée sur le rapport de ces deux types de statistique et montre de bons résultats sur des données réelles donnant une classification de sujets dans quatre types de sous-groupes : le plongeur extrême, le plongeur, le non plongeur et le plongeur inversé. Elle suggère l'existence possible d'un sous-groupe de plongeur ultra extrême.

References

Degaute, J. P., Cauter, E. V., Borne, P., and Linkowski, P. (1994). Twenty-four-hour blood pressure and heart rate profiles in humans. *Hypertension* **23,** 244–253.

Hirotsu, C. (1982). Use of cumulative efficient scores for testing ordered alternatives in discrete models. *Biometrika* **69,** 567–577.

Hirotsu, C. (1983). An approach to defining the pattern of interaction effects in a two-way layout. *Annals of the Institute of Statistical Mathematics A* **35,** 77–90.

Hirotsu, C. (1991). An approach to comparing treatments based on repeated measures. *Biometrika* **78,** 583–594.

Hirotsu, C. and Marumo, K. (2002). Changepoint analysis as a method for isotonic inference. *Scandinavian Journal of Statistics* **29,** 125–138.

Johnson, D. E. and Graybill, F. A. (1972). An analysis of a two-way model with interaction and no replication. *Journal of the American Statistical Association* **67,** 862–868.

Kobrin, I., Oigman, W., Kumar, A., Venture, H. O., Messeri, F. H., Frohlic, E. D., and Dunn, F. G. (1984). Diurnal variation of blood pressure in elderly patients with essential hypertension. *Journal of the American Geriatric Society* **32,** 896–899.

Mann, S., Altman, D. G., Raftery, E. B., and Bannister, R. (1983). Circadian variation of blood pressure in autonomic failure. *Circulation* **68,** 477–483.

Millar-Craig, M. W., Bishop, C. N., and Raftery E. B. (1978). Circadian variation of blood-pressure. *Lancet* **1,** 795–797.

O'Brien, E., Sheridan, J., and O'Malley, K. (1988). Dippers and non-dippers. *Lancet* **2,** 397.

Ohkubo, T., Imai, Y., and Tsuji, I., et al (1997). Relation between nocturnal decline in blood pressure and mortality. *American Journal of Hypertension* **10,** 1201–1207.

Rao, C. R. (1973). *Linear Statistical Inference and Its Application,* 2nd edition. New York: Wiley.

Sen, A. and Srivastava, M. S. (1975). Some one-sided tests for change in level. *Technometrics* **17,** 61–64.

Shimizu, K. (1994). Time series analysis of diurnal blood pressure variation—Its basis and clinical application to the hypertensive aged subjects (in Japanese). *Journal of Medicine, Keio University* **71,** 249–262.

Yamamoto, Y., Akiguchi, I., Oiwa, K., Satoi, H., and Kimura, J. (1995). Diminished nocturnal blood pressure decline and lesion site in cerebrovascular disease. *Stroke* **26,** 829–833.

APPENDIX A

The matrix $\boldsymbol{Q}_t$ is formed from the matrix $\boldsymbol{L}'_t$ as $\boldsymbol{Q}'_t = \text{diag}(\eta_k^{1/2}\delta'_k)(\boldsymbol{L}'_t\boldsymbol{L}_t)^{-1}\boldsymbol{L}'_t$, where $\text{diag}(d_k)$ is a diagonal matrix for standardization with $d_k = \eta_k^{1/2}\delta'_k$ as its kth diagonal element, $\eta_k - 1 = 2(k-1)(t-2-k)/(3(t-1))$, $\delta'_k = -[6t(t^2-1)/k(k+1)(t-k)(t-k-1)\{(t-1)(2k+1) -2(k^2-1)\}]^{1/2}$ $k = 1,\ldots,t-2$. The motivation of $\boldsymbol{Q}'_t$ is due to the complete class lemma of Hirotsu (1982), which asserts that a class of admissible tests for the ordered alternative $H_2(2)$ is formed by all the tests that are increasing in every element of $(\boldsymbol{L}'_t\boldsymbol{L}_t)^{-1}\boldsymbol{L}'_t(\boldsymbol{y}_i - \boldsymbol{y}_{i'})$ and with a convex acceptance region. Since our alternative is two-sided, it is natural to make sum of squares of the elements of $(\boldsymbol{L}'_t\boldsymbol{L}_t)^{-1}\boldsymbol{L}'_t(\boldsymbol{y}_i - \boldsymbol{y}_{i'})$. The constants $\eta_k^{1/2}\delta'_k$ are for normalizing, so that the statistic $\chi^{\dagger 2}(i;i')$ is expanded as

$$\chi^{\dagger 2}(i;i') = \sigma^2\left\{\frac{2t(t+1)}{1\cdot 2\cdot 3\cdot 4}\chi^2_{(2)} + \frac{2t(t+1)}{2\cdot 3\cdot 4\cdot 5}\chi^2_{(3)}\right.$$
$$\left. + \cdots + \frac{2t(t+1)}{(t-2)(t-1)t(t+1)}\chi^2_{(t-1)}\right\}, \quad (A.1)$$

and easy to understand, where $\chi^2_{(k)}$ is a chi-squared statistic with 1 d.f. detecting a departure in the direction of the Chebyshev's kth-order orthogonal polynomial in time and independent of each other, $k = 2,\ldots,t-1$; see Hirotsu (1986) for details. The coefficient

$$\gamma_{(k)} = 2t(t+1)/\{(k-1)k(k+1)(k+2)\} \quad (A.2)$$

is the kth latent root of $\boldsymbol{Q}'_t\boldsymbol{Q}_t$, declining rapidly in k so that $\chi^{\dagger 2}(i;i')$ is detecting mainly the quadratic and cubic depar-

tures. Anyway, the matrix $(\boldsymbol{L}'_t\boldsymbol{L}_t)^{-1}\boldsymbol{L}'_t$ is the key for those characteristics. It is interesting to note that its τth row represents a contrast, suggesting a slope change at the $(\tau+1)$th period. As a simple example, we give

$$\left(\boldsymbol{L}'_5\boldsymbol{L}_5\right)^{-1}\boldsymbol{L}'_5 = \begin{pmatrix} \frac{1}{5}(2 & -2 & -1 & 0 & 1) \\ \frac{1}{10}(4 & -1 & -6 & -1 & 4) \\ \frac{1}{5}(1 & 0 & -1 & -2 & 2) \end{pmatrix}$$

with the first, second, and third rows suggesting a slope change at the second, third, and fourth period, respectively. Recently, the relationship between the convexity hypothesis and the slope change model has been proved more generally for one-way layout, in Hirotsu and Marumo (2002). The present approach is extending it to an interaction problem.

APPENDIX B

The component S_i^- of S^- is defined by $S_i^- = \sum_i \boldsymbol{v}'_i\boldsymbol{v}_i$, where $\boldsymbol{v}_i = \text{diag}\{(\eta_k\delta_k'^2)^{-1/2}\}\boldsymbol{L}'_t(\boldsymbol{y}_i - \bar{\boldsymbol{y}})$. This is simply the weighted sum of squares of the second-order differences in the subsequent measurements, and is less affected by the systematic components. The weights are chosen so as to make the statistic have an inverse characteristic to those sum of squares derived in Sections 4.1 to 4.3. To show this, it suffices to prove that $\boldsymbol{L}_t\text{diag}\{(\eta_k\delta_k'^2)^{-1}\}\boldsymbol{L}'_t$ is a Moor-Penrose inverse of the matrix $\boldsymbol{Q}_t\boldsymbol{Q}'_t$. It is, however, very easy to verify

$$\boldsymbol{Q}_t\boldsymbol{Q}'_t\boldsymbol{L}_t\text{diag}\left\{\left(\eta_k\delta_k'^2\right)^{-1}\right\}\boldsymbol{L}'_t\boldsymbol{Q}_t\boldsymbol{Q}'_t = \boldsymbol{Q}_t\boldsymbol{Q}'_t,$$

and other three conditions of Rao (1973, Chapter 1) for a Moor-Penrose inverse. Then, under the null hypothesis, we obviously have

$$S^- = \sigma^2\left\{\gamma_{(2)}^{-1}\chi^2_{(2)}(n-1) + \cdots + \gamma_{(t-1)}^{-1}\chi^2_{(t-1)}(n-1)\right\} \quad (B.1)$$

with $\chi^2_{(k)}(n-1)$ independent chi-squared variables, as defined with respect to (A.1), but with $n-1$ d.f., since S^- is the sum of squares over the n subjects. Now, the higher order terms obtain relatively larger weights, which is what we intended, since a change represented by a higher order polynomial is less systematic and should thus be more appropriate for evaluating the noise variation.