Information Criterion-Based Nonhierarchical Clustering

(Last Modified: April 13, 2015)

Isamu NAGAI[1*], Katsuyuki TAKAHASHI[2] and Hirokazu YANAGIHARA[3]

[1] *Department of International Liberal Studies, School of International Liberal Studies, Chukyo University*
101 Tokodachi, Kaizu-cho, Toyota-shi, Aichi 470-0393, Japan

[2] *Department of Social System and Management,*
Graduate School of Systems and Information Engineering, University of Tsukuba,
1-1-1 Tennodai, Ibaraki 305-8571, Japan

[3] *Department of Mathematics, Graduate School of Science, Hiroshima University*
1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8626, Japan

Abstract

In the analysis of actual data, it is important to determine whether there are clusters in the data. This can be done using one of several methods of cluster analysis, which can be roughly divided into hierarchical and nonhierarchical clustering methods. Nonhierarchical clustering can be applied to more types of data than can hierarchical clustering (see e.g., Saito & Yadohisa, 2006), and hence, in this paper, we focus on nonhierarchical clustering. In nonhierarchical clustering, the results heavily depend on the number of clusters, and thus it is very important to select the appropriate number of clusters. In the present paper, we propose a new information criterion for selecting the number of clusters. In order to do so, we clarify the relationship between nonhierarchical clustering and the multivariate linear regression model.

Key words: AIC; Cluster analysis; Information criterion; k-means procedure; Multivariate linear regression model; Nonhierarchical clustering.

1. Introduction

In practice, we often determine whether clusters exist before further analyzing a data set. However, this is highly intuitive, and a formal cluster analysis is one way to avoid subjectivity.

In a cluster analysis, n individuals with p-dimensional data are divided into several clusters. This can be done with either hierarchical clustering or nonhierarchical clustering. We will briefly illustrate these methods in the Section 2. Further details of cluster analysis can be found in the literature (e.g., Bijnen, 1973; Romesburg, 1984; Hastie, Tibshirani, and Friedman, 2008; and Everitt, Landau, Leese, and Stahl, 2011). Saito and Yadohisa (2006) point out that nonhierarchical clustering can deal with more types of data than can hierarchical clustering. Hence, in the present paper, we focus on nonhierarchical clustering. One popular method for nonhierarchical clustering is the k-means procedure that was proposed by MacQueen (1967). In nonhierarchical clustering, the number of clusters must be decided by the user, and, since the results are strongly affected by this, it is important to choose appropriately.

*Corresponding author E-mail: *inagai@lets.chukyo-u.ac.jp*

The number of clusters is often selected in an arbitrary manner or by an empirical rule, which may be based on a scatter plot or on various properties of the data. Several studies have proposed methods for selecting the number of clusters (see, e.g., Bozdogan, 1986), but these methods have assumed that the data clusters are hierarchical. This assumption is not always satisfied. In this paper, we propose a new information criterion that can be used for selecting the number of clusters without assuming that they are hierarchical.

To do this, we begin by clarifying the relation between nonhierarchical clustering and the multivariate linear regression model. Based on this relation, we will propose a criterion based on the Akaike information criterion (AIC; Akaike, 1973); as mentioned above, our proposed criterion does not assume a hierarchy in the data. Moreover, this criterion allows us to regard the number of explanatory variables as the number of clusters. The AIC consists of adding the negative twofold maximum log-likelihood to twice the number of independent parameters. In the multivariate linear regression model, the number of independent parameters is $m(2l + m + 1)$, where l is the number of explanatory variables, and m is the dimension of the response variables. However, in nonhierarchical clustering, there are more parameters. In a cluster analysis, the location of each individual data point is an independent parameter. Thus, by adding such a term to the AIC, we propose a new criterion for selecting the number of clusters.

The remainder of the present paper is organized as follows: In Section 2, we briefly illustrate cluster analysis and several cluster criteria for the clustering of data. We show the equivalent condition for renewing the clustered data, based on each cluster criterion. Further, we prove this equivalence in the Appendix. In Section 3, we show the relationship between cluster analysis and the multivariate linear regression model. Then, from an illustration of the AIC, we propose a new AIC-based criterion that includes information about the location of each data point. In Section 4, we perform a numerical simulation to demonstrate the new criterion.

2. Cluster Analysis

Let $\boldsymbol{y}_i$, $i = 1, \ldots, n$, be a p-dimensional data vector, where n is the number of individuals. One of various cluster analysis methods is often used to determine the clusters that exist in $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ (see, e.g., Hastie, Tibshirani, & Friedman, 2009, Section 14.3).

These methods can be roughly divided, as follows:

- Hierarchical clustering: Initially, each data point is regarded as a cluster, so there are n clusters. A clustering method is then used to combine clusters until there is only one cluster.
- Nonhierarchical clustering: The number of clusters is determined prior to beginning the cluster analysis. The data points $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ are divided into clusters either randomly or by using one of

various methods. The contents of each cluster are then evaluated using some cluster criterion, and the process is repeated until convergence is reached.

In general, nonhierarchical clustering can be applied to more types of data than can hierarchical clustering (see, e.g., Saito & Yadohisa, 2006). We note that the k-means procedure (MacQueen, 1967) is popular method for nonhierarchical clustering.

For nonhierarchical clustering, various cluster criteria have been proposed for renewing clusters (see, e.g., Marriott, 1982; Krzanowski & Marriott, 1995), and we will begin by introducing them. Let k be the number of clusters, let n_j be the number of individuals belonging to the jth cluster. Here, we note that $n = n_1 + \cdots + n_k$. Moreover, let C_j be the set of indices of the individuals in the jth cluster. For example, consider a clustered set in which $k = 2$; $\boldsymbol{y}_1$, $\boldsymbol{y}_2$, and $\boldsymbol{y}_4$ belong to the first cluster; and $\boldsymbol{y}_3$ and $\boldsymbol{y}_5$ belong to the second cluster. Then, $n_1 = 3$, $n_2 = 2$, $C_1 = \{1, 2, 4\}$, and $C_2 = \{3, 5\}$. In the present paper, we always assume $n_j \geq 1$ for all j; that is, each cluster includes more than one individual. We summarize the $C_1, \ldots, C_k$ as $G(k) = \{C_1, \ldots, C_k\}$. In addition, let

$$\boldsymbol{W}_j = \sum_{i \in C_j} (\boldsymbol{y}_i - \bar{\boldsymbol{y}}_j^{(k)})(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_j^{(k)})', \ \boldsymbol{W}(G(k)) = \sum_{j=1}^{k} \boldsymbol{W}_j,$$

where $\bar{\boldsymbol{y}}_j^{(k)} = \sum_{i \in C_j} \boldsymbol{y}_i / n_j$, and $\bar{\boldsymbol{y}}_j^{(k)}$ is the sample mean in the jth cluster. We note that $\boldsymbol{W}_j$ ($j = 1, \ldots, k$) and $\boldsymbol{W}(G(k))$ are symmetric matrices. Throughout this paper, we will also assume that $\det(\boldsymbol{W}_j) \neq 0$ for any $G(k)$, k, and j. Note that this assumption means that $\det(\boldsymbol{W}_j) > 0$, since $\boldsymbol{W}_j$ is a nonnegative-definite matrix for any $G(k)$, k, and j (for the definition of a nonnegative-definite matrix, see, e.g., Gentle, 2007, Section 8.3). From this assumption, we obtain $\det(\boldsymbol{W}(G(k))) > 0$ for any $G(k)$ and k (see, e.g., Lütkepohl, 1996, p. 55). Using $\boldsymbol{W}_j$ and $\boldsymbol{W}(G(k))$, Table 1 lists the various cluster criteria that been proposed previously.

Table 1; Cluster criteria for renewing clustered data

(i)	$\mathrm{tr}(\boldsymbol{W}(G(k)))$
(ii)	$\det(\boldsymbol{W}(G(k)))$
(iii)	$\sum_{j=1}^{k} \det(\boldsymbol{W}_j)^{1/p}$
(iv)	$\prod_{j=1}^{k} \det(\boldsymbol{W}_j)^{n_j}$
(v)	$n \log\{\det(\boldsymbol{W}(G(k)))\} - 2 \sum_{j=1}^{k} n_j \log(n_j)$
(vi)	$\sum_{j=1}^{k} [n_j \log\{\det(\boldsymbol{W}_j)\} - 2 n_j \log(n_j)]$

Of the cluster criteria listed here, (i) and (ii) are the most frequently used, since they are very simple. Criteria (iii) to (vi) were proposed by Krzanowski and Marriott (1995), and we will refer to any of these six as $CC(\boldsymbol{W}(G(k)))$. The algorithm for performing nonhierarchical clustering with any of the criteria $CC(\boldsymbol{W}(G(k)))$ is as follows:

1. Choose k, and select a cluster criterion $CC(\boldsymbol{W}(G(k)))$.
2. Divide $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ into k clusters, either randomly or by using a nonhierarchical clustering method, such as the k-means procedure.
3. Move $\boldsymbol{y}_r$, which is in the sth cluster, into the tth cluster ($s \neq t$) if $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$ holds for the selected criterion, where $G'(k) = \{C_1, \ldots, C_{s-1}, C'_s, C_{s+1}, \ldots, C_{t-1}, C'_t, C_{t+1}, \ldots, C_k\}$, C'_s is derived by deleting r from C_s, and C'_t is derived by adding r into C_t. Here, we note that $C'_s = C_s \setminus \{r\}$ and $C'_t = C_t \cup \{r\}$. When $\boldsymbol{y}_r$ is moved from the sth cluster into the tth cluster, we renew $G(k)$, C_s, and C_t as $G'(k)$, C'_s, and C'_t, respectively.
4. Repeat the previous renewal procedure until the selected $CC(\boldsymbol{W}(G(k)))$ converges to a minimum value. This produces the optimal clustering $\hat{G}(k)$. That is, the optimal clustering $\hat{G}(k)$ is derived from $\hat{G}(k) = \mathrm{argmin}_{G(k)}\, CC(\boldsymbol{W}(G(k)))$ for the selected cluster criterion and the given k.

Marriott (1982) showed the variation that occurred when the rth individual $\boldsymbol{y}_r$ is added to the jth cluster. From that, for each criterion, we can derive the condition for moving $\boldsymbol{y}_r$ from the sth cluster to the tth cluster. In Table 2, letting $\boldsymbol{a} = \{n_t/(n_t+1)\}^{1/2}(\boldsymbol{y}_r - \bar{\boldsymbol{y}}_t)$ and $\boldsymbol{b} = \{n_s/(n_s-1)\}^{1/2}(\boldsymbol{y}_r - \bar{\boldsymbol{y}}_s)$, we summarize the equivalent condition with $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$ for each criterion. From the definitions of the various matrices, we can see that for (i), $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$ is equivalent to the renewal condition (2.1). Yanagihara and Yoshimoto (2005) showed for (ii), $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$ is equivalent to the renewal condition (2.2). We prove in the Appendix that the remaining renewal conditions (2.3), (2.4), (2.5), and (2.6) are equivalent to the corresponding cluster criteria.

Table 2; The conditions for renewal of each cluster based on the corresponding $CC(\boldsymbol{W}(G(k)))$

$CC(\boldsymbol{W}(G(k)))$	renewal condition	
(i)	$\boldsymbol{a}'\boldsymbol{a} < \boldsymbol{b}'\boldsymbol{b}$	(2.1)
(ii)	$(\boldsymbol{a}-\boldsymbol{b})'\boldsymbol{W}(G(k))^{-1}(\boldsymbol{a}+\boldsymbol{b}) < \boldsymbol{a}'\boldsymbol{W}(G(k))^{-1}(\boldsymbol{a}\boldsymbol{b}'-\boldsymbol{b}\boldsymbol{a}')\boldsymbol{W}(G(k))^{-1}\boldsymbol{b}$	(2.2)
(iii)	$\boldsymbol{a}'\boldsymbol{W}_t^{-1}\boldsymbol{a}\det(\boldsymbol{W}_t) < \boldsymbol{b}'\boldsymbol{W}_s^{-1}\boldsymbol{b}\det(\boldsymbol{W}_s)$	(2.3)
(iv)	$(1+\boldsymbol{a}'\boldsymbol{W}_t^{-1}\boldsymbol{a})^{n_t+1}\det(\boldsymbol{W}_t) < (1-\boldsymbol{b}'\boldsymbol{W}_s^{-1}\boldsymbol{b})^{-(n_s-1)}\det(\boldsymbol{W}_s)$	(2.4)
(v)	$1+(\boldsymbol{a}-\boldsymbol{b})'\boldsymbol{W}(G(k))^{-1}(\boldsymbol{a}+\boldsymbol{b})-\boldsymbol{a}'\boldsymbol{W}(G(k))^{-1}(\boldsymbol{a}\boldsymbol{b}'-\boldsymbol{b}\boldsymbol{a}')\boldsymbol{W}(G(k))^{-1}\boldsymbol{b} < \left\{\frac{n_t}{n_s}\left(1-\frac{1}{n_s}\right)^{n_s-1}\left(1+\frac{1}{n_t}\right)^{n_t+1}\right\}^{2/n}$	(2.5)
(vi)	$(1-\boldsymbol{b}'\boldsymbol{W}_s^{-1}\boldsymbol{b})^{n_s-1}(1+\boldsymbol{a}'\boldsymbol{W}_t^{-1}\boldsymbol{a})^{n_t+1}\det(\boldsymbol{W}_t\boldsymbol{W}_s^{-1}) < \left\{\frac{n_t}{n_s}\left(1-\frac{1}{n_s}\right)^{n_s-1}\left(1+\frac{1}{n_t}\right)^{n_t+1}\right\}^2$	(2.6)

3. Criterion for Selecting the Number of Clusters

3.1. Relationship between cluster analysis and the multivariate linear regression model

Prior to proposing our new information criterion for selecting the number of clusters k, we will clarify the relationship between cluster analysis and the multivariate linear regression model.

Let $\boldsymbol{e}_j^{(r)}$ be an r-dimensional vector in which the jth element is one and the other elements are zero. Then, when the ith individual $\boldsymbol{y}_i$ is in the jth cluster (that is, $i \in C_j$), we set $\boldsymbol{x}_i^{(k)} = \boldsymbol{e}_j^{(k)}$ ($i = 1, \ldots, n;\ j = 1, \ldots, k$). Note that $\boldsymbol{x}_i^{(k)}$ also depends on k, since it depends on the results of the clustering. Furthermore, let $\boldsymbol{\xi}_j^{(k)}$ ($j = 1, \ldots, k$) be an unknown p-dimensional vector which means that the unknown center vector of the jth cluster when $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ are divided into k clusters. When $i \in C_j$ ($j = 1, \ldots, k$), $\boldsymbol{y}_i$ is closer to $\boldsymbol{\xi}_j^{(k)}$ than $\boldsymbol{\xi}_{j'}^{(k)}$ ($j' \neq j$). Thus, we can assume that $\boldsymbol{y}_i$ can be derived from the following model:

$$\boldsymbol{y}_i = \boldsymbol{\xi}_j^{(k)} + \boldsymbol{\varepsilon}_i = \boldsymbol{\Xi}^{(k)\prime} \boldsymbol{x}_i^{(k)} + \boldsymbol{\varepsilon}_i \quad (i = 1, \ldots, n), \tag{3.1}$$

where $\boldsymbol{\Xi}^{(k)} = (\boldsymbol{\xi}_1^{(k)}, \ldots, \boldsymbol{\xi}_k^{(k)})'$ is a $k \times p$ matrix. Suppose that $\boldsymbol{\varepsilon}_i \perp\!\!\!\perp \boldsymbol{\varepsilon}_j$ ($i \neq j$) and $\boldsymbol{\varepsilon}_i \sim N_p(\boldsymbol{0}_p, \boldsymbol{\Sigma})$, where $\boldsymbol{0}_p$ is a p-dimensional zero vector, and $\boldsymbol{\Sigma}$ is an unknown $p \times p$ covariance matrix with $\det(\boldsymbol{\Sigma}) \neq 0$. Then the model (3.1) can be expressed as $\boldsymbol{Y} \sim N_{n\times p}(\boldsymbol{X}^{(k)}\boldsymbol{\Xi}^{(k)}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)$, where $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)'$, $\boldsymbol{X}^{(k)} = (\boldsymbol{x}_1^{(k)}, \ldots, \boldsymbol{x}_n^{(k)})'$, and $\otimes$ indicates the Kronecker product (for the definition of the Kronecker product, see, e.g., Muirhead, 1982, p. 73). Note that $\mathrm{rank}(\boldsymbol{X}^{(k)}) = k$ for any k, since we assume that $n_j \geq 1$ holds for $j = 1, \ldots, k$. Hence, using dummy vectors $\boldsymbol{x}_1^{(k)}, \ldots, \boldsymbol{x}_n^{(k)}$ that are based on the clustered data, when k is provided, cluster analysis is equivalent to the multivariate linear regression model. Thus, by comparing several model selection criteria for the multivariate linear regression model for each k, we may select the number of clusters.

When we have $\boldsymbol{Y} \sim N_{n\times p}(\boldsymbol{X}^{(k)}\boldsymbol{\Xi}^{(k)}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)$, based on $\hat{G}(k)$ under the provided k, the log-likelihood function is derived, as follows:

$$g(\boldsymbol{\Xi}^{(k)}, \boldsymbol{\Sigma}|\boldsymbol{Y}, \boldsymbol{X}^{(k)}) = -\frac{1}{2}\left(np\log(2\pi) + n\log\{\det(\boldsymbol{\Sigma})\} + \mathrm{tr}\left\{(\boldsymbol{Y} - \boldsymbol{X}^{(k)}\boldsymbol{\Xi}^{(k)})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}^{(k)}\boldsymbol{\Xi}^{(k)})'\right\}\right).$$

Then, we can also define $\hat{G}(k)$ based on $\hat{\boldsymbol{X}}^{(k)} = \mathrm{argmax}_{\boldsymbol{X}^{(k)} \in \mathcal{M}(n,k)} \max_{\boldsymbol{\Xi}^{(k)}, \boldsymbol{\Sigma}} g(\boldsymbol{\Xi}^{(k)}, \boldsymbol{\Sigma}|\boldsymbol{Y}, \boldsymbol{X}^{(k)})$, where $\mathcal{M}(n,k) = \{\boldsymbol{A} \in \mathbb{M}(n,k) | \boldsymbol{e}_i^{(n)\prime}\boldsymbol{A}\boldsymbol{e}_j^{(k)} = 0 \text{ or } 1, \sum_{j=1}^k \boldsymbol{e}_i^{(n)\prime}\boldsymbol{A}\boldsymbol{e}_j^{(k)} = 1 \text{ for any } i\}$ and $\mathbb{M}(n,k)$ is an $n \times k$ real matrix. Here, $\boldsymbol{X}^{(k)} \in \mathcal{M}(n,k)$ means each individual belongs to only one cluster, and, for a given k, all clusters have more than one individual.

The maximum-likelihood estimators for $\boldsymbol{\Xi}^{(k)}$ and $\boldsymbol{\Sigma}$ are derived as $\hat{\boldsymbol{\Xi}}^{(k)} = (\boldsymbol{X}^{(k)\prime}\boldsymbol{X}^{(k)})^{-1}\boldsymbol{X}^{(k)\prime}\boldsymbol{Y}$ and $\hat{\boldsymbol{\Sigma}}^{(k)} = n^{-1}\boldsymbol{Y}'\{\boldsymbol{I}_n - \boldsymbol{X}^{(k)}(\boldsymbol{X}^{(k)\prime}\boldsymbol{X}^{(k)})^{-1}\boldsymbol{X}^{(k)\prime}\}\boldsymbol{Y}$, respectively. Thus, the maximized log-likelihood function can be obtained as $g(\hat{\boldsymbol{\Xi}}^{(k)}, \hat{\boldsymbol{\Sigma}}^{(k)}|\boldsymbol{Y}, \boldsymbol{X}^{(k)}) = -[np\{\log(2\pi) + 1\} + n\log\{\det(\hat{\boldsymbol{\Sigma}}^{(k)})\}]/2$. Then, we can also define $\hat{G}(k)$ based on $\hat{\boldsymbol{X}}^{(k)} = \mathrm{argmax}_{\boldsymbol{X}^{(k)} \in \mathcal{M}(n,k)}\, g(\hat{\boldsymbol{\Xi}}^{(k)}, \hat{\boldsymbol{\Sigma}}^{(k)}|\boldsymbol{Y}, \boldsymbol{X}^{(k)})$. Hence, k is likely to be optimized when we minimize $n\det(\hat{\boldsymbol{\Sigma}}^{(k)})$. Therefore, in the next subsection, we will propose an information criterion that is based on this term.

Additionally, since $\bar{\boldsymbol{y}}_j^{(k)} = \hat{\boldsymbol{\Xi}}^{(k)\prime}\boldsymbol{x}_i^{(k)}$, and $\boldsymbol{I}_n - \boldsymbol{X}^{(k)}(\boldsymbol{X}^{(k)\prime}\boldsymbol{X}^{(k)})^{-1}\boldsymbol{X}^{(k)\prime}$ is a symmetric and idempotent matrix, the relationship between $\hat{\boldsymbol{\Sigma}}^{(k)}$ and $\boldsymbol{W}(\hat{G}(k))$ can be derived, as follows:

$$\begin{aligned} n\hat{\boldsymbol{\Sigma}}^{(k)} &= (\boldsymbol{Y} - \boldsymbol{X}^{(k)}\hat{\boldsymbol{\Xi}}^{(k)})'(\boldsymbol{Y} - \boldsymbol{X}^{(k)}\hat{\boldsymbol{\Xi}}^{(k)}) \\ &= \sum_{j=1}^{k}\sum_{i\in C_j}(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_j^{(k)})(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_j^{(k)})' \\ &= \boldsymbol{W}(\hat{G}(k)). \end{aligned}$$

3.2. New AIC-type criterion for selecting the number of clusters

Note that in the ordinary multivariate linear regression model, the only independent parameters are in $\boldsymbol{\Xi}$ and $\boldsymbol{\Sigma}$. For the model (3.1), based on $g(\hat{\boldsymbol{\Xi}}^{(k)}, \hat{\boldsymbol{\Sigma}}^{(k)}|\boldsymbol{Y}, \boldsymbol{X}^{(k)})$, and neglecting the constant term, we propose the following AIC-based information criterion:

$$\mathrm{AIC}(\hat{G}(k)) = n\log\{\det(\boldsymbol{W}(\hat{G}(k)))\} + 2p\left(k + \frac{p+1}{2}\right). \tag{3.2}$$

This criterion is frequently used for selecting explanatory variables in the multivariate linear regression model. (For more details of the AIC, see, e.g., Konishi and Kitagawa, 2008; Rao, Toutenburg, Shalabh, and Heumann, 2008).

However, when minimizing the above criterion, we often select a large number of clusters k. The reason for this is that $n\log\{\det(\boldsymbol{W}(\hat{G}(k)))\}$ decreases more rapidly than $2p\{k+(p+1)/2\}$ increases as k becomes large.

Furthermore, in the ordinary multivariate linear regression model, the AIC is considered by combining $g(\hat{\boldsymbol{\Xi}}^{(k)}, n^{-1}\boldsymbol{W}(\hat{G}(k))|\boldsymbol{Y}, \boldsymbol{X}^{(k)})$ with the number of independent parameters, which are only in $\boldsymbol{\Xi}$ and $\boldsymbol{\Sigma}$, and neglecting the constant terms. On the other hand, in cluster analysis, we can regard the location of the ith individual as an independent parameter. Thus, clustering the data is equivalent to assigning a value of 0 or 1 to $k-1$ parameters for each individual. Since the individual belongs to the kth cluster when all $k-1$ parameters are zeros, only $k-1$ parameters can be chosen; that is, there are $k-1$ independent parameters for each individual. This means that there are $n(k-1)$ new independent parameters, which correspond to the location of the n individuals.

However, we note that the weight for the new parameter $n(k-1)$ may not be equal to 2, which is the weight for ordinary parameters. Hence, we will let α be a nonnegative tuning parameter, and we propose the following AIC-type criterion:

$$\mathrm{AIC}(\hat{G}(k)|\alpha) = \mathrm{AIC}(\hat{G}(k)) + \alpha n(k-1). \tag{3.3}$$

Note that $\mathrm{AIC}(\hat{G}(k)|0) = \mathrm{AIC}(\hat{G}(k))$, and $\mathrm{AIC}(\hat{G}(k)|2)$ corresponds to this criterion when the weight for the new parameter is 2. In the numerical studies presented in the next section, we will choose α. It is reasonable to expect that more clusters will be selected when α is small, and that fewer will be selected when α is large.

Using the k-means procedure for a given k, $\hat{G}(k)$ can be derived, and then we can regard the model (3.1) as the true model. The number of clusters k is then selected by $\mathrm{AIC}(\hat{G}(k)|\alpha)$ for a fixed α. In order to select k, we used the following algorithm:

1. Set α and the candidate number of clusters $\mathcal{K}$; for example, $\mathcal{K} = \{1, 2, 3, 5, 10, 15\}$.
2. Obtain $\hat{G}(k)$ based on some criterion by using the algorithm in Section 2 for each $k \in \mathcal{K}$.
3. Compare (3.3) for each $k \in \mathcal{K}$.
4. Select k by minimizing (3.3) for the given α.

In this algorithm and in (3.3), the tuning parameter α is fixed. However, in the numerical studies in the next section, we will derive the α that is optimized for selecting the number of clusters.

4. Numerical Studies

4.1. Simulation

In this subsection, we evaluate the performance of the proposed criterion for selecting k by presenting the results of some simulations.

Let k^* be the true number of clusters, let $\boldsymbol{D}_p = \mathrm{diag}(1, \ldots, p)$ be a $p \times p$ diagonal matrix, and let $\boldsymbol{\Delta}_p(\rho)$ be a $p \times p$ matrix in which the (i, j)th element is $\rho^{|i-j|}$. Then $\boldsymbol{Y}$, which is the data, is generated from $N_{n\times p}(\boldsymbol{\Theta}, \boldsymbol{\Sigma}^* \otimes \boldsymbol{I}_n)$ for each repetition, where $\boldsymbol{\Sigma}^* = \boldsymbol{D}_p^{1/2}\boldsymbol{\Delta}_p(\rho)\boldsymbol{D}_p^{1/2}$, $\boldsymbol{\Theta} = 5\delta(\boldsymbol{\theta}_1\boldsymbol{1}'_{n_1}, \ldots, \boldsymbol{\theta}_{k^*}\boldsymbol{1}'_{n_{k^*}})'$ is an $n \times p$ matrix, δ is a scale parameter, and $\boldsymbol{1}_r$ is an r-dimensional vector in which each of the elements is one. Here, $\boldsymbol{\theta}_j$ $(j = 1, \ldots, k^*)$ controls a p-dimensional vector at the true center of each true cluster. We set the $\boldsymbol{\theta}_j$ as follows: $\boldsymbol{\theta}_1 = \boldsymbol{0}_p$, $\boldsymbol{\theta}'_2\boldsymbol{e}_i^{(p)} = \sin(i)$, $\boldsymbol{\theta}'_3\boldsymbol{e}_i^{(p)} = \log((-1)^i + i + 1)$, and $\boldsymbol{\theta}'_4\boldsymbol{e}_i^{(p)} = (-1)^i + 0.5$ $(i = 1, \ldots, p)$. We note that δ controls the scale of $\boldsymbol{\Theta}$. The candidate number of clusters was set to $\mathcal{K} = \{1, 2, 3, \ldots, 10\}$, and the weight for the new penalty term α was set to $0, 0.1, 0.5, 1, 1.5,$ and 2.

In each iteration, the data $\boldsymbol{Y}$ is was divided into k $(\in \mathcal{K})$ clusters by using the k-means procedure; we used the 'kmeans' function in the R programming language. The clustered data were renewed based on the condition (2.2), which evaluates the results of clustering based on $\det(\boldsymbol{W}(\hat{G}(k)))$. Note that $\det(\boldsymbol{W}(\hat{G}(k)))$ is also in the first term of $\mathrm{AIC}(\hat{G}(k)|\alpha)$ in (3.3) and the other criteria. Using the clustering results, $\boldsymbol{W}(\hat{G}(k))$ and $\boldsymbol{X}^{(k)}$ were derived for each k. $\mathrm{AIC}(\hat{G}(k)|\alpha)$ was also derived for each α. In order to compare the various criteria for selecting the number of clusters, we computed the Bayesian information criterion (BIC; Schwarz, 1978), the consistent AIC (CAIC; Bozdogan, 1987), and the corrected AIC (AIC$^{\mathrm{C}}$; Sugiura,1978):

$$\mathrm{BIC}(\hat{G}(k)) = n\log\{\det(\boldsymbol{W}(\hat{G}(k)))\} + p\log(n)\left(k + \frac{p+1}{2}\right),$$

$$\mathrm{CAIC}(\hat{G}(k)) = \mathrm{BIC}(\hat{G}(k)) + p\left(k + \frac{p+1}{2}\right),$$

$$\mathrm{AIC}^{\mathrm{C}}(\hat{G}(k)) = \mathrm{AIC}(\hat{G}(k)) + \frac{2p(k+p+1)}{n-k-p-1}\left(k + \frac{p+1}{2}\right).$$

These criteria were initially proposed for selecting the explanatory variables in the multivariate linear regression model. For each of the $\mathcal{K}$, we select $\hat{k}$ by minimizing each criterion; the result of clustering $\hat{G}(\hat{k})$ is also derived in each repetition. The results from each criterion are compared by using the predicted error (PE), defined as follows:

$$\mathrm{PE} = \frac{1}{p}\log\{\det(\hat{\boldsymbol{\Sigma}}^{(\hat{k})})\} + \log(2\pi) + \frac{1}{p}\mathrm{tr}\left(\boldsymbol{\Sigma}^{*}\hat{\boldsymbol{\Sigma}}^{(\hat{k})^{-1}}\right) + \frac{1}{np}\mathrm{tr}\left\{(\boldsymbol{\mathcal{M}} - \hat{\boldsymbol{Y}})'(\boldsymbol{\mathcal{M}} - \hat{\boldsymbol{Y}})\hat{\boldsymbol{\Sigma}}^{(\hat{k})^{-1}}\right\},$$

where $\hat{\boldsymbol{\Sigma}}^{(\hat{k})}$ is obtained by $n^{-1}\boldsymbol{W}(\hat{G}(\hat{k}))$, $\hat{\boldsymbol{Y}} = \boldsymbol{X}^{(\hat{k})}(\boldsymbol{X}^{(\hat{k})\prime}\boldsymbol{X}^{(\hat{k})})^{-1}\boldsymbol{X}^{(\hat{k})\prime}\boldsymbol{Y}$, and $\boldsymbol{X}^{(\hat{k})}$ is derived from the cluster results $\hat{G}(\hat{k})$ for each repetition. After 10,000 repetitions, the average PE values and the probability (%) of correctly selecting the number of clusters were used for comparing these criteria.

We present the (rounded) results in Tables 3 to 12. In Tables 3 to 7, for each case, the minimum value is in bold, and next smallest value is in italics. In Tables 8 to 12, for each case, the maximum value is in bold, and next largest is in italics. That is, in each table, the best score is in bold, and the second best is in italics. In these tables, for simplicity, BIC($\hat{G}(k)$), CAIC($\hat{G}(k)$), and AIC$^{\mathrm{C}}$($\hat{G}(k)$) are written as BIC, CAIC, and AIC$^{\mathrm{C}}$, respectively.

Please insert Tables 3 to 12 around here.

We first consider the results based on the PEs that are listed in Tables 3 to 7. We will focus on the results for $k^* = 2$, as shown in Table 3. As p increases, the values for AIC($\hat{G}(k)$) (= AIC($\hat{G}(k)|0$)), AIC($\hat{G}(k)|0.1$), BIC($\hat{G}(k)$), CAIC($\hat{G}(k)$), and AIC$^{\mathrm{C}}$($\hat{G}(k)$) always decrease. In addition, as p increases, the values for AIC($\hat{G}(k)|1$) and AIC($\hat{G}(k)|1.5$) increase, except when $\rho = 0.95$, and they decrease when $\rho = 0.95$. As δ increases, the results for AIC($\hat{G}(k)|\alpha$) ($\alpha = 0.5, 1, 1.5, 2$) almost always decrease. The results for the other criteria also tend to decrease as δ increases, at least in some situations. The results for AIC($\hat{G}(k)|\alpha$) ($\alpha = 1, 1.5, 2$) almost always decrease as ρ increases. We note that the results for AIC($\hat{G}(k)$), AIC($\hat{G}(k)|0.1$), BIC($\hat{G}(k)$), CAIC($\hat{G}(k)$), and AIC$^{\mathrm{C}}$($\hat{G}(k)$) almost always follow the same tendencies as ρ increases. However, in several situations, the results for the other criteria tend to increase as ρ increases. Comparing the results, we see that AIC($\hat{G}(k)|1$) always yields the best results, followed by AIC($\hat{G}(k)|1.5$). Furthermore, both AIC($\hat{G}(k)|2$) and AIC($\hat{G}(k)|0.5$) work well in almost every situation. On the other hand, AIC($\hat{G}(k)$) and other the criteria cannot be used directly for selecting the number of clusters based on the PE values with $k^* = 2$.

Thus, we will focus on the PEs when $k^* = 4$, which are shown in Tables 4 to 7. As p increases, the results for AIC($\hat{G}(k)$), AIC($\hat{G}(k)|0.1$), BIC($\hat{G}(k)$), CAIC($\hat{G}(k)$), and AIC$^{\mathrm{C}}$($\hat{G}(k)$) always decrease.

The results of AIC($\hat{G}(k)|2$) are small when $\rho = 0.95$. The results of AIC($\hat{G}(k)|1$) and AIC($\hat{G}(k)|1.5$) are also small when $\rho = 0.95$, except in the case $(n_1, n_2, n_3, n_4, \delta) = (30, 50, 30, 30, 1)$. As δ increases, the results for AIC($\hat{G}(k)$), AIC($\hat{G}(k)|0.1$), AIC($\hat{G}(k)|2$), BIC($\hat{G}(k)$), CAIC($\hat{G}(k)$), and AIC$^{\mathrm{C}}$($\hat{G}(k)$) always decrease. The results for AIC($\hat{G}(k)|0.5$) and AIC($\hat{G}(k)|1.5$) also become small in most cases. In many cases, the results for AIC($\hat{G}(k)|1$) decrease as δ increases. The results for AIC($\hat{G}(k)|2$) decrease somewhat as ρ increases, and the results for AIC($\hat{G}(k)|1.5$) decrease in many cases. In addition, the results for AIC($\hat{G}(k)$), AIC($\hat{G}(k)|0.1$), BIC($\hat{G}(k)$), CAIC($\hat{G}(k)$), and AIC$^{\mathrm{C}}$($\hat{G}(k)$) almost always show the same tendency as ρ increases. We note that in all situations considered, either AIC($\hat{G}(k)|0.5$) or AIC($\hat{G}(k)|1$) gave the best results. As when $k^* = 2$, we see that BIC($\hat{G}(k)$), CAIC($\hat{G}(k)$), AIC$^{\mathrm{C}}$($\hat{G}(k)$), and AIC($\hat{G}(k)$) cannot be used directly to select the number of clusters.

Next, we consider the probability (%) of correctly selecting the number of categories; this is shown in Tables 8 to 12. In the results in Table 8, when $k^* = 2$, AIC($\hat{G}(k)|1$) and AIC($\hat{G}(k)|1.5$) almost always select the correct number of clusters, and AIC($\hat{G}(k)|0.5$) is always either the best or second best method. Tables 9 to 12 show the results when $k^* = 4$. Based on these results, we see that AIC($\hat{G}(k)|1$) is almost always the best method, followed by AIC($\hat{G}(k)|0.5$).

These results indicate the necessity of the term $\alpha n(k-1)$, which is the parameter for the location of each individual. Further, from these results, we recommend using either AIC($\hat{G}(k)|0.5$), AIC($\hat{G}(k)|1$), or AIC($\hat{G}(k)|1.5$) for selecting the number of clusters.

4.2. An analysis of real data

For a cluster analysis of actual data, we used the 'iris' data set (Fisher, 1936), which is built into the R language, and is frequently used as test data for cluster analyses. The iris data set has 150 individuals data points, which are based on three types of iris, and hence have three natural clusters. For each individual, the following information was recorded: sepal length, sepal width, petal length, petal width, and the name of the type of iris. For the various information criteria, we used the various length and width values to select the number of clusters (types).

In order to compare the criteria, we selected the number of clusters independently for each trial. For each trial, we randomly deleted one individual from each group. Then, based on the remaining data, we selected the number of clusters by minimizing each of the criteria. We note that each trail was based on the data from 147 individuals, and we repeated the clustering process 10,000 times.

We set the candidate number clusters to be $\mathcal{K} = \{1, 2, 3, 4, 5\}$. When we used AIC($\hat{G}(k)|\alpha$), given in (3.3), we set α to be $0, 0.1, 0.5, 1, 1.5, 2$.

In the Table 13, we list how many times each candidate number of clusters was selected. Furthermore, as in Tables 3 to 12, for simplicity, BIC($\hat{G}(k)$), CAIC($\hat{G}(k)$), and AIC$^{\mathrm{C}}$($\hat{G}(k)$) are written as BIC, CAIC, and AIC$^{\mathrm{C}}$, respectively.

Please insert Table 13 around here.

$\mathrm{AIC}(\hat{G}(k))$, $\mathrm{BIC}(\hat{G}(k))$, $\mathrm{CAIC}(\hat{G}(k))$, $\mathrm{AIC}^{\mathrm{c}}(\hat{G}(k))$, and $\mathrm{AIC}(\hat{G}(k)|0.1)$ tend to select a large number of clusters, and thus the data are divided more minutely. $\mathrm{CAIC}(\hat{G}(k))$ selects fewer clusters than does $\mathrm{BIC}(\hat{G}(k))$, and $\mathrm{BIC}(\hat{G}(k))$ tends to select fewer than does $\mathrm{AIC}^{\mathrm{c}}(\hat{G}(k))$. Thus, overall, $\mathrm{AIC}(\hat{G}(k)|1.5)$ and $\mathrm{AIC}(\hat{G}(k)|2)$ tend to select fewer clusters. Moreover, $\mathrm{AIC}(\hat{G}(k)|1)$ selected the true number of clusters the most frequently. Thus, $\mathrm{AIC}(\hat{G}(k)|1)$ is the best method for selecting the number of clusters, and, as can be seen in the table, the second best methods is $\mathrm{AIC}(\hat{G}(k)|0.5)$.

Thus, we recommend using $\mathrm{AIC}(\hat{G}(k)|0.5)$ or $\mathrm{AIC}(\hat{G}(k)|1)$ for selecting the number of clusters.

5. Conclusions

In practice, prior to formally analyzing a data set, it is common to begin by determining subjectively whether the data are clustered. This is then followed by a formal cluster analysis, which does not depend on the analyst's intuition. It is well known that there are two types of cluster analysis methods: hierarchical and nonhierarchical. We briefly discussed these in Section 2. Compared to hierarchical clustering, nonhierarchical clustering can be applied to more types of data (Saito and Yadohisa, 2006), and so, in this paper, we have focused on nonhierarchical clustering.

We note that for some nonhierarchical clustering methods, it is necessary for the user to provide the number of clusters, although there is no method for selecting this for an arbitrary data set. However, in Section 3 and (3.1), we showed the relationship between clustering analysis and the multivariate linear regression model, and using this, we proposed a new AIC-type criterion (3.3) for selecting the number of clusters. This criterion was derived by adding a new term $n(k-1)$ with the nonnegative weight α to the AIC (3.2).

By conducting numerical studies, we showed that the added term $n(k-1)$ is needed to reduce the predicted error and to select the correct number of clusters. By inspecting the simulation results, we recommend using $\mathrm{AIC}(\hat{G}(k)|0.5)$, $\mathrm{AIC}(\hat{G}(k)|1)$, or $\mathrm{AIC}(\hat{G}(k)|1.5)$ to select the number of clusters. Furthermore, from the results of analyzing the 'iris' data set (using built in data set in the R language), we note that $\mathrm{AIC}(\hat{G}(k)|1)$ is the best criterion for selecting the number of clusters.

Based on numerical studies, we recommend using $\mathrm{AIC}(\hat{G}(k)|0.5)$ or $\mathrm{AIC}(\hat{G}(k)|1)$ for selecting the number of nonhierarchical clusters.

Appendix: Proof of the Renewal Condition for the Cluster Criterion in Nonhierarchical Clustering

In Section 2, we illustrated the cluster criterion for nonhierarchical clustering, and we then considered the cluster criteria (iii), (iv), (v), and (vi), as listed in Table 1. We now consider whether $\boldsymbol{y}_r$ in the sth cluster moves to the tth cluster ($s \neq t$), based on each criterion. Here, we recall some notation that was previously defined: $G(k) = \{C_1, \ldots, C_k\}$, C_i has the indices of the individuals in the ith cluster, $G'(k) = \{C_1, \ldots, C_{s-1}, C_s', C_{s+1}, \ldots, C_{t-1}, C_t', C_{t+1}, \ldots, C_k\}$, C_s' is derived by deleting r from C_s, and C_t' is derived by adding r to C_t. Note that $n_s \geq 2$, since $n_j \geq 1$ is always assumed. Then, the condition for moving $\boldsymbol{y}_r$ in the sth cluster to the tth cluster is $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$ in each cluster criterion. In this section, we prove that, for each cluster criterion, this condition coincides with the renewal conditions listed in Table 2: (2.3), (2.4), (2.5), and (2.6).

To begin, we prepare the relationship between $\boldsymbol{W}(G(k))$ and $\boldsymbol{W}(G'(k))$. Here, we also recall that $\boldsymbol{W}_j = \sum_{i \in C_j} (\boldsymbol{y}_i - \bar{\boldsymbol{y}}_j)(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_j)'$ and $\boldsymbol{W}(G(k)) = \sum_{j=1}^k \boldsymbol{W}_j$, where $\bar{\boldsymbol{y}}_j = \sum_{i \in C_j} \boldsymbol{y}_i / n_j$, and n_j is the number of individuals in the jth criterion. Note that the difference between $G(k)$ and $G'(k)$ is only the content of the sth cluster and the tth cluster. Furthermore, let $\boldsymbol{W}_{j'} = \sum_{i \in C_j'} (\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{j'})(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{j'})'$, where $\bar{\boldsymbol{y}}_{j'}$ is the sample mean of C_j'. Here, $\bar{\boldsymbol{y}}_{\ell'}$ $(\ell' \neq s', t')$ is the same as $\bar{\boldsymbol{y}}_\ell$. Then, from the definition of $\boldsymbol{W}(G(k))$, we can see that

$$\boldsymbol{W}(G'(k)) = \sum_{\substack{j=1,\ldots,k \\ j \neq s,t}} \sum_{i \in C_j} (\boldsymbol{y}_i - \bar{\boldsymbol{y}}_j)(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_j)' + \boldsymbol{W}_{t'} + \boldsymbol{W}_{s'}, \tag{A.1}$$

since $\boldsymbol{W}_{t'} = \sum_{i \in C_t'} (\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{t'})(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{t'})'$ $\boldsymbol{W}_{s'} = \sum_{i \in C_s'} (\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{s'})(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{s'})'$, and $\bar{\boldsymbol{y}}_{\ell'} = \bar{\boldsymbol{y}}_\ell$ for $\ell \neq s, t$. Since C_s' and C_t' are derived by deleting and adding r, we have $\bar{\boldsymbol{y}}_{s'} = \sum_{i \in C_s'} \boldsymbol{y}_i / (n_s - 1)$ and $\bar{\boldsymbol{y}}_{t'} = \sum_{i \in C_t'} \boldsymbol{y}_i / (n_t + 1)$, and we obtain the following results:

$$\bar{\boldsymbol{y}}_{t'} = \frac{n_t \bar{\boldsymbol{y}}_t + \boldsymbol{y}_r}{n_t + 1} \tag{A.2}$$

$$\bar{\boldsymbol{y}}_{s'} = \frac{n_s \bar{\boldsymbol{y}}_s - \boldsymbol{y}_r}{n_s - 1}, \tag{A.3}$$

since $\bar{\boldsymbol{y}}_s = \sum_{i \in C_s} \boldsymbol{y}_i / n_s$ and $\bar{\boldsymbol{y}}_t = \sum_{i \in C_t} \boldsymbol{y}_i / n_t$. Hence, we can calculate $\boldsymbol{W}_{t'}$ and $\boldsymbol{W}_{s'}$ in equation (A.1) by using these results.

Using (A.2), we derive $\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{t'} = \boldsymbol{y}_i - \bar{\boldsymbol{y}}_t - (\boldsymbol{y}_r - \bar{\boldsymbol{y}}_t)/(n_t + 1)$. Then, the following result is obtained:

$$\begin{aligned}
\boldsymbol{W}_{t'} &= \sum_{i \in C_t} (\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{t'})(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{t'})' + (\boldsymbol{y}_r - \bar{\boldsymbol{y}}_{t'})(\boldsymbol{y}_r - \bar{\boldsymbol{y}}_{t'})' \\
&= \sum_{i \in C_t} \left\{ \boldsymbol{y}_i - \bar{\boldsymbol{y}}_t - \frac{\boldsymbol{y}_r - \bar{\boldsymbol{y}}_t}{n_t + 1} \right\} \left\{ \boldsymbol{y}_i - \bar{\boldsymbol{y}}_t - \frac{\boldsymbol{y}_r - \bar{\boldsymbol{y}}_t}{n_t + 1} \right\}' + \left(\frac{n_t}{n_t + 1} \right)^2 (\boldsymbol{y}_r - \bar{\boldsymbol{y}}_t)(\boldsymbol{y}_r - \bar{\boldsymbol{y}}_t)' \\
&= \boldsymbol{W}_t + \boldsymbol{a}\boldsymbol{a}',
\end{aligned}$$

since $\boldsymbol{a} = \{n_t/(n_t+1)\}^{1/2}(\boldsymbol{y}_r - \bar{\boldsymbol{y}}_t)$, $\sum_{i \in C_t} (\boldsymbol{y}_i - \bar{\boldsymbol{y}}_t)(\boldsymbol{y}_r - \bar{\boldsymbol{y}}_t)' = \boldsymbol{O}_p$, and $\sum_{i \in C_t} (\boldsymbol{y}_r - \bar{\boldsymbol{y}}_t)(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_t)' = \boldsymbol{O}_p$, where $\boldsymbol{O}_p$ is a $p \times p$ zero matrix. By a similar calculation, using (A.3), we derive

$$\begin{aligned}
\boldsymbol{W}_{s'} &= \sum_{i \in C_s} (\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{s'})(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_{s'})' - (\boldsymbol{y}_r - \bar{\boldsymbol{y}}_{s'})(\boldsymbol{y}_r - \bar{\boldsymbol{y}}_{s'}) \\
&= \sum_{i \in C_s} \left\{ \boldsymbol{y}_i - \bar{\boldsymbol{y}}_s + \frac{\boldsymbol{y}_r - \bar{\boldsymbol{y}}_s}{n_s - 1} \right\} \left\{ \boldsymbol{y}_i - \bar{\boldsymbol{y}}_s + \frac{\boldsymbol{y}_r - \bar{\boldsymbol{y}}_s}{n_s - 1} \right\}' - \left(\frac{n_s}{n_s - 1} \right)^2 (\boldsymbol{y}_r - \bar{\boldsymbol{y}}_s)(\boldsymbol{y}_r - \bar{\boldsymbol{y}}_s)' \\
&= \boldsymbol{W}_s - \boldsymbol{b}\boldsymbol{b}',
\end{aligned}$$

since $\boldsymbol{b} = \{n_s/(n_s-1)\}^{1/2}(\boldsymbol{y}_r-\bar{\boldsymbol{y}}_s)$, $\sum_{i\in C_s}(\boldsymbol{y}_i-\bar{\boldsymbol{y}}_s)(\boldsymbol{y}_r-\bar{\boldsymbol{y}}_s)' = \boldsymbol{O}_p$, and $\sum_{i\in C_s}(\boldsymbol{y}_r-\bar{\boldsymbol{y}}_s)(\boldsymbol{y}_i-\bar{\boldsymbol{y}}_s)' = \boldsymbol{O}_p$.

From (A.1), these results imply

$$\boldsymbol{W}(G'(k)) = \boldsymbol{W}(G(k)) + \boldsymbol{a}\boldsymbol{a}' - \boldsymbol{b}\boldsymbol{b}'.$$

This directly shows that $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$, which is based on (i), coincides with the renewal condition (2.1).

Letting $\boldsymbol{E} = \boldsymbol{W}(G(k))^{-1}(\boldsymbol{a},\boldsymbol{b})\{(1,0)'(1,0) + (0,1)'(0,-1)\}(\boldsymbol{a},\boldsymbol{b})'$, since $0 < \det(\boldsymbol{W}(G(k)))$ for any $G(k)$, and k is assumed, the following equation holds:

$$\det(\boldsymbol{W}(G'(k))) = \det(\boldsymbol{W}(G(k)))\det(\boldsymbol{I}_p + \boldsymbol{E}). \tag{A.4}$$

Let λ_i be the ith eigenvalue of $\boldsymbol{E}$, then $\det(\boldsymbol{I}_p + \boldsymbol{E}) = \prod_{i=1}^{p}(1+\lambda_i)$. From Lütkepuhl (1996, p. 65), the nonzero eigenvalues in $\lambda_1,\ldots,\lambda_p$ are equal to ν_1 and ν_2, which are the eigenvalues of a 2×2 matrix $\boldsymbol{F} = \{(1,0)'(1,0) + (0,1)'(0,-1)\}(\boldsymbol{a},\boldsymbol{b})'\boldsymbol{W}_G^{-1}(\boldsymbol{a},\boldsymbol{b})$. Thus, $\det(\boldsymbol{I}_p + \boldsymbol{E}) = 1 + \nu_1 + \nu_2 + \nu_1\nu_2$. Since ν_1 and ν_2 are eigenvalues of $\boldsymbol{F}$, from the Cayley-Hamilton theorem, ν_1 and ν_2 are the solutions to the following quadratic equation:

$$\nu^2 - \mathrm{tr}(\boldsymbol{F})\nu + \det(\boldsymbol{F}) = 0.$$

Thus, from the relationship between the solution and the coefficients of the above equation, we obtain the following equations:

$$\begin{aligned}\nu_1 + \nu_2 &= \mathrm{tr}(\boldsymbol{F}) = (\boldsymbol{a}-\boldsymbol{b})'\boldsymbol{W}(G(k))^{-1}(\boldsymbol{a}+\boldsymbol{b}),\\ \nu_1\nu_2 &= \det(\boldsymbol{F}) = -\boldsymbol{a}'\boldsymbol{W}(G(k))^{-1}(\boldsymbol{a}\boldsymbol{b}' - \boldsymbol{b}\boldsymbol{a}')\boldsymbol{W}(G(k))^{-1}\boldsymbol{b}.\end{aligned}$$

Hence, we obtain

$$\det(\boldsymbol{I}_p + \boldsymbol{E}) = 1 + (\boldsymbol{a}-\boldsymbol{b})'\boldsymbol{W}(G(k))^{-1}(\boldsymbol{a}+\boldsymbol{b}) - \boldsymbol{a}'\boldsymbol{W}(G(k))^{-1}(\boldsymbol{a}\boldsymbol{b}' - \boldsymbol{b}\boldsymbol{a}')\boldsymbol{W}(G(k))^{-1}\boldsymbol{b}. \tag{A.5}$$

We directly obtain that the renewal condition (2.2) coincides with $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$ based on (ii) (this was previously shown by Yanagihara and Yoshimoto, 2005).

Using these results, we can prove $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$ based on the cluster criterion (iii) is equivalent to the renewal condition (2.3). When we use the cluster criterion (iii), since it is organized as a summation of $\det(\boldsymbol{W}_i)^{1/p}$, satisfying $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$ coincides with satisfying $\det(\boldsymbol{W}_{s'})^{1/p} + \det(\boldsymbol{W}_{t'})^{1/p} < \det(\boldsymbol{W}_s)^{1/p} + \det(\boldsymbol{W}_t)^{1/p}$. This condition is equivalent to $1 < \exp\{\det(\boldsymbol{W}_s)^{1/p} - \det(\boldsymbol{W}_{s'})^{1/p} + \det(\boldsymbol{W}_t)^{1/p} - \det(\boldsymbol{W}_{t'})^{1/p}\} = \exp\{(\det(\boldsymbol{W}_s) - \det(\boldsymbol{W}_{s'}) + \det(\boldsymbol{W}_t) - \det(\boldsymbol{W}_{t'}))/p\}$. Thus, we consider the condition that $0 < \det(\boldsymbol{W}_s) - \det(\boldsymbol{W}_{s'}) + \det(\boldsymbol{W}_t) - \det(\boldsymbol{W}_{t'})$.

From Schott (2005, Problem 7.14), the following equation holds:

$$\det(\boldsymbol{P} + \boldsymbol{Q}) = \{1 + \mathrm{tr}(\boldsymbol{P}^{-1}\boldsymbol{Q})\}\det(\boldsymbol{P}),$$

for any nonsingular matrix $\boldsymbol{P}$ and some appropriate dimensional matrix $\boldsymbol{Q}$ with $\text{rank}(\boldsymbol{Q}) = 1$. Since $\boldsymbol{W}_{t'} = \boldsymbol{W}_t + \boldsymbol{a}\boldsymbol{a}'$, $\boldsymbol{W}_s = \boldsymbol{W}_{s'} + \boldsymbol{b}\boldsymbol{b}'$, $\text{rank}(\boldsymbol{a}\boldsymbol{a}') = 1$, $\text{rank}(\boldsymbol{b}\boldsymbol{b}') = 1$, and we assume $\boldsymbol{W}_i$ is always a nonsingular matrix, we obtain

$$\det(\boldsymbol{W}_{t'}) = (1 + \boldsymbol{a}'\boldsymbol{W}_t^{-1}\boldsymbol{a}) \det(\boldsymbol{W}_t). \tag{A.6}$$

$$\det(\boldsymbol{W}_s) = (1 + \boldsymbol{b}'\boldsymbol{W}_{s'}^{-1}\boldsymbol{b}) \det(\boldsymbol{W}_{s'}). \tag{A.7}$$

Hence, we obtain

$$\det(\boldsymbol{W}_s) - \det(\boldsymbol{W}_{s'}) + \det(\boldsymbol{W}_t) - \det(\boldsymbol{W}_{t'}) = \boldsymbol{b}'\boldsymbol{W}_{s'}^{-1}\boldsymbol{b} \det(\boldsymbol{W}_{s'}) - \boldsymbol{a}'\boldsymbol{W}_t^{-1}\boldsymbol{a} \det(\boldsymbol{W}_t).$$

From Schott (2005, Corollary 1.7.2), the following equation holds for any nonsingular matrix $\boldsymbol{R}$:

$$(\boldsymbol{R} - \boldsymbol{c}\boldsymbol{c}')^{-1} = \boldsymbol{R}^{-1} + \frac{\boldsymbol{R}^{-1}\boldsymbol{c}\boldsymbol{c}'\boldsymbol{R}^{-1}}{1 - \boldsymbol{c}'\boldsymbol{R}^{-1}\boldsymbol{c}}, \tag{A.8}$$

for some vector $\boldsymbol{c}$ that is of suitable dimension, and such that $\boldsymbol{R} - \boldsymbol{c}\boldsymbol{c}'$ is a nonsingular matrix. We note that $\det(\boldsymbol{R} - \boldsymbol{c}\boldsymbol{c}') \neq 0$ is equivalent to $1 - \boldsymbol{c}'\boldsymbol{R}\boldsymbol{c} \neq 0$ (see, e.g., Siotani, Hayakawa, and Fujikoshi, 1985, A.1.3).

Here, since $\boldsymbol{W}_{s'} = \boldsymbol{W}_s - \boldsymbol{b}\boldsymbol{b}'$ is assumed to be a nonsingular matrix, we obtain $1 - \boldsymbol{b}'\boldsymbol{W}_s^{-1}\boldsymbol{b}' \neq 0$ (see, e.g., Siotani, Hayakawa, and Fujikoshi, 1985, A.1.3). Thus, from (A.8) and $\det(\boldsymbol{W}_s) \neq 0$, we have

$$\boldsymbol{W}_{s'}^{-1} = \boldsymbol{W}_s^{-1} + \frac{\boldsymbol{W}_s^{-1}\boldsymbol{b}\boldsymbol{b}'\boldsymbol{W}_s^{-1}}{1 - \boldsymbol{b}'\boldsymbol{W}_s^{-1}\boldsymbol{b}}.$$

Hence, we obtain $\boldsymbol{b}'\boldsymbol{W}_{s'}^{-1}\boldsymbol{b} = \boldsymbol{b}'\boldsymbol{W}_s^{-1}\boldsymbol{b}/(1 - \boldsymbol{b}'\boldsymbol{W}_s^{-1}\boldsymbol{b})$ and $1 + \boldsymbol{b}'\boldsymbol{W}_{s'}^{-1}\boldsymbol{b} = (1 - \boldsymbol{b}'\boldsymbol{W}_s^{-1}\boldsymbol{b})^{-1}$ (> 0). Using this and (A.7), we see that the renewal condition $0 < \boldsymbol{b}'\boldsymbol{W}_{s'}^{-1}\boldsymbol{b} \det(\boldsymbol{W}_{s'}) - \boldsymbol{a}'\boldsymbol{W}_t^{-1}\boldsymbol{a} \det(\boldsymbol{W}_t)$ is equivalent to

$$\boldsymbol{a}'\boldsymbol{W}_t^{-1}\boldsymbol{a} \det(\boldsymbol{W}_t) < \frac{\boldsymbol{b}'\boldsymbol{W}_s^{-1}\boldsymbol{b}}{1 - \boldsymbol{b}'\boldsymbol{W}_s^{-1}\boldsymbol{b}} \det(\boldsymbol{W}_{s'}).$$

From (A.7) and $1+\boldsymbol{b}'\boldsymbol{W}_{s'}^{-1}\boldsymbol{b} = (1-\boldsymbol{b}'\boldsymbol{W}_s^{-1}\boldsymbol{b})^{-1}$, we can prove the renewal condition (2.3) is equivalent to $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$, based on the cluster criterion (iii).

Next, we prove that the renewal condition (2.4) is equivalent to $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$ when we use the cluster criterion (iv). Recall that C_s' is derived by deleting r from C_s, and C_t' is derived by adding r to C_t. Hence, since the differences between $G(k)$ and $G'(k)$ are in the sth and tth clusters, we obtain $CC(\boldsymbol{W}(G(k))) = \det(\boldsymbol{W}_s)^{n_s} \det(\boldsymbol{W}_t)^{n_t} z$ and $CC(\boldsymbol{W}(G'(k))) = \det(\boldsymbol{W}_{s'})^{n_s - 1} \det(\boldsymbol{W}_{t'})^{n_t + 1} z$, where $z = \prod_{\substack{i=1,\ldots,k \\ i \neq s,t}} \det(\boldsymbol{W}_i)^{n_i}$ can be considered to be a constant. Note that $z > 0$, since $\det(\boldsymbol{W}_i) > 0$ for any i and $G(k)$. Thus, the condition for $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$ based on (iv) coincides with $\det(\boldsymbol{W}_{s'})^{n_s - 1} \det(\boldsymbol{W}_{t'})^{n_t + 1} < \det(\boldsymbol{W}_s)^{n_s} \det(\boldsymbol{W}_t)^{n_t}$. Since $\det(\boldsymbol{W}_i) > 0$ for any i and $G(k)$, from (A.6) and (A.7), this inequality is equivalent to

$$(1 + \boldsymbol{a}'\boldsymbol{W}_t^{-1}\boldsymbol{a})^{n_t + 1} \det(\boldsymbol{W}_t) < (1 + \boldsymbol{b}'\boldsymbol{W}_{s'}^{-1}\boldsymbol{b})^{n_s} \det(\boldsymbol{W}_{s'}).$$

Hence, since $1 + \boldsymbol{b}'\boldsymbol{W}_{s'}^{-1}\boldsymbol{b} = (1 - \boldsymbol{b}'\boldsymbol{W}_s^{-1}\boldsymbol{b})^{-1}$ and using (A.7), we can prove that (2.4) is coincides with $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$, based on the cluster criterion (iv).

Next, we prove that the renewal condition (2.5) coincides with $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$ when we use the cluster criterion (v). Based on (v), since $G(k)$ and $G'(k)$ differ in the sth and tth clusters, we can express this as

$$\begin{aligned} & n\{\log(\det(\boldsymbol{W}(G'(k)))) - \log(\det(\boldsymbol{W}(G(k))))\} \\ & < 2\{(n_s - 1)\log(n_s - 1) + (n_t + 1)\log(n_t + 1) - (n_s \log(n_s) + n_t \log(n_t))\}. \end{aligned}$$

The right-hand side in the above inequation is $n \log\{\det(\boldsymbol{I}_p + \boldsymbol{E})\}$, from (A.4). On the other hand, the left-hand side is equal to $2\log\{(n_s - 1)^{n_s - 1}(n_t + 1)^{n_t + 1} n_s^{-n_s} n_t^{-n_t}\}$. Thus, the above inequality coincides with

$$\det(\boldsymbol{I}_p + \boldsymbol{E}) < \left\{ \frac{(n_s - 1)^{n_s - 1}(n_t + 1)^{n_t + 1}}{n_s^{n_s} n_t^{n_t}} \right\}^{2/n}.$$

Hence, we obtain that the renewal condition (2.5) is equivalent to $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$, based on (v) from (A.5).

Finally, we prove that $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$, based on the cluster criterion (vi), is equivalent to (2.6). Considering the difference between $G(k)$ and $G'(k)$, this inequation is equivalent to

$$\begin{aligned} & (n_s - 1)\log(\det(\boldsymbol{W}_{s'})) + (n_t + 1)\log(\det(\boldsymbol{W}_{t'})) - 2\{(n_s - 1)\log(n_s - 1) + (n_t + 1)\log(n_t + 1)\} \\ & < n_s \log(\det(\boldsymbol{W}_s)) + n_t \log(\det(\boldsymbol{W}_t)) - 2\{n_s \log(n_s) + n_t \log(n_t)\}. \end{aligned}$$

Using (A.7), we obtain $\det(\boldsymbol{W}_{s'}) = (1 - \boldsymbol{b}'\boldsymbol{W}_s^{-1}\boldsymbol{b})\det(\boldsymbol{W}_s)$, since $1 + \boldsymbol{b}'\boldsymbol{W}_{s'}^{-1}\boldsymbol{b} = (1 - \boldsymbol{b}\boldsymbol{W}_s^{-1}\boldsymbol{b})^{-1}$. Furthermore, using (A.6), the above inequation is equivalent to

$$\begin{aligned} & (n_s - 1)\log(1 - \boldsymbol{b}'\boldsymbol{W}_s^{-1}\boldsymbol{b}) - \log(\det(\boldsymbol{W}_s)) + (n_t + 1)\log(1 + \boldsymbol{a}'\boldsymbol{W}_t^{-1}\boldsymbol{a}) + \log(\det(\boldsymbol{W}_t)) \\ & < 2\left\{ n_s \log\left(1 - \frac{1}{n_s}\right) - \log(n_s - 1) + n_t \log\left(1 + \frac{1}{n_t}\right) + \log(n_t + 1) \right\}. \end{aligned}$$

Then, since $\det(\boldsymbol{W}_s)^{-1} = \det(\boldsymbol{W}_s^{-1})$, this inequation can be rewritten, as follows:

$$\begin{aligned} & \log\left\{(1 - \boldsymbol{b}'\boldsymbol{W}_s^{-1}\boldsymbol{b})^{n_s - 1}(1 + \boldsymbol{a}'\boldsymbol{W}_t^{-1}\boldsymbol{a})^{n_t + 1}\det(\boldsymbol{W}_t\boldsymbol{W}_s^{-1})\right\} \\ & < \log\left[\left(1 - \frac{1}{n_s}\right)^{n_s}\left(1 + \frac{1}{n_t}\right)^{n_t}\frac{n_t + 1}{n_s - 1}\right]^2. \end{aligned}$$

Hence, we obtain that the renewal condition (2.6) coincides with $CC(\boldsymbol{W}(G'(k))) < CC(\boldsymbol{W}(G(k)))$, based on (vi).

References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory* (eds. B. N. Petrov & F. Csáki), 267–281, Akadémiai Kiadó, Budapest.

[2] Bijnen, E. J. (1973). *Cluster Analysis*, Tilburg University Press.

[3] Bozdogan, H. (1986). Multi-sample cluster analysis as an alternative to multiple comparison procedures. *Bulletin of Informatics and Cybernetics*, **22**, 95–130.

[4] Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika*, **52**, 345–370.

[5] Everitt, B. S. Landau, S., Leese, M. & Stahl, D. (2011). *Cluster Analysis* (5th ed.), Wiley Series in Probability and Statistics.

[6] Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, 179–188.

[7] Gentle, J. E. (2007). *Matrix Algebra* Theory, Computations, and Applications in Statistics, Springer Science+Business Media.

[8] Hastie, T. Tibshirani, R. & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.), Springer Science+Business Media.

[9] Konishi, S. & Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*, Springer Science+Business Media.

[10] Krzanowski, W. J. & Marriott, F. H. C. (1995). *Multivariate Analysis, Part 2: Classification, Covariance Structure, and Repeated Measurements*, John Wiley and Sons.

[11] Lütkepohl, H. (1996). *Handbook of Matrices*, John Wiley and Sons.

[12] MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (ed. J. Neyman), **1**, 281–298, Berkeley.

[13] Marriott, F. H. C. (1982). Optimization methods of cluster analysis. *Biometrika*, **69**, 417–421.

[14] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, John Wiley and Sons.

[15] Rao, C. R., Toutenburg, H. Shalabh & Heumann, C. (2008). *Linear Models and Generalization, Least Squares and Alternatives* (3rd Extended Ed.), Springer-Verlag Berlin Heidelberg.

[16] Romesburg, H. C. (1984). *Cluster Analysis for Researchers*, Lifetime Learning Publications Belmont, California.

[17] Saito, T. & Yadohisa, H. (2006). *Analysis of Relevant Data -Multidimensional Sscaling and Cluster Analysis*, Kyoritsu shuppan (in Japanese).

[18] Schott, J. R. (2005). *Matrix Analysis for Statistics*, John Wiley and Sons.

[19] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

[20] Siotani, M., Hayakawa, T. & Fujikoshi, Y. (1985). *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook*, American Science Press, Columbus, Ohio.

[21] Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics, Theory and Methods*, **7**, 13–26.

[22] Yanagihara, H. & Yoshimoto, A. (2005). Clustering individual growth patterns in even-aged forest stand. *Formath*, **4**, 49–70 (in Japanese).

Table 3; PEs based on each criterion ($k^* = 2$)

				$\mathrm{AIC}(\hat{G}(k)\|\alpha)$								
				the values of α								
(n_1, n_2)	p	δ	ρ	0	0.1	0.5	1	1.5	2	BIC	CAIC	$\mathrm{AIC^C}$
$(30, 30)$	2	1	0.25	12.32	12.18	4.49	**3.54**	**3.54**	*4.12*	12.23	12.17	12.28
			0.5	15.12	15.02	5.22	**3.62**	**3.62**	*3.95*	15.08	15.01	15.10
			0.95	26.33	25.63	*3.06*	**2.71**	**2.71**	3.13	26.02	25.61	26.22
		2	0.25	11.37	11.16	**3.24**	**3.24**	**3.24**	**3.24**	11.26	*11.15*	11.33
			0.5	13.30	13.19	*3.69*	**3.11**	**3.11**	**3.11**	13.27	13.18	13.29
			0.95	23.85	23.38	*4.42*	**2.13**	**2.13**	**2.13**	23.57	23.35	23.78
	5	1	0.25	6.46	6.42	4.48	**3.91**	**3.91**	*4.21*	6.33	6.01	6.41
			0.5	6.63	6.58	*4.12*	**3.65**	**3.65**	**3.65**	6.40	5.61	6.57
			0.95	10.64	10.40	2.17	**2.15**	*2.15*	2.18	10.11	8.99	10.37
		2	0.25	6.45	6.42	*4.53*	**3.81**	**3.81**	**3.81**	6.33	5.93	6.42
			0.5	7.06	7.04	*3.74*	**3.65**	**3.65**	**3.65**	6.90	6.40	7.03
			0.95	7.18	6.87	**2.01**	**2.01**	**2.01**	**2.01**	6.03	*4.59*	6.82
$(30, 50)$	2	1	0.25	10.09	9.93	*4.14*	**3.21**	**3.21**	**3.21**	10.03	9.98	10.07
			0.5	14.12	14.13	*3.97*	**3.19**	**3.19**	4.00	14.14	14.14	14.13
			0.95	19.23	18.88	5.34	**2.63**	**2.63**	*2.99*	19.11	19.01	19.23
		2	0.25	8.51	8.42	*3.91*	**3.19**	**3.19**	**3.19**	8.48	8.45	8.51
			0.5	10.47	*10.41*	**3.13**	**3.13**	**3.13**	**3.13**	10.46	10.43	10.47
			0.95	14.53	14.12	*2.48*	**2.09**	**2.09**	**2.09**	14.36	14.21	14.49
	5	1	0.25	6.18	6.17	4.33	**3.92**	**3.92**	*4.19*	6.13	6.02	6.18
			0.5	5.95	5.88	*4.03*	**3.73**	**3.73**	**3.73**	5.75	5.25	5.92
			0.95	7.22	7.11	*2.18*	**2.01**	**2.01**	**2.01**	6.87	6.00	7.19
		2	0.25	5.60	5.57	*3.94*	**3.81**	**3.81**	**3.81**	5.53	5.40	5.59
			0.5	6.06	6.00	*3.81*	**3.66**	**3.66**	**3.66**	5.93	5.73	6.04
			0.95	6.12	5.98	**1.98**	**1.98**	**1.98**	**1.98**	5.79	*5.17*	6.07
$(50, 30)$	2	1	0.25	9.71	9.61	4.34	**3.58**	**3.58**	*4.09*	9.68	9.64	9.70
			0.5	11.97	11.84	4.15	**3.37**	**3.37**	*3.91*	11.93	11.88	11.96
			0.95	18.38	18.17	**2.31**	**2.31**	**2.31**	*2.99*	18.30	18.23	18.35
		2	0.25	13.06	12.95	*3.93*	**3.27**	**3.27**	**3.27**	13.02	12.98	13.05
			0.5	10.52	10.43	*3.44*	**3.08**	**3.08**	**3.08**	10.49	10.45	10.52
			0.95	14.07	13.84	*2.57*	**2.05**	**2.05**	**2.05**	14.00	13.89	14.07
	5	1	0.25	5.55	5.50	*3.98*	**3.80**	**3.80**	**3.80**	5.41	5.23	5.53
			0.5	6.07	6.05	*3.99*	**3.73**	**3.73**	**3.73**	6.01	5.89	6.06
			0.95	8.05	7.85	*2.90*	**2.30**	**2.30**	**2.30**	7.51	6.44	7.98
		2	0.25	6.32	6.30	*3.96*	**3.88**	**3.88**	**3.88**	6.26	6.15	6.31
			0.5	6.17	6.12	*3.98*	**3.65**	**3.65**	**3.65**	6.04	5.75	6.15
			0.95	6.72	6.56	*2.18*	**1.98**	**1.98**	**1.98**	6.35	5.65	6.66
$(50, 50)$	2	1	0.25	9.15	8.99	*5.45*	**3.45**	**3.45**	**3.45**	9.09	9.06	9.14
			0.5	11.69	11.64	4.51	**3.39**	**3.39**	*3.97*	11.69	11.68	11.69
			0.95	16.04	15.68	*3.72*	**2.05**	**2.05**	**2.05**	15.99	15.86	16.04
		2	0.25	8.38	*8.32*	**3.18**	**3.18**	**3.18**	**3.18**	8.37	8.35	8.38
			0.5	9.93	9.90	*3.84*	**3.06**	**3.06**	**3.06**	9.93	9.93	9.93
			0.95	13.90	*13.65*	**2.04**	**2.04**	**2.04**	**2.04**	13.85	13.79	13.90
	5	1	0.25	6.22	6.19	*4.64*	**3.88**	**3.88**	**3.88**	6.17	6.06	6.21
			0.5	5.91	5.88	*4.07*	**3.82**	**3.82**	**3.82**	5.85	5.76	5.90
			0.95	7.71	7.48	*2.60*	**1.99**	**1.99**	**1.99**	7.22	6.38	7.68
		2	0.25	5.56	5.54	*3.80*	**3.79**	**3.79**	**3.79**	5.52	5.45	5.55
			0.5	5.92	5.92	**3.59**	**3.59**	**3.59**	**3.59**	5.90	*5.84*	5.92
			0.95	6.91	6.73	**1.97**	**1.97**	**1.97**	**1.97**	6.55	*6.08*	6.90

Table 4; PEs based on each criterion ($k^* = 4$; 1/4)

				$\mathrm{AIC}(\hat{G}(k)\|\alpha)$								
				the values of α								
(n_1, n_2, n_3, n_4)	p	δ	ρ	0	0.1	0.5	1	1.5	2	BIC	CAIC	$\mathrm{AIC^C}$
(30, 30, 30, 30)	2	0.5	0.25	7.01	7.00	**3.79**	*3.85*	4.63	4.63	7.01	7.01	7.01
			0.5	7.74	7.72	*3.62*	**3.41**	**3.41**	4.64	7.75	7.75	7.74
			0.95	6.84	6.57	*3.50*	**3.43**	3.56	4.60	6.71	6.67	6.82
		1	0.25	5.98	5.96	*3.58*	**3.54**	3.64	3.76	5.99	5.99	5.98
			0.5	5.75	5.74	*3.48*	**3.48**	3.55	3.73	5.76	5.76	5.76
			0.95	5.58	5.51	**3.11**	*3.22*	3.24	3.32	5.55	5.54	5.58
	5	0.5	0.25	4.89	4.88	**3.94**	*3.96*	4.25	4.73	4.87	4.84	4.89
			0.5	4.86	4.84	**4.07**	*4.09*	4.58	4.58	4.83	4.80	4.86
			0.95	3.46	3.40	2.99	**2.85**	*2.97*	3.81	3.39	3.36	3.45
		1	0.25	4.48	4.45	**3.88**	*3.90*	4.01	4.12	4.44	4.37	4.48
			0.5	4.26	4.24	**3.68**	*3.68*	3.80	3.86	4.22	4.16	4.26
			0.95	2.79	2.73	*2.02*	**2.02**	2.02	2.05	2.71	2.62	2.79
(30, 30, 30, 50)	2	0.5	0.25	7.68	7.64	*3.78*	**3.78**	4.55	4.55	7.68	7.68	7.68
			0.5	8.27	8.25	*4.23*	**3.90**	4.54	4.54	8.28	8.28	8.27
			0.95	7.80	7.81	*3.50*	**3.43**	4.37	4.37	7.82	7.82	7.81
		1	0.25	5.24	5.22	**3.50**	*3.54*	3.60	3.93	5.24	5.24	5.24
			0.5	5.98	5.95	**3.48**	*3.49*	3.55	3.66	5.98	5.98	5.98
			0.95	4.84	4.79	*3.19*	**3.17**	3.33	3.45	4.83	4.82	4.84
	5	0.5	0.25	4.54	4.51	**3.93**	*3.94*	4.01	4.69	4.51	4.48	4.54
			0.5	4.64	4.62	**3.96**	*3.97*	4.33	4.57	4.62	4.59	4.64
			0.95	3.55	3.47	2.94	**2.89**	*2.94*	3.74	3.46	3.40	3.54
		1	0.25	4.48	4.46	*3.86*	**3.86**	3.95	4.06	4.46	4.42	4.48
			0.5	4.13	4.10	**3.65**	**3.65**	*3.65*	3.76	4.10	4.04	4.13
			0.95	2.70	2.65	*2.03*	**2.03**	2.03	2.07	2.64	2.58	2.70
(30, 30, 50, 30)	2	0.5	0.25	7.23	7.16	*3.98*	**3.64**	4.54	4.60	7.22	7.21	7.23
			0.5	7.60	7.57	*3.90*	**3.71**	4.60	4.60	7.60	7.60	7.60
			0.95	7.59	7.46	*3.33*	**3.21**	**3.21**	4.57	7.54	7.52	7.59
		1	0.25	5.39	5.38	**3.48**	*3.53*	3.57	3.79	5.40	5.40	5.39
			0.5	4.82	4.79	**3.33**	*3.46*	3.51	3.69	4.82	4.82	4.82
			0.95	3.85	3.81	**3.08**	*3.22*	3.41	3.53	3.84	3.84	3.85
	5	0.5	0.25	4.42	4.39	**3.89**	*3.92*	4.03	4.71	4.39	4.34	4.42
			0.5	4.73	4.69	*4.14*	**4.11**	4.35	4.59	4.69	4.65	4.73
			0.95	3.44	3.41	3.06	**2.83**	*2.95*	3.03	3.40	3.38	3.44
		1	0.25	4.37	4.35	**3.85**	*3.87*	4.01	4.11	4.34	4.30	4.37
			0.5	4.30	4.28	**3.66**	*3.67*	3.78	3.91	4.27	4.23	4.30
			0.95	2.66	2.61	*2.02*	**2.02**	2.02	2.08	2.60	2.54	2.66
(30, 30, 50, 50)	2	0.5	0.25	7.10	7.08	*4.09*	**3.82**	4.51	4.51	7.10	7.10	7.10
			0.5	7.07	7.02	*3.82*	**3.75**	4.51	4.51	7.06	7.06	7.07
			0.95	8.08	8.07	*3.79*	**3.70**	4.49	4.49	8.10	8.11	8.08
		1	0.25	5.50	5.47	*3.60*	**3.60**	3.68	4.03	5.50	5.50	5.50
			0.5	5.81	5.80	3.65	**3.54**	*3.58*	3.88	5.82	5.82	5.81
			0.95	3.72	3.64	**3.02**	*3.09*	3.39	3.90	3.70	3.69	3.72
	5	0.5	0.25	4.61	4.60	**4.01**	*4.06*	4.38	4.69	4.60	4.59	4.61
			0.5	4.54	4.53	**3.69**	*3.71*	3.76	4.55	4.53	4.52	4.54
			0.95	3.98	3.91	3.38	**3.18**	*3.25*	3.87	3.91	3.88	3.97
		1	0.25	4.40	4.39	**3.86**	*3.88*	4.01	4.06	4.39	4.37	4.40
			0.5	4.29	4.28	**3.64**	*3.65*	3.75	3.84	4.28	4.26	4.29
			0.95	2.68	2.62	*2.00*	**2.00**	2.00	2.04	2.62	2.57	2.68

Table 5; PEs based on each criterion ($k^* = 4$; 2/4)

				$\mathrm{AIC}(\hat{G}(k)\|\alpha)$								
				the values of α								
(n_1, n_2, n_3, n_4)	p	δ	ρ	0	0.1	0.5	1	1.5	2	BIC	CAIC	$\mathrm{AIC^C}$
$(30, 50, 30, 30)$	2	0.5	0.25	6.62	6.57	**3.50**	*3.50*	4.66	4.66	6.61	6.61	6.62
			0.5	7.48	7.45	**3.62**	**3.62**	*4.64*	*4.64*	7.48	7.48	7.48
			0.95	6.85	6.64	*2.75*	**2.75**	2.75	4.62	6.78	6.76	6.84
		1	0.25	5.76	5.73	*3.52*	**3.50**	3.52	3.80	5.76	5.76	5.76
			0.5	5.22	5.19	**3.34**	*3.42*	3.45	3.54	5.22	5.22	5.22
			0.95	4.32	4.20	**2.96**	*3.02*	3.21	3.35	4.28	4.26	4.31
	5	0.5	0.25	4.55	4.54	**3.99**	*4.02*	4.10	4.72	4.53	4.52	4.55
			0.5	4.58	4.55	**3.75**	*3.83*	*3.83*	4.57	4.55	4.52	4.58
			0.95	4.02	3.97	3.72	**2.99**	**2.99**	*2.99*	3.97	3.96	4.01
		1	0.25	4.39	4.37	**3.86**	*3.89*	4.03	4.03	4.36	4.32	4.39
			0.5	4.23	4.21	**3.66**	*3.70*	3.88	3.88	4.20	4.17	4.23
			0.95	2.81	2.74	*2.03*	**2.03**	2.03	2.09	2.74	2.66	2.81
$(30, 50, 30, 50)$	2	0.5	0.25	6.84	6.81	*3.89*	**3.83**	4.49	4.58	6.84	6.83	6.84
			0.5	7.35	7.35	**3.57**	*3.58*	4.58	4.58	7.36	7.36	7.35
			0.95	7.39	7.28	*3.69*	**3.21**	**3.21**	4.40	7.35	7.33	7.39
		1	0.25	5.30	5.26	**3.51**	*3.52*	3.62	3.98	5.29	5.29	5.30
			0.5	5.33	5.31	3.61	**3.48**	*3.54*	3.95	5.34	5.33	5.33
			0.95	4.06	3.92	**3.01**	*3.03*	3.22	3.36	4.00	3.99	4.06
	5	0.5	0.25	4.60	4.59	**3.92**	*3.94*	4.08	4.69	4.59	4.57	4.60
			0.5	4.81	4.78	**3.83**	*3.86*	4.07	4.56	4.78	4.77	4.81
			0.95	3.37	3.33	3.02	**2.80**	*2.84*	2.97	3.33	3.32	3.37
		1	0.25	4.33	4.32	**3.83**	*3.85*	4.00	4.00	4.32	4.30	4.33
			0.5	4.13	4.11	**3.62**	*3.62*	3.81	3.81	4.11	4.09	4.13
			0.95	2.73	2.68	*2.02*	**2.02**	2.02	2.08	2.68	2.65	2.72
$(30, 50, 50, 30)$	2	0.5	0.25	6.77	6.74	*4.38*	**3.60**	4.64	4.64	6.77	6.77	6.77
			0.5	6.93	6.88	*3.73*	**3.69**	4.13	4.64	6.92	6.92	6.93
			0.95	7.62	7.55	3.73	**3.41**	*3.48*	4.19	7.60	7.59	7.62
		1	0.25	5.37	5.34	*3.56*	**3.53**	3.60	3.92	5.37	5.37	5.37
			0.5	5.15	5.13	**3.34**	*3.52*	3.57	3.94	5.15	5.15	5.15
			0.95	4.16	4.11	*2.98*	**2.89**	3.21	3.35	4.15	4.14	4.16
	5	0.5	0.25	4.63	4.62	**3.96**	*3.99*	4.16	4.71	4.62	4.61	4.63
			0.5	4.46	4.42	**3.82**	*3.87*	4.31	4.57	4.43	4.41	4.45
			0.95	3.63	3.55	*2.98*	**2.83**	*2.98*	2.99	3.55	3.53	3.63
		1	0.25	4.37	4.35	**3.84**	*3.89*	4.02	4.02	4.36	4.34	4.37
			0.5	4.20	4.18	**3.66**	*3.72*	3.90	3.90	4.19	4.16	4.20
			0.95	2.65	2.58	2.01	**2.01**	*2.01*	2.01	2.58	2.52	2.65
$(30, 50, 50, 50)$	2	0.5	0.25	7.02	7.01	*3.95*	**3.72**	4.49	4.57	7.03	7.03	7.02
			0.5	6.96	6.92	**3.64**	**3.64**	*4.45*	4.57	6.96	6.96	6.96
			0.95	7.65	7.64	*3.40*	**3.26**	3.69	4.24	7.66	7.67	7.65
		1	0.25	5.15	5.15	**3.47**	*3.54*	3.65	3.95	5.16	5.16	5.15
			0.5	5.30	5.27	**3.42**	*3.49*	3.66	3.80	5.30	5.30	5.30
			0.95	4.26	4.19	**2.98**	*2.98*	3.26	3.33	4.25	4.24	4.26
	5	0.5	0.25	4.55	4.53	**3.93**	*3.95*	4.08	4.70	4.54	4.53	4.55
			0.5	4.50	4.47	**3.86**	*3.87*	4.20	4.56	4.47	4.46	4.50
			0.95	3.36	3.33	*3.15*	**3.06**	3.16	3.16	3.34	3.33	3.36
		1	0.25	4.26	4.24	**3.81**	*3.84*	3.98	4.00	4.25	4.23	4.26
			0.5	4.22	4.21	**3.65**	*3.66*	3.84	3.87	4.21	4.20	4.22
			0.95	2.73	2.70	*2.24*	**2.23**	2.40	2.56	2.71	2.69	2.73

Table 6; PEs based on each criterion ($k^* = 4$; 3/4)

				$\mathrm{AIC}(\hat{G}(k)\|\alpha)$								
				the values of α								
(n_1, n_2, n_3, n_4)	p	δ	ρ	0	0.1	0.5	1	1.5	2	BIC	CAIC	$\mathrm{AIC^C}$
$(50, 30, 30, 30)$	2	0.5	0.25	6.60	6.57	*3.85*	**3.47**	4.64	4.64	6.60	6.60	6.60
			0.5	7.28	7.25	*3.89*	**3.70**	**3.70**	4.63	7.28	7.27	7.28
			0.95	7.85	7.72	*3.25*	**3.06**	3.30	4.59	7.87	7.84	7.86
		1	0.25	5.23	5.22	**3.48**	*3.49*	3.56	3.64	5.24	5.24	5.24
			0.5	5.35	5.29	*3.46*	**3.43**	3.51	3.60	5.34	5.34	5.35
			0.95	4.55	4.45	3.45	**3.22**	*3.23*	3.50	4.51	4.49	4.55
	5	0.5	0.25	4.65	4.63	**4.00**	*4.03*	4.19	4.71	4.63	4.61	4.65
			0.5	4.56	4.54	**3.76**	**3.76**	*4.01*	4.55	4.54	4.52	4.56
			0.95	3.52	3.47	*2.77*	**2.76**	2.85	3.31	3.46	3.39	3.52
		1	0.25	4.42	4.39	**3.83**	*3.83*	3.95	3.96	4.39	4.36	4.41
			0.5	4.24	4.22	**3.65**	*3.65*	3.67	3.81	4.22	4.19	4.24
			0.95	2.69	2.63	*2.00*	**2.00**	2.00	2.02	2.63	2.57	2.69
$(50, 30, 30, 50)$	2	0.5	0.25	7.06	7.05	**3.70**	*3.73*	4.58	4.58	7.06	7.06	7.06
			0.5	6.84	6.83	**3.37**	*3.37*	4.56	4.56	6.85	6.84	6.84
			0.95	7.73	7.63	*3.93*	**3.69**	4.52	4.52	7.70	7.70	7.73
		1	0.25	5.65	5.63	*3.59*	**3.52**	3.61	3.71	5.66	5.66	5.65
			0.5	5.85	5.81	3.65	**3.49**	*3.57*	3.66	5.86	5.85	5.85
			0.95	3.66	3.59	**2.92**	*3.07*	3.28	3.43	3.63	3.63	3.66
	5	0.5	0.25	4.67	4.65	**3.98**	*3.99*	4.02	4.69	4.65	4.63	4.67
			0.5	4.44	4.42	**3.82**	*3.82*	4.55	4.55	4.42	4.40	4.44
			0.95	3.38	3.31	**3.02**	*3.03*	3.15	3.86	3.31	3.29	3.37
		1	0.25	4.32	4.31	**3.80**	*3.81*	3.92	3.96	4.31	4.29	4.32
			0.5	4.14	4.13	**3.62**	*3.62*	3.64	3.81	4.13	4.11	4.14
			0.95	2.67	2.63	*2.01*	**2.01**	**2.01**	2.01	2.64	2.60	2.67
$(50, 30, 50, 30)$	2	0.5	0.25	7.50	7.46	**3.66**	*3.66*	4.65	4.65	7.50	7.50	7.50
			0.5	7.00	6.97	**3.45**	*3.46*	4.63	4.63	7.00	7.00	7.00
			0.95	6.81	6.70	2.98	**2.86**	*2.90*	4.60	6.77	6.76	6.81
		1	0.25	5.00	4.98	**3.36**	*3.48*	3.51	3.64	5.00	5.00	5.00
			0.5	5.43	5.38	*3.49*	**3.46**	3.51	3.60	5.43	5.42	5.43
			0.95	4.38	4.26	3.22	**3.17**	*3.19*	3.32	4.34	4.32	4.38
	5	0.5	0.25	4.48	4.45	**3.96**	*3.97*	4.37	4.71	4.46	4.43	4.48
			0.5	4.76	4.73	**3.80**	*3.80*	4.33	4.56	4.74	4.72	4.76
			0.95	3.00	2.97	*2.82*	**2.77**	2.89	3.42	2.97	2.96	3.00
		1	0.25	4.39	4.37	**3.82**	*3.85*	4.00	4.05	4.37	4.36	4.39
			0.5	4.22	4.21	**3.65**	*3.67*	3.75	3.87	4.21	4.20	4.22
			0.95	2.71	2.66	**2.15**	*2.17*	2.18	2.24	2.67	2.62	2.71
$(50, 30, 50, 50)$	2	0.5	0.25	6.71	6.70	*3.69*	**3.49**	4.56	4.56	6.71	6.71	6.71
			0.5	7.63	7.61	*4.01*	**3.94**	4.55	4.55	7.63	7.63	7.63
			0.95	6.15	6.06	**3.15**	*3.18*	*3.18*	4.52	6.13	6.12	6.15
		1	0.25	5.16	5.14	**3.47**	*3.53*	3.59	3.69	5.17	5.16	5.16
			0.5	5.46	5.40	*3.52*	**3.51**	3.53	3.67	5.45	5.45	5.46
			0.95	4.18	4.12	**2.73**	*2.84*	2.85	3.30	4.17	4.17	4.18
	5	0.5	0.25	4.62	4.59	**4.01**	*4.04*	4.41	4.69	4.60	4.58	4.61
			0.5	4.54	4.52	**3.78**	*3.80*	4.16	4.56	4.52	4.51	4.54
			0.95	3.01	2.98	**2.87**	*2.88*	3.00	3.87	2.99	2.98	3.01
		1	0.25	4.39	4.38	**3.82**	*3.84*	3.98	4.04	4.38	4.36	4.39
			0.5	4.12	4.11	**3.63**	*3.65*	3.73	3.89	4.11	4.10	4.12
			0.95	2.59	2.56	*2.01*	**2.01**	2.02	2.05	2.57	2.54	2.59

Table 7; PEs based on each criterion ($k^* = 4$; 4/4)

(n_1, n_2, n_3, n_4)	p	δ	ρ	AIC($\hat{G}(k)\|\alpha$) the values of α: 0	0.1	0.5	1	1.5	2	BIC	CAIC	AIC$^{\mathrm{C}}$
$(50, 50, 30, 30)$	2	0.5	0.25	6.82	6.78	**3.54**	**3.54**	*4.68*	*4.68*	6.82	6.82	6.82
			0.5	7.13	7.09	**3.43**	**3.43**	*4.67*	*4.67*	7.13	7.12	7.13
			0.95	7.29	7.13	*3.60*	**2.99**	**2.99**	4.63	7.24	7.23	7.28
		1	0.25	4.95	4.92	**3.41**	*3.45*	3.46	3.46	4.95	4.95	4.95
			0.5	5.47	5.42	**3.39**	*3.40*	3.44	3.59	5.46	5.46	5.47
			0.95	3.95	3.84	**2.81**	*2.90*	3.09	3.19	3.92	3.91	3.94
	5	0.5	0.25	4.56	4.55	**3.90**	*3.94*	4.17	4.72	4.55	4.53	4.56
			0.5	4.52	4.49	*3.94*	**3.94**	4.04	4.55	4.49	4.47	4.52
			0.95	3.51	3.47	3.04	**2.79**	*3.02*	3.05	3.47	3.44	3.51
		1	0.25	4.35	4.34	**3.82**	*3.85*	3.99	3.99	4.34	4.32	4.35
			0.5	4.34	4.32	**3.66**	*3.67*	3.72	3.86	4.32	4.30	4.34
			0.95	2.78	2.75	*2.49*	**2.47**	2.62	2.81	2.76	2.74	2.78
$(50, 50, 30, 50)$	2	0.5	0.25	6.66	6.67	**3.83**	*3.84*	4.63	4.63	6.67	6.67	6.66
			0.5	7.25	7.21	**3.55**	*3.55*	3.63	4.62	7.26	7.26	7.25
			0.95	8.13	7.99	3.36	**2.96**	*2.98*	4.58	8.11	8.09	8.13
		1	0.25	5.72	5.68	**3.47**	*3.50*	3.57	3.62	5.72	5.72	5.72
			0.5	5.67	5.64	3.46	**3.43**	*3.45*	3.61	5.67	5.67	5.67
			0.95	4.16	4.08	3.23	**3.12**	*3.17*	3.33	4.14	4.13	4.16
	5	0.5	0.25	4.51	4.50	**3.85**	*3.87*	3.93	4.70	4.51	4.50	4.51
			0.5	4.44	4.42	**3.71**	*3.73*	3.75	4.55	4.42	4.41	4.44
			0.95	3.48	3.44	3.14	**2.94**	*3.09*	3.33	3.45	3.43	3.48
		1	0.25	4.28	4.26	**3.81**	*3.82*	3.90	3.99	4.27	4.25	4.28
			0.5	4.21	4.20	**3.63**	*3.63*	3.64	3.79	4.20	4.19	4.21
			0.95	2.71	2.67	2.00	**2.00**	*2.00*	2.00	2.68	2.65	2.71
$(50, 50, 50, 30)$	2	0.5	0.25	7.08	7.03	*3.75*	**3.64**	4.70	4.70	7.07	7.07	7.08
			0.5	6.42	6.37	**3.52**	**3.52**	*4.68*	*4.68*	6.41	6.41	6.42
			0.95	7.96	7.96	3.30	**3.13**	*3.15*	4.66	7.98	7.98	7.96
		1	0.25	5.53	5.50	*3.53*	**3.49**	3.62	3.74	5.53	5.53	5.53
			0.5	5.67	5.65	**3.33**	*3.41*	3.44	3.62	5.67	5.67	5.67
			0.95	4.04	3.96	**3.08**	*3.14*	3.16	3.29	4.01	4.00	4.04
	5	0.5	0.25	4.53	4.52	**3.91**	*3.92*	4.05	4.71	4.53	4.52	4.53
			0.5	4.61	4.59	**3.93**	*4.02*	*4.02*	4.58	4.59	4.58	4.61
			0.95	3.44	3.39	*3.06*	**2.92**	3.16	3.16	3.39	3.38	3.44
		1	0.25	4.30	4.29	**3.81**	*3.83*	3.95	4.00	4.30	4.29	4.30
			0.5	4.19	4.17	**3.63**	*3.66*	3.84	3.84	4.18	4.16	4.19
			0.95	2.69	2.65	2.00	**2.00**	*2.00*	2.00	2.66	2.63	2.69
$(50, 50, 50, 50)$	2	0.5	0.25	6.52	6.51	**3.48**	*3.48*	4.63	4.63	6.52	6.53	6.52
			0.5	7.23	7.24	*3.77*	**3.50**	4.63	4.63	7.24	7.24	7.23
			0.95	6.84	6.69	*3.51*	**3.20**	**3.20**	4.59	6.82	6.81	6.84
		1	0.25	5.20	5.19	**3.50**	*3.57*	3.67	4.00	5.20	5.21	5.20
			0.5	5.45	5.41	*3.53*	**3.47**	3.56	3.72	5.44	5.44	5.45
			0.95	4.13	4.05	**3.03**	*3.07*	3.22	3.32	4.11	4.10	4.13
	5	0.5	0.25	4.48	4.48	**3.91**	*3.94*	4.17	4.72	4.48	4.48	4.48
			0.5	4.46	4.44	**3.79**	*3.82*	3.83	4.58	4.45	4.43	4.46
			0.95	3.55	3.54	3.34	**3.08**	*3.24*	3.89	3.54	3.53	3.55
		1	0.25	4.39	4.38	**3.84**	*3.85*	3.96	4.02	4.38	4.38	4.39
			0.5	4.11	4.09	**3.60**	*3.61*	3.67	3.81	4.10	4.09	4.11
			0.95	2.57	2.53	**1.99**	**1.99**	*2.00*	2.03	2.55	2.52	2.57

Table 8; Probability (%) of selecting correct number of clusters with each criterion ($k^* = 2$)

				AIC($\hat{G}(k)\|\alpha$)								
				the values of α								
(n_1, n_2)	p	δ	ρ	0	0.1	0.5	1	1.5	2	BIC	CAIC	AIC$^{\mathrm{C}}$
$(30, 30)$	2	1	0.25	0.00	0.00	*31.57*	**100.0**	**100.0**	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	*0.48*	**100.0**	**100.0**	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	*53.15*	**100.0**	**100.0**	0.00	0.00	0.00	0.00
		2	0.25	*0.00*	*0.00*	**100.0**	**100.0**	**100.0**	**100.0**	*0.00*	*0.00*	*0.00*
			0.5	0.00	0.00	*60.22*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.95	0.00	0.00	*17.79*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
	5	1	0.25	0.00	0.00	*46.98*	**100.0**	**100.0**	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	*16.52*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.95	*0.00*	*0.00*	**91.16**	**91.16**	**91.16**	**91.16**	*0.00*	*0.00*	*0.00*
		2	0.25	0.00	0.00	*5.92*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.5	0.00	0.00	*87.92*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.95	*0.00*	*0.00*	**100.0**	**100.0**	**100.0**	**100.0**	*0.00*	*0.00*	*0.00*
$(30, 50)$	2	1	0.25	0.00	0.00	*20.19*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.5	0.00	0.00	*17.16*	**100.0**	**100.0**	0.00	0.00	0.00	0.00
			0.95	*0.00*	*0.00*	*0.00*	**100.0**	**100.0**	*0.00*	*0.00*	*0.00*	*0.00*
		2	0.25	0.00	0.00	*21.17*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.5	*0.00*	*0.00*	**100.0**	**100.0**	**100.0**	**100.0**	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	*32.46*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
	5	1	0.25	0.00	0.00	*0.94*	**100.0**	**100.0**	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	*6.75*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.95	0.00	0.00	*27.21*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
		2	0.25	0.00	0.00	*41.66*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.5	0.00	0.00	*30.43*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.95	*0.00*	*0.00*	**100.0**	**100.0**	**100.0**	**100.0**	*0.00*	*0.00*	*0.00*
$(50, 30)$	2	1	0.25	*0.00*	*0.00*	*0.00*	**100.0**	**100.0**	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	*0.00*	**100.0**	**100.0**	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	*0.00*	*0.00*	**100.0**	**100.0**	**100.0**	*0.00*	*0.00*	*0.00*	*0.00*
		2	0.25	0.00	0.00	*23.24*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.5	0.00	0.00	*32.67*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.95	0.00	0.00	*64.39*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
	5	1	0.25	0.00	0.00	*31.02*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.5	0.00	0.00	*35.57*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.95	0.00	0.00	*5.65*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
		2	0.25	0.00	0.00	*88.93*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.5	0.00	0.00	*20.85*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.95	0.00	0.00	*31.71*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
$(50, 50)$	2	1	0.25	0.00	0.00	*29.76*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.5	0.00	0.00	*53.25*	**100.0**	**100.0**	0.00	0.00	0.00	0.00
			0.95	*0.00*	*0.00*	*0.00*	**100.0**	**100.0**	**100.0**	*0.00*	*0.00*	*0.00*
		2	0.25	*0.00*	*0.00*	**100.0**	**100.0**	**100.0**	**100.0**	*0.00*	*0.00*	*0.00*
			0.5	0.00	0.00	*31.06*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.95	*0.00*	*0.00*	**100.0**	**100.0**	**100.0**	**100.0**	*0.00*	*0.00*	*0.00*
	5	1	0.25	0.00	0.00	*30.46*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.5	0.00	0.00	*53.16*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.95	0.00	0.00	*24.63*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
		2	0.25	0.00	0.00	*97.21*	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.5	*0.00*	*0.00*	**100.0**	**100.0**	**100.0**	**100.0**	*0.00*	*0.00*	*0.00*
			0.95	*0.00*	*0.00*	**100.0**	**100.0**	**100.0**	**100.0**	*0.00*	*0.00*	*0.00*

Table 9; Probability (%) of selecting correct number of clusters with each criterion ($k^* = 4$; 1/4)

				$\mathrm{AIC}(\hat{G}(k)\|\alpha)$								
				the values of α								
(n_1, n_2, n_3, n_4)	p	δ	ρ	0	0.1	0.5	1	1.5	2	BIC	CAIC	$\mathrm{AIC^C}$
$(30, 30, 30, 30)$	2	1	0.25	*0.00*	*0.00*	**99.96**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**84.47**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	**61.07**	*49.12*	0.00	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	**57.00**	*17.44*	8.57	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**61.31**	*13.40*	7.17	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	**63.70**	*8.38*	8.35	4.45	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**91.24**	**91.24**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	0.00	0.00	*84.47*	**85.73**	0.00	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	0.99	*25.27*	**53.80**	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	56.69	*56.83*	**57.32**	56.69	0.00	0.27	0.00
			0.5	0.00	0.00	**60.79**	**60.79**	**60.79**	**60.79**	0.00	*0.83*	0.00
			0.95	0.00	0.00	*61.35*	*61.35*	*61.35*	**61.39**	0.11	5.60	0.00
$(30, 30, 30, 50)$	2	1	0.25	*0.00*	*0.00*	**0.26**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	0.00	0.00	**88.59**	*0.05*	0.00	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	**88.54**	*79.27*	0.00	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	**67.60**	*10.02*	7.75	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**66.12**	*9.36*	4.99	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	*39.25*	**65.80**	12.15	6.47	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**88.18**	**88.18**	**88.18**	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	0.00	0.00	*43.66*	**79.67**	0.00	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	0.00	*8.91*	**55.07**	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	61.33	*61.35*	**61.43**	61.33	0.00	0.00	0.00
			0.5	0.00	0.00	**78.97**	**78.97**	**78.97**	**78.97**	0.00	*1.13*	0.00
			0.95	0.00	0.22	*55.25*	*55.25*	**55.28**	*55.25*	0.26	3.01	0.00
$(30, 30, 50, 30)$	2	1	0.25	*0.00*	*0.00*	**89.60**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**95.83**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	*0.00*	*0.00*	**78.61**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
		2	0.25	0.00	0.00	**69.74**	*8.08*	4.92	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**68.55**	*11.06*	8.67	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	*47.36*	**70.20**	5.36	0.00	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**64.02**	**64.02**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**51.79**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	2.60	*23.79*	**29.18**	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	48.46	*48.65*	**49.47**	48.46	0.00	0.07	0.00
			0.5	0.00	0.00	**46.49**	**46.49**	**46.49**	**46.49**	0.00	*0.03*	0.00
			0.95	0.00	0.04	49.15	49.15	*49.17*	**49.43**	0.05	0.47	0.00
$(30, 30, 50, 50)$	2	1	0.25	*0.00*	*0.00*	**97.42**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**93.58**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	*68.09*	**94.28**	0.00	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	**74.22**	*17.91*	14.60	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**74.49**	*11.15*	9.17	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	*58.54*	**78.39**	7.56	0.00	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**70.90**	**70.90**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**80.64**	**80.64**	**80.64**	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	0.45	*53.73*	**88.26**	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	56.63	*56.75*	**57.48**	56.63	0.00	0.00	0.00
			0.5	*0.00*	*0.00*	**60.55**	**60.55**	**60.55**	**60.55**	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.04	56.74	56.74	*56.75*	**56.80**	0.03	0.43	0.00

Table 10; Probability (%) of selecting correct number of clusters with each criterion ($k^* = 4$; 2/4)

				AIC($\hat{G}(k)\|\alpha$)								
				the values of α								
(n_1, n_2, n_3, n_4)	p	δ	ρ	0	0.1	0.5	1	1.5	2	BIC	CAIC	AIC$^{\mathrm{C}}$
$(30, 50, 30, 30)$	2	1	0.25	*0.00*	*0.00*	**0.03**	**0.03**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
			0.95	0.00	0.00	*0.02*	**0.03**	0.00	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	**54.06**	*2.95*	1.53	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**48.67**	*6.52*	3.86	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	*42.45*	**51.27**	13.74	4.83	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**67.82**	**67.82**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**67.39**	**67.39**	**67.39**	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	*0.00*	*0.00*	*0.00*	**0.01**	**0.01**	*0.00*	*0.00*	*0.00*	*0.00*
		2	0.25	*0.00*	*0.00*	**46.47**	**46.47**	**46.47**	**46.47**	*0.00*	*0.00*	*0.00*
			0.5	0.00	0.00	**51.42**	**51.42**	**51.42**	**51.42**	0.00	*0.07*	0.00
			0.95	0.00	0.02	**61.87**	**61.87**	**61.87**	**61.87**	0.02	*1.42*	0.00
$(30, 50, 30, 50)$	2	1	0.25	*0.00*	*0.00*	**31.17**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	0.00	0.00	**74.17**	*0.10*	0.00	0.00	0.00	0.00	0.00
			0.95	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		2	0.25	0.00	0.00	**55.81**	*14.40*	7.92	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**56.75**	*13.33*	8.41	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	*36.63*	**56.30**	11.62	4.33	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**74.39**	**74.39**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**68.15**	**68.15**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	15.02	66.54	**75.85**	*69.50*	0.00	0.00	0.00
		2	0.25	*0.00*	*0.00*	**56.43**	**56.43**	**56.43**	**56.43**	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**61.04**	**61.04**	**61.04**	**61.04**	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.01	**56.68**	**56.68**	**56.68**	**56.68**	0.01	*0.27*	0.00
$(30, 50, 50, 30)$	2	1	0.25	*0.00*	*0.00*	**26.85**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**34.89**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	**69.13**	*33.79*	0.00	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	**61.17**	*13.17*	6.12	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**60.49**	*13.54*	8.74	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	*19.96*	**59.94**	11.77	4.40	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**76.44**	**76.44**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	0.00	0.00	*63.13*	**71.58**	6.39	0.00	0.00	0.00	0.00
			0.95	*0.00*	*0.00*	**0.01**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
		2	0.25	0.00	0.00	43.33	*43.34*	**43.36**	43.33	0.00	0.00	0.00
			0.5	*0.00*	*0.00*	**41.18**	**41.18**	**41.18**	**41.18**	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.01	**74.55**	**74.55**	**74.55**	**74.55**	0.01	*3.16*	0.00
$(30, 50, 50, 50)$	2	1	0.25	0.00	0.00	**84.47**	*0.01*	0.00	0.00	0.00	0.00	0.00
			0.5	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
			0.95	0.00	0.00	**75.77**	*36.80*	0.00	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	**65.86**	*14.06*	5.01	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**64.22**	*18.43*	4.83	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	*60.93*	**64.60**	8.45	4.03	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**78.49**	**78.49**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**81.90**	**81.90**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	*8.36*	**72.34**	0.00	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	53.73	*53.81*	**54.29**	53.73	0.00	0.00	0.00
			0.5	0.00	0.00	51.36	*51.38*	**52.36**	51.36	0.00	0.00	0.00
			0.95	0.00	0.00	0.33	*21.33*	**42.21**	**42.21**	0.00	0.00	0.00

Table 11; Probability (%) of selecting correct number of clusters with each criterion ($k^* = 4$; 3/4)

				AIC($\hat{G}(k)\|\alpha$)								
				the values of α								
(n_1, n_2, n_3, n_4)	p	δ	ρ	0	0.1	0.5	1	1.5	2	BIC	CAIC	AIC$^{\mathrm{C}}$
$(50, 30, 30, 30)$	2	1	0.25	*0.00*	*0.00*	**88.13**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**56.26**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	*0.44*	**19.50**	0.00	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	**50.31**	*11.51*	5.73	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**49.83**	*12.43*	6.60	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	**20.55**	17.38	*18.00*	0.00	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**76.46**	**76.46**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**100.0**	**100.0**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	0.18	**44.02**	*42.92*	0.00	0.00	0.00	0.00
		2	0.25	*0.00*	*0.00*	**65.27**	**65.27**	**65.27**	**65.27**	*0.00*	*0.00*	*0.00*
			0.5	0.00	0.00	**65.39**	**65.39**	**65.39**	**65.39**	0.00	*0.02*	0.00
			0.95	0.00	0.00	**68.15**	**68.15**	**68.15**	**68.15**	0.01	*1.08*	0.00
$(50, 30, 30, 50)$	2	1	0.25	*0.00*	*0.00*	**7.02**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**0.02**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	**31.11**	*0.08*	0.00	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	**55.09**	*16.52*	7.99	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**53.33**	*13.99*	7.93	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	*33.28*	**58.03**	10.39	6.48	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**93.35**	**93.35**	**93.35**	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**97.28**	**97.28**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	0.04	**61.95**	*59.06*	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	*66.98*	*66.98*	**67.01**	*66.98*	0.00	0.00	0.00
			0.5	0.00	0.00	**66.39**	**66.39**	**66.39**	**66.39**	0.00	*0.01*	0.00
			0.95	0.00	0.01	**72.63**	**72.63**	**72.63**	**72.63**	0.01	*0.31*	0.00
$(50, 30, 50, 30)$	2	1	0.25	*0.00*	*0.00*	**0.02**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**1.62**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	**62.06**	*22.53*	0.00	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	**87.92**	*11.63*	9.34	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**58.21**	*10.54*	6.57	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	*56.47*	**58.83**	12.05	7.23	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**96.78**	**96.78**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**99.65**	**99.65**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	0.01	*16.65*	**38.48**	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	52.31	*52.37*	**52.65**	52.31	0.00	0.00	0.00
			0.5	0.00	0.00	*56.93*	**56.95**	*56.93*	*56.93*	0.00	0.00	0.00
			0.95	0.00	0.00	3.86	43.46	*43.47*	**66.79**	0.00	0.00	0.00
$(50, 30, 50, 50)$	2	1	0.25	*0.00*	*0.00*	**73.97**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**99.16**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	*0.00*	*0.00*	**72.58**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
		2	0.25	0.00	0.00	**63.71**	*12.52*	7.19	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**75.24**	*6.21*	5.66	0.00	0.00	0.00	0.00
			0.95	0.00	0.01	67.88	*68.12*	**69.10**	7.82	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**76.21**	**76.21**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	0.00	0.00	*90.67*	**93.40**	0.36	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	0.17	*65.45*	**98.27**	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	55.16	*55.24*	**55.73**	55.16	0.00	0.00	0.00
			0.5	0.00	0.00	53.64	*53.75*	**54.32**	53.64	0.00	0.00	0.00
			0.95	0.17	0.38	**55.07**	**55.07**	**55.07**	**55.07**	0.27	*0.57*	0.17

Table 12; Probability (%) of selecting correct number of clusters with each criterion ($k^* = 4$; 4/4)

				$\mathrm{AIC}(\hat{G}(k)\|\alpha)$								
				the values of α								
(n_1, n_2, n_3, n_4)	p	δ	ρ	0	0.1	0.5	1	1.5	2	BIC	CAIC	$\mathrm{AIC^C}$
$(50, 50, 30, 30)$	2	1	0.25	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
			0.5	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
			0.95	*0.00*	*0.00*	**0.01**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
		2	0.25	0.00	0.00	**43.52**	*0.49*	0.24	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**43.60**	*13.04*	10.35	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	*38.46*	**43.29**	9.04	3.56	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**77.02**	**77.02**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	0.00	0.00	*51.53*	**95.61**	0.00	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	*0.30*	**40.78**	0.00	0.00	0.00	0.00	0.00
		2	0.25	*0.00*	*0.00*	**55.30**	**55.30**	**55.30**	**55.30**	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**56.90**	**56.90**	**56.90**	**56.90**	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	0.27	10.85	*44.87*	**59.55**	0.00	0.00	0.00
$(50, 50, 30, 50)$	2	1	0.25	*0.00*	*0.00*	**0.80**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	0.00	0.00	**1.11**	*0.48*	0.00	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	**26.06**	*0.96*	0.00	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	**49.89**	*9.56*	3.86	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**50.60**	*14.03*	12.58	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	**17.78**	*12.63*	12.58	4.87	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**90.25**	**90.25**	**90.25**	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**82.98**	**82.98**	**82.98**	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	0.17	**46.62**	*24.34*	0.00	0.00	0.00	0.00
		2	0.25	*0.00*	*0.00*	**62.53**	**62.53**	**62.53**	**62.53**	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**64.96**	**64.96**	**64.96**	**64.96**	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.01	**72.92**	**72.92**	**72.92**	**72.92**	0.00	*0.04*	0.00
$(50, 50, 50, 30)$	2	1	0.25	*0.00*	*0.00*	**87.17**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**0.05**	**0.05**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	**40.12**	*2.91*	1.15	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	**52.14**	*19.11*	8.69	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**53.16**	*15.10*	12.46	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	**50.45**	9.96	*10.53*	5.02	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**83.33**	**83.33**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**80.86**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
		2	0.25	*0.00*	*0.00*	**52.36**	**52.36**	**52.36**	**52.36**	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**53.18**	**53.18**	**53.18**	**53.18**	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	**51.43**	**51.43**	**51.43**	**51.43**	0.00	*0.02*	0.00
$(50, 50, 50, 50)$	2	1	0.25	*0.00*	*0.00*	**0.02**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	0.00	0.00	**83.76**	*0.51*	0.00	0.00	0.00	0.00	0.00
			0.95	*0.00*	*0.00*	**59.13**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
		2	0.25	0.00	0.00	**58.41**	*15.92*	7.98	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**56.66**	*12.56*	6.25	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	*59.79*	**60.57**	8.52	4.21	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**87.99**	**87.99**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	0.00	0.00	*68.67*	**82.74**	**82.74**	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	0.01	*35.82*	**36.55**	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	55.82	*55.83*	**56.29**	55.82	0.00	0.00	0.00
			0.5	*0.00*	*0.00*	**62.45**	**62.45**	**62.45**	**62.45**	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	**60.07**	**60.07**	**60.07**	**60.07**	0.00	*0.01*	0.00

Table 13; The number of times each number of clusters was selected by each criterion for the 'iris' data

	AIC($\hat{G}(k)\|\alpha$)						BIC	CAIC	AIC$^{\mathrm{C}}$
	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 2$			
1	0	0	0	0	0	0	0	0	0
2	0	0	0	1648	10000	10000	0	0	0
3 ($= k^*$)	0	0	*1042*	**7987**	0	0	0	0	0
4	973	1516	3858	365	0	0	1091	1577	973
5	9027	8484	5100	0	0	0	8909	8423	9027