Information criterion-based non-hierarchical clustering

Isamu Nagai*

Department of International Liberal Studies,
School of International Liberal Studies,
Chukyo University,
Aichi, Nagoya, Japan
Email: inagai@lets.chukyo-u.ac.jp
*Corresponding author

Katsuyuki Takahashi

Department of Social System and Management,
Graduate School of Systems and Information Engineering,
University of Tsukuba,
Ibaraki, Tsukuba, Japan
Email: k0620872@sk.tsukuba.ac.jp

Hirokazu Yanagihara

Department of Mathematics,
Graduate School of Science,
Hiroshima University,
Hiroshima, Higashi-Hiroshima, Japan
Email: yanagi-hiro@hiroshima-u.ac.jp

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 and 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. Bozdogan (1986) and Manning et al. (2009, Section 16.4.1) used formal information criteria, e.g., Aakaike's information criterion (AIC) and so on, for selecting the number of clusters. In this paper, we verify that such formal information criteria work poorly for selecting the number of clusters by conducting numerical examinations. Hence, we extend a formal AIC by adding a new penalty term, and search for an additional penalty with an acceptable selection-performance through numerical experiments.

Keywords: Aakaike's information criterion; AIC; cluster analysis; information criterion; *k*-means procedure; multivariate linear regression model; non-hierarchical clustering.

Reference to this paper should be made as follows: Nagai, I., Takahashi, K. and Yanagihara, H. (2017) ‘Information criterion-based non-hierarchical clustering’, *Int. J. Knowledge Engineering and Soft Data Paradigms*, Vol. 6, No. 1, pp.1–43.

Biographical notes: Isamu Nagai received his PhD in Hiroshima University. He is an Associate Professor at the Department of International Liberal Studies, School of International Liberal Studies, Chukyo University.

Katsuyuki Takahashi graduated the Department of Social System and Management, Graduate School of Systems and Information Engineering, University of Tsukuba.

Hirokazu Yanagihara received his PhD in Hiroshima University. He is a Professor at the Department of Mathematics, Graduate School of Science, Hiroshima University.

1 Introduction

In practice, we often determine whether clusters exist before further analysing 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 non-hierarchical 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 et al., 2009; Everitt et al., 2011). Saito and Yadohisa (2006) points out that non-hierarchical clustering can deal with more types of data than can hierarchical clustering. Hence, in the present paper, we focus on non-hierarchical clustering. One popular method for non-hierarchical clustering is the k-means procedure that was proposed by MacQueen (1967). In non-hierarchical 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.

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. Bozdogan (1986) and Manning et al. (2009; Section 16.4.1) used formal information criteria, e.g., Aakaike’s information criterion (AIC; Akaike, 1973), Bayesian information criterion (BIC; Schwarz, 1978) and consistent AIC (CAIC; Bozdogan, 1986) for selecting the number of clusters. Unfortunately, such formal information criteria work poorly for selecting the number of clusters. We will verify such a fact through numerical experiments. This may be caused from an undervaluation of a penalty term of a formal information criterion. Hence, we extend a formal AIC by adding a new penalty term ‘$\alpha n(k - 1)$’, where k is the number of clusters and α is some nonnegative value. Since it is very difficult to obtain a theoretical best value of α, through numerical experiments, we search for a value of which maintains an acceptable selection-performance.

The remainder of the present paper is organised as follows: In Section 2, we briefly illustrate cluster analysis and several cluster criteria for the clustering of data. In Section 3, we show the relationship between cluster analysis and the multivariate linear regression model. Adding the new penalty term $\alpha n(k-1)$, we extend a formal AIC. In Section 4, we verify that formal information criteria work poorly for selecting the number of clusters and search for a value of α which maintains an acceptable selection-performance by conducting numerical experiments. In Appendix, we prove the useful and equivalent conditions for renewing clusters for each cluster criterion.

2 Cluster analysis

Let 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 $y_1, \ldots, y_n$ (see, e.g., Hastie et al., 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.
- Non-hierarchical clustering: the number of clusters is determined prior to beginning the cluster analysis. The data points $y_1, \ldots, y_n$ 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, non-hierarchical clustering can be applied to more types of data than can hierarchical clustering (see, e.g., Saito and Yadohisa, 2006). We note that the k-means procedure (MacQueen, 1967) is popular method for non-hierarchical clustering.

For non-hierarchical clustering, various cluster criteria have been proposed for renewing clusters (see, e.g., Marriott, 1982; Krzanowski and 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 j^{th} 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 j^{th} cluster. For example, consider a clustered set in which $k = 2$, y_1, y_2, and y_4 belong to the first cluster; and y_3 and 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. In addition, let $G(k) = \{C_1, \ldots, C_k\}$ and:

$$W_j = \sum_{i \in C_j} \left(y_i - \bar{y}_j^{(k)} \right) \left(y_i - \bar{y}_j^{(k)} \right)', \quad W(G(k)) = \sum_{j=1}^{k} W_j,$$

where $\bar{y}_j^{(k)}$ is the sample mean in the j^{th} cluster, i.e., $\bar{y}_j^{(k)} = \sum_{i \in C_j} y_i / n_j$. Using W_j and $W(G(k))$, Table A1 in Appendix lists the various cluster criteria that been proposed previously. We note that W_j ($j = 1, \ldots, k$) and $W(G(k))$ are symmetric matrices. In particular, we assume that $G(k) \in \mathcal{G}(k)$, where $\mathcal{G}(k)$ denotes a set of cluster partition such that det $W(G(k)) \neq 0$ for any $G(k)$ and k when we use (ii) or (v) in Table A1, Appendix. When we use other criteria, we assume $\det(W_j) \neq 0$ ($j = 1, \ldots, k$) instead of assuming $\det(W(G(k))) \neq 0$. If we assume this, we obtain $\det(W(G(k))) \neq 0$. This fact means that we obtain $\det(W(G(k))) > 0$ for any $G(k)$ and k [see, e.g., Lütkepohl, (1996), p.55].

Of the cluster criteria listed there, (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 $CC(W(G(k)))$. The algorithm for performing non-hierarchical clustering with any of the criteria $CC(W(G(k)))$ is as follows:

1 Choose k, and select a cluster criterion $CC(W(G(k)))$.

2 Divide $y_1, \ldots, y_n$ into k clusters, either randomly or by using a non-hierarchical clustering method, such as the k-means procedure.

3 Move y_r, which is in the s^{th} cluster, into the t^{th} cluster ($s \neq t$) if $CC(W(G'(k))) < CC(W(G(k)))$ 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 y_r is moved from the s^{th} cluster into the t^{th} 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 $CC(W(G(k)))$ converges to a minimum value. This produces the optimal clustering $\hat{G}(k)$. That is, the optimal clustering $\hat{G}(k)$ is derived from $\hat{G}(k) = \arg\min_{G(k)\in\mathcal{G}(k)} CC(W(G(k)))$ for the selected cluster criterion and the given k.

Marriott (1982) showed the variation that occurred when the r^{th} individual y_r is added to the j^{th} cluster. Using this result, we derive the equivalent conditions with $CC(W(G'(k))) < CC(W(G(k)))$ for each cluster criterion in Appendix. When we use this equivalent condition, we can easily check that the renewal condition is satisfied or not.

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.

In the cluster analysis under the provided k, we can assume that y_i will derived from the following model independently when $i \in C_j$ ($j = 1, \ldots, k$):

$$y_i = \xi_j^{(k)} + \varepsilon_i \, (i = 1, \ldots, n), \tag{1}$$

where $\xi_j^{(k)}$ $(j = 1, \ldots, k)$ is an unknown p-dimensional vector which means that the unknown centre vector of the true j^{th} cluster population, and $\varepsilon_1, \ldots, \varepsilon_n$ are independent p-dimensional error vectors with the mean 0_p and some covariance matrix, where 0_p is a p-dimensional zero vector. Let $e_j^{(r)}$ be an r-dimensional standard basis vector in which the j^{th} element is one and the other elements are zero. By using the standard basis vector as a vector of dummy variables, i.e., $x_i^{(k)} = e_j^{(k)}$ when $i \in C_j$ ($i = 1, \ldots, n, j = 1, \ldots, k$), the model (1) can be rewritten as:

$$y_i = \Xi^{(k)\prime} x_i^{(k)} + \varepsilon_i \; (i = 1, \ldots, n), \tag{2}$$

where $\Xi^{(k)} = (\xi_1^{(k)}, \ldots, \xi_k^{(k)})'$ is a $k \times p$ matrix. Here, if we assume that $\varepsilon_i \sim N_p(0_p, \Sigma)$, where Σ is an unknown $p \times p$ covariance matrix with $\det(\Sigma) \neq 0$, the model (2) can be expressed as:

$$Y \sim N_{n\times p}\left(X^{(k)}\Xi^{(k)}, \Sigma \otimes I_n\right), \tag{3}$$

where $Y = (y_1, \ldots, y_n)'$, $X^{(k)} = (x_1^{(k)}, \ldots, x_n^{(k)})'$, and $\otimes$ indicates the Kronecker product [for the definition of the Kronecker product, see, e.g., Muirhead, (1982), p.73]. In the cluster analysis, we can assume that all clusters have more than one individual. This means that all the basis vectors $e_1^{(k)}, \ldots, e_k^{(k)}$ should appear in $x_1^{(k)}, \ldots, x_n^{(k)}$ when k is provided. Hence, we naturally assume that $\mathrm{rank}(X^{(k)}) = k$. Needless to say, the matrix $X^{(k)}$ also expresses a cluster partition when the number of clusters is k as well as $G(k)$.

Here, we will consider a method for determining a cluster partition by maximising the log-likelihood function of (3). It is well known that the log-likelihood function of the model (3) is given by:

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

By maximising the above log-likelihood function with respect to $(\Xi^{(k)}, \Sigma)$, maximum likelihood estimators of $\Xi^{(k)}$ and Σ under the fixed $X^{(k)}$ are derived as:

$$\begin{aligned} \hat{\Xi}^{(k)}\left(X^{(k)}\right) &= \left(X^{(k)\prime}X^{(k)}\right)^{-1}X^{(k)\prime}Y, \\ \hat{\Sigma}^{(k)}\left(X^{(k)}\right) &= \frac{1}{n}Y'\left\{I_n - X^{(k)}\left(X^{(k)\prime}X^{(k)}\right)^{-1}X^{(k)\prime}\right\}Y. \end{aligned}$$

These imply the maximum log-likelihood under the fixed $X^{(k)}$ as:

$$\begin{aligned} & g\left(\hat{\Xi}^{(k)}\left(X^{(k)}\right), \hat{\Sigma}^{(k)}\left(X^{(k)}\right) \mid Y, X^{(k)}\right) \\ & = -\frac{n}{2}\left\{p\log(2\pi) + p + \log\left\{\det\left(\hat{\Sigma}^{(k)}\left(X^{(k)}\right)\right)\right\}\right\} \\ & = -\frac{np}{2}\{\log(2\pi) + 1 - \log(n)\} - \frac{n}{2}\log\left\{\det\left(n\hat{\Sigma}^{(k)}\left(X^{(k)}\right)\right)\right\} = \mathcal{L}\left(X^{(k)}\right). \end{aligned} \tag{4}$$

Then, we have:

$$g\left(\Xi^{(k)}, \Sigma \mid Y, X^{(k)}\right) \leq \mathcal{L}\left(X^{(k)}\right).$$

This indicates that the cluster partition maximising the log-likelihood function of the model (3) is $X^{(k)}$ maximising $\mathcal{L}(X^{(k)})$. Hence, the optimal cluster partition derived from the maximisation of the log-likelihood function of the model (3) is given by

$$\hat{X}^{(k)} = \arg \max_{X^{(k)} \in \mathrm{M}(n,k)} \mathcal{L}\left(X^{(k)}\right), \tag{5}$$

where M(n, k) is the set of all full-rank $n \times k$ matrices of which the row vectors are one of $\{e_1^{(k)}, \ldots, e_k^{(k)}\}$. Note that $-np\{\log(2\pi) + 1 - \log(n)\} / 2$ in (4) is independent of $X^{(k)}$ and $\log(x)$ is a strictly monotonically increasing function with respect to x. Thus, the equation in (5) can be rewritten as:

$$\hat{X}^{(k)} = \arg \min_{X^{(k)} \in \mathrm{M}(n,k)} \det\left\{n\hat{\Sigma}\left(X^{(k)}\right)\right\}.$$

Note that the number of individuals belonging to the j^{th} cluster n_j is presented by:

$$n_j = \sum_{i=1}^{n} I\left(x_i^{(k)} = e_j^{(k)}\right),$$

where $I(A)$ is an indicator function i.e., $I(A) = 1$ if A is true and $I(A) = 0$ if A is not true. Then, we derive:

$$\begin{aligned}
\hat{\Xi}^{(k)}\left(X^{(k)}\right) &= \operatorname{diag}\left(n_1, \ldots, n_k\right)^{-1} \sum_{i=1}^{n} \begin{pmatrix} I\left(x_i^{(k)} = e_1^{(k)}\right) y_i' \\ \vdots \\ I\left(x_i^{(k)} = e_k^{(k)}\right) y_i' \end{pmatrix} \\
&= \operatorname{diag}\left(n_1, \ldots, n_k\right)^{-1} \sum_{i=1}^{n} \begin{pmatrix} I\left(i \in C_1\right) y_i' \\ \vdots \\ I\left(i \in C_k\right) y_i' \end{pmatrix} = \begin{pmatrix} \overline{y}_1^{(k)\prime} \\ \vdots \\ \overline{y}_k^{(k)\prime} \end{pmatrix}.
\end{aligned}$$

This indicates that:

$$n\hat{\Sigma}^{(k)}\left(X^{(k)}\right) = \sum_{j=1}^{k} \sum_{i \in C_j} \left(y_i - \overline{y}_j^{(k)}\right)\left(y_i - \overline{y}_j^{(k)}\right)' = W\left(G(k)\right).$$

This fact means that a method for determining a cluster partition by maximising the loglikelihood of the model (3) with respect to $X^{(k)}$ is equivalent to that by minimising $\det(W(G(k)))$ with respect to $G(k)$. From this fact, we can see that $n\hat{\Sigma}^{(k)}(\hat{X}^{(k)}) = W(\hat{G}(k))$.

3.2 Formal information criterion for selecting the number of clusters

In the previous subsection, we show that a method for determining a cluster partition $G(k)$ by minimising $\det(W(G(k)))$ is equivalent to that by maximising the log-likelihood function of the ordinary multivariate linear regression model (3). Hence, a selection method of the number of clusters based on the minimisation of an information criterion comes immediately to mind. By neglecting the constant term, an information criterion in the model (3) is given by:

$$IC\left(G(k)\right) = n \log\left\{\det\left(W\left(G(k)\right)\right)\right\} + m(k),$$

where $m(k)$ is a positive constant expressing a penalty for the complexity of the model (3) which depends on the number of clusters but is independent of the cluster partition. Since the optimal cluster partition $\hat{G}(k)$ is derived from the minimisation of det{$W(G(k))$}, the optimal number of clusters is obtained by minimising IC($G(k)$) as:

$$\hat{k} = \arg\min_{k \in \mathcal{K}} \mathrm{IC}\left(\hat{G}(k)\right) = \arg\min_{k \in \mathcal{K}} \min_{G(k) \in \mathcal{G}(k)} \mathrm{IC}\left(G(k)\right),$$

where $\mathcal{K}$ denotes the candidate number of clusters; for example, $\mathcal{K} = \{1, 2, 3, 5, 10, 15\}$. Hence, $\mathrm{IC}(\hat{G}(k))$ is regarded as a criterion for selecting the number of clusters.

If we regard the cluster partition as a hyper parameter, the number of independent parameters in the model (4) is $p\{k + (p + 1) / 2\}$. Then, famous four information criteria, AIC proposed by Akaike (1973), Bayesian information criterion (BIC) proposed by Schwarz (1978), a consistent AIC (CAIC) proposed by Bozdogan (1987), and a bias-corrected AIC (AIC$^{\mathrm{C}}$) proposed by Bedrick and Tsai (1994), may be used as IC($G(k)$). The formal AIC, BIC, CAIC and AIC$^{\mathrm{C}}$ of the model (4) are given as follows:

$$\mathrm{AIC}\left(G(k)\right) = n \log\left\{\det\left(W\left(G(k)\right)\right)\right\} + 2p\left(k + \frac{p+1}{2}\right), \tag{6}$$

$$\mathrm{BIC}\left(G(k)\right) = n \log\left\{\det\left(W\left(G(k)\right)\right)\right\} + p\log(n)\left(k + \frac{p+1}{2}\right), \tag{7}$$

$$\mathrm{CAIC}\left(G(k)\right) = \mathrm{BIC}\left(G(k)\right) + p\left(k + \frac{p+1}{2}\right), \tag{8}$$

$$\mathrm{AIC}^{\mathrm{C}}\left(G(k)\right) = \mathrm{AIC}\left(G(k)\right) + \frac{2p(k+p+1)}{n-k-p-1}\left(k + \frac{p+1}{2}\right). \tag{9}$$

For more details of the AIC, see, e.g., Konishi and Kitagawa (2008), and Rao et al. (2008). Needless to say, the above four criteria are frequently used for selecting explanatory variables in the multivariate linear regression model.

Recall that det($W(G(k))$) in each criterion is corresponding to the cluster criterion (ii) in Table A1. Hence, we should not use the formal information criteria $\mathrm{AIC}(\hat{G}(k))$, $\mathrm{BIC}(\hat{G}(k))$, $\mathrm{CAIC}(\hat{G}(k))$ and $\mathrm{AIC}^{\mathrm{C}}(\hat{G}(k))$ as a criterion for selecting the number of clusters if a cluster criterion other than (ii) is used for determining the cluster partition.

3.3 New AIC-type criterion for selecting the number of clusters

In the previous subsection, we illustrated the formal information criteria in (6), (7), (8) and (9) for selecting the number of clusters. Unfortunately, it is not expected that the formal information criteria work well for selecting the number of clusters. In general, a probability to cause the false partition in k-means clustering does not converge to 0 even when $k \geq k^*$, where k^* is the true number of clusters. This will imply that:

$$n \log\left\{\det\left(W\left(\hat{G}(k)\right)\right)\right\} - n \log\left\{\det\left(W\left(\hat{G}(k^*)\right)\right)\right\} = O_p(n) \ as\ n \to \infty,$$

for all $k \in \mathcal{K} \setminus \{k^*\}$. We illustrate the reason why the above equation is derived through a simple example, i.e., $k^* = 1$, $k = 2$, $p = 1$ and the true model is $N(0, 1)$. Since we consider the case of $p = 1$, i.e., an univariate case, $W(\hat{G}(k))$ becomes scalar. Hence we use $w(\hat{G}(k))$ instead of $W(\hat{G}(k))$. Needless to say, $\det(w(\hat{G}(k))) = w(\hat{G}(k))$. Let $y_1, \ldots, y_{n_1}$ be observations which belong to the first cluster and $y_{n_1+1}, \ldots, y_n$ be observations which belong to the second cluster where $n = n_1 + n_2$. Then, $w(\hat{G}(2)) / n$ is given by:

$$\frac{1}{n} w\left(\hat{G}(2)\right) = \frac{n_1}{n}\left\{\frac{1}{n_1}\sum_{i=1}^{n_1}\left(y_i - \overline{y}_1^{(2)}\right)^2\right\} + \frac{n_2}{n}\left\{\frac{1}{n_2}\sum_{i=1}^{n_2}\left(y_{n_1+i} - \overline{y}_2^{(2)}\right)^2\right\},$$

where

$$\overline{y}_1^{(2)} = \frac{1}{n_1}\sum_{i=1}^{n_1} y_i, \quad \overline{y}_2^{(2)} = \frac{1}{n_2}\sum_{i=1}^{n_2} y_{n_1+i}.$$

In this case, the distributions of $y_1, \ldots, y_{n_1}$ and $y_{n_1+1}, \ldots, y_n$ will converge to the half normal distributions and n_j / n ($j = 1, 2$) will converge to $1 / 2$. Notice that the variance of the half normal distribution is $1 - 2 / \pi$. It follows from this variance and the law of large numbers that:

$$\frac{1}{n} w\left(\hat{G}(2)\right) \xrightarrow{p} \frac{1}{2}\left(1 - \frac{2}{\pi}\right) + \frac{1}{2}\left(1 - \frac{2}{\pi}\right) = \left(1 - \frac{2}{\pi}\right) \; as \; n \to \infty.$$

Moreover, it is clear that $w(\hat{G}(1)) / n \xrightarrow{p} 1$. Hence, we derive:

$$\begin{aligned}&\frac{1}{n}\left\{n \log\left\{w\left(\hat{G}(2)\right)\right\} - n \log\left\{w\left(\hat{G}(1)\right)\right\}\right\}\\&= \log\left\{w\left(\hat{G}(2)/n\right)\right\} - \log\left\{w\left(\hat{G}(1)/n\right)\right\} \xrightarrow{p} \log\left(1 - \frac{2}{\pi}\right) \approx -1.0123.\end{aligned}$$

This indicates $n \log\{w(\hat{G}(2))\} - n \log\{w(\hat{G}(1))\} = O_p(n)$ as $n \to \infty$. Furthermore, we know that $n \log\{\det(W(\hat{G}(k+1)))$ is smaller than $n \log\{\det(W(\hat{G}(k)))$ at almost all cases. Therefore, the maximum value in $\mathcal{K}$ tends to be chosen as $\hat{k}$ by $\mathrm{AIC}(\hat{G}(k))$, $\mathrm{BIC}(\hat{G}(k))$, $\mathrm{CAIC}(\hat{G}(k))$ and $\mathrm{AIC^C}(\hat{G}(k))$, because the orders of these penalty terms are smaller than $O(n)$.

This defect will be improved by using an penalty term of which the order is $O(n)$. Here, 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 l^{th} cluster when all $k - 1$ parameters are zeros, only $k - 1$ parameters can be chosen; that is, we will regard that there are $k - 1$ independent parameters for each individual. This means that there are new $n(k - 1)$ independent parameters, which correspond to the location of the n individuals. In the present paper, we try using this term in order to be penalised on the number of clusters.

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

$$\mathrm{AIC}\left(G(k)\middle|\alpha\right) = \mathrm{AIC}\left(G(k)\right) + \alpha n(k-1). \tag{10}$$

Note that AIC($G(k)$ | 0) = AIC($G(k)$), and AIC($G(k)$ | 2) corresponds to the criterion when the weight for the new parameters coincides with that for the present parameters. In the numerical studies presented in the next section, we will chose 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.

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 D_p = diag(1, …, p) be a $p \times p$ diagonal matrix, and let $\Delta_p(\rho)$ be a $p \times p$ matrix in which the (i, j)th element is $\rho^{|i-j|}$. Then Y, which is the data, is generated from $N_{n\times p}(\Theta, \Sigma^* \otimes I_n)$ for each repetition, where $\Sigma^* = D_p^{1/2}\Delta_p(\rho)D_p^{1/2}$, $\Theta = 5\delta(\theta_1 1_{n_1}', \ldots, \theta_{k^*} 1_{n_{k^*}}')'$ is an $n \times p$ matrix, δ is a scale parameter, and 1_r is an r-dimensional vector in which each of the elements is one. Here, θ_j (j = 1, …, k^*) controls a p-dimensional vector at the true center of each true cluster. We set the θ_j as follows: $\theta_1 = 0_p$, $\theta_2' e_i^{(p)} = \sin(i)$, $\theta_3' e_i^{(p)} = \log((-1)^i + i + 1)$, and $\theta_4' e_i^{(p)} = (-1)^i + 0.5 (i = 1, \ldots, p)$. We note that δ controls the scale of Θ. 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 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 based on $\det(W(\hat{G}(k)))$. Note that $\det(W(\hat{G}(k)))$ is also in the first term of $\mathrm{AIC}(\hat{G}(k) \mid \alpha)$ in (10) and the formal other criteria that are in Section 3.2. Using the clustering results, $W(\hat{G}(k))$ was derived for each k. Then, $\mathrm{AIC}(\hat{G}(k) \mid \alpha)$ was also derived for each α. In order to compare with the formal criteria $\mathrm{BIC}(\hat{G}(k))$, $\mathrm{CAIC}(\hat{G}(k))$, and $\mathrm{AIC^C}(\hat{G}(k))$ in Section 3.2 for selecting the number of clusters, we select k by using these criteria and $\mathrm{AIC}(\hat{G}(k) \mid \alpha)$ for several α.

We select $\hat{k}$ by minimising 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:

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

where $\mathcal{M} = \Theta, \hat{\Sigma}^{(\hat{k})}(\hat{X}^{(\hat{k})})$ is obtained by $n^{-1}W(\hat{G}(\hat{k}))$, $\hat{Y} = \hat{X}^{(\hat{k})}(\hat{X}^{(\hat{k})\prime}\hat{X}^{(\hat{k})})^{-1}\hat{X}^{(\hat{k})\prime}Y$, and $\hat{X}^{(\hat{k})}$ is derived from the cluster results $\hat{G}(\hat{k})$ for each repetition. Here, the expectation of above PE is the risk function. The idea of formal criteria is selecting model which makes the risk function smaller. After 10,000 repetitions, the average PE values and the probability (%) of correctly selecting the number of clusters were used for comparing these criteria. This average values of PEs are regarded as the risk function's value.

We present the (rounded) results in Tables 1 to 14. In Tables 1 to 7, for each case, the minimum value is in bold, and next smallest value is in italics. In Tables 8 to 14, 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, $\mathrm{BIC}(\hat{G}(k))$, $\mathrm{CAIC}(\hat{G}(k))$, and $\mathrm{AIC^C}(\hat{G}(k))$ are written as BIC, CAIC, and $\mathrm{AIC^C}$, respectively.

We first consider the results based on the PEs that are listed in Tables 1 to 7. We will focus on the results for $k^* = 2$, as shown in Tables 1 and 2. As δ increases, the results for $\mathrm{AIC}(\hat{G}(k)\,|\,\alpha)$ (α = 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 $\mathrm{AIC}(\hat{G}(k)\,|\,\alpha)$ (α = 1, 1.5, 2) almost always decrease as ρ increases. We note that the results for $\mathrm{AIC}(\hat{G}(k))$, $\mathrm{AIC}(\hat{G}(k)\,|\,0.1)$, $\mathrm{BIC}(\hat{G}(k))$, $\mathrm{CAIC}(\hat{G}(k))$, and $\mathrm{AIC^C}(\hat{G}(k))$ 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 all results based on PEs in $k^* = 2$, we see that $\mathrm{AIC}(\hat{G}(k)\,|\,1)$ always yields the best results, followed by $\mathrm{AIC}(\hat{G}(k)\,|\,1.5)$. Furthermore, both $\mathrm{AIC}(\hat{G}(k)\,|\,2)$ and $\mathrm{AIC}(\hat{G}(k)\,|\,0.5)$ work well in almost every situation. On the other hand, $\mathrm{AIC}(\hat{G}(k))$ and other the criteria cannot be used directly for selecting the number of clusters based on the PE values with $k^* = 2$.

Next, we will focus on the PEs when $k^* = 4$, which are shown in Tables 3 to 7. The results of $\mathrm{AIC}(\hat{G}(k)\,|\,2)$ are small when ρ = 0.95. The results of $\mathrm{AIC}(\hat{G}(k)\,|\,1)$ and $\mathrm{AIC}(\hat{G}(k)\,|\,1.5)$ are also small when ρ = 0.95, except in the case (n_1, n_2, n_3, n_4, δ) = (30, 50, 30, 30, 1). As δ increases, the results for all criteria are similar as the case of $k^* = 2$ in many cases. The results for $\mathrm{AIC}(\hat{G}(k)\,|\,2)$ decrease somewhat as ρ increases, and the results for $\mathrm{AIC}(\hat{G}(k)\,|\,1.5)$ decrease in many cases. As similar as $k^* = 2$, several criteria have the same tendencies as ρ increases in almost always. We note that in all situations considered, either $\mathrm{AIC}(\hat{G}(k)\,|\,0.5)$ or $\mathrm{AIC}(\hat{G}(k)\,|\,1)$ gave the best results. As when $k^* = 2$, we see that $\mathrm{BIC}(\hat{G}(k))$, $\mathrm{CAIC}(\hat{G}(k))$, $\mathrm{AIC^C}(\hat{G}(k))$ and $\mathrm{AIC}(\hat{G}(k))$ cannot be used directly to select the number of clusters.

Table 1 PEs based on each criterion ($k^* = 2$; 1 / 2)

				AIC($\hat{G}(k) \mid \alpha$)								
				The values of α								
(n_1, n_2)	p	δ	ρ	*0*	*0.1*	*0.5*	*1*	*1.5*	*2*	*BIC*	*CAIC*	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

Table 1 PEs based on each criterion ($k^* = 2$; 1 / 2) (continued)

				AIC($\hat{G}(k) \mid \alpha$)								
				The values of α								
(n_1, n_2)	*p*	δ	ρ	*0*	*0.1*	*0.5*	*1*	*1.5*	*2*	*BIC*	*CAIC*	*AIC^C*
(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 2 PEs based on each criterion ($k^* = 2$; 2 / 2)

				AIC($\hat{G}(k) \mid \alpha$)								
				The values of α								
(n_1, n_2)	p	δ	ρ	*0*	*0.1*	*0.5*	*1*	*1.5*	*2*	*BIC*	*CAIC*	AIC^C
(100, 100)	2	1	0.25	7.96	7.92	4.22	**3.36**	**3.36**	4.11	7.96	7.95	7.96
			0.5	9.45	9.39	4.81	**3.26**	**3.26**	3.93	9.45	9.44	9.45
			0.95	15.67	15.52	4.09	**2.41**	**2.41**	3.00	15.67	15.67	15.67
		2	0.25	8.14	*8.07*	**3.17**	**3.17**	**3.17**	3.17	8.13	8.13	8.14
			0.5	9.34	*9.26*	**3.05**	**3.05**	**3.05**	3.05	9.33	9.33	9.34
			0.95	13.13	12.89	*2.58*	**2.03**	**2.03**	2.03	13.10	13.09	13.13
	5	1	0.25	5.26	5.24	*4.05*	**3.80**	**3.80**	3.80	5.25	5.23	5.26
			0.5	5.76	5.78	*4.04*	**3.62**	**3.62**	3.62	5.77	5.78	5.76
			0.95	6.50	6.41	**1.97**	**1.97**	**1.97**	1.97	6.45	6.39	6.50
		2	0.25	5.22	5.21	*3.85*	**3.78**	**3.78**	3.78	5.22	5.21	5.22
			0.5	5.33	5.32	**3.60**	**3.60**	**3.60**	3.60	5.33	5.31	5.33
			0.95	5.93	5.86	**1.96**	**1.96**	**1.96**	1.96	5.90	5.83	5.93
(500, 500)	2	1	0.25	7.94	7.93	4.64	**3.33**	**3.33**	4.10	7.94	7.94	7.94
			0.5	9.32	9.28	4.93	**3.26**	**3.26**	3.92	9.31	9.31	9.32
			0.95	16.60	16.53	4.16	**2.43**	**2.43**	3.00	16.60	16.60	16.60
		2	0.25	7.39	7.36	*3.64*	**3.15**	**3.15**	3.15	7.39	7.39	7.39
			0.5	8.11	8.02	*3.29*	**3.04**	**3.04**	3.04	8.11	8.11	8.11
			0.95	10.81	10.68	*2.27*	**2.02**	**2.02**	2.02	10.80	10.80	10.81
	5	1	0.25	4.99	4.99	*4.11*	**3.82**	**3.82**	3.82	4.99	4.99	4.99
			0.5	5.22	5.21	*3.98*	**3.65**	**3.65**	3.65	5.22	5.22	5.22
			0.95	5.84	5.77	*2.31*	**1.94**	**1.94**	1.94	5.83	5.83	5.84
		2	0.25	4.92	4.92	*3.95*	**3.75**	**3.75**	3.75	4.92	4.92	4.92
			0.5	5.14	5.13	*3.78*	**3.57**	**3.57**	3.57	5.14	5.14	5.14
			0.95	5.43	5.39	*2.05*	**1.94**	**1.94**	1.94	5.43	5.43	5.43

Table 3 PEs based on each criterion ($k^* = 4$; 1 / 5)

				AIC($\hat{G}(k)\mid\alpha$)								
				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*^{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

Table 3 PEs based on each criterion ($k^* = 4$; 1 / 5) (continued)

				$\mathrm{AIC}(\hat{G}(k)\,|\,\alpha)$								
				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*	AIC^C
(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 4 PEs based on each criterion ($k^* = 4$; 2 / 5)

				$\mathrm{AIC}(\hat{G}(k)\mid\alpha)$								
				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^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

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

				$\mathrm{AIC}(\hat{G}(k) \mid \alpha)$								
				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^C
(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 5 PEs based on each criterion ($k^* = 4$; 3 / 5)

(n_1, n_2, n_3, n_4)	p	δ	ρ	$\mathrm{AIC}(\hat{G}(k) \mid \alpha)$ The values of α: *0*	*0.1*	*0.5*	*1*	*1.5*	*2*	*BIC*	*CAIC*	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

Table 5 PEs based on each criterion ($k^* = 4$; 3 / 5) (continued)

				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*^{C}
(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 6 PEs based on each criterion ($k^* = 4$; 4 / 5)

				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^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

Table 6 PEs based on each criterion ($k^* = 4$; 4 / 5) (continued)

				$\mathrm{AIC}(\hat{G}(k)\,|\,\alpha)$								
				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*	AIC^C
(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 7 PEs based on each criterion ($k^* = 4$; 5 / 5)

				$\mathrm{AIC}(\hat{G}(k) \mid \alpha)$								
				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^C
(100, 100, 100, 100)	2	1	0.25	6.25	6.22	**3.51**	**3.51**	4.63	4.63	6.25	6.25	6.25
			0.5	6.71	6.63	*3.64*	**3.58**	4.62	4.62	6.71	6.70	6.71
			0.95	7.00	6.92	*3.81*	**3.35**	4.39	4.59	6.99	6.99	7.00
		2	0.25	5.12	5.12	**3.47**	*3.50*	3.57	3.67	5.13	5.13	5.12
			0.5	4.95	4.94	**3.33**	*3.44*	3.54	3.76	4.95	4.95	4.95
			0.95	4.21	4.12	**2.95**	*3.00*	3.22	3.30	4.20	4.20	4.21
	5	1	0.25	4.53	4.52	**3.87**	*3.88*	4.15	4.71	4.53	4.53	4.53
			0.5	4.45	4.44	**3.81**	*3.81*	4.29	4.57	4.45	4.45	4.45
			0.95	3.70	3.67	3.34	**3.08**	3.22	3.89	3.68	3.68	3.70
		2	0.25	4.22	4.21	**3.79**	*3.81*	3.93	4.01	4.22	4.22	4.22
			0.5	4.03	4.02	**3.60**	*3.60*	3.71	3.75	4.03	4.03	4.03
			0.95	2.44	2.40	**1.99**	*1.99*	2.01	2.09	2.42	2.42	2.44
(500, 500, 500, 500)	2	1	0.25	6.22	6.21	**3.51**	**3.51**	4.63	4.63	6.22	6.22	6.22
			0.5	6.60	6.59	*3.64*	**3.53**	4.62	4.62	6.60	6.60	6.60
			0.95	6.70	6.64	*3.27*	**3.11**	3.11	4.59	6.70	6.70	6.70
		2	0.25	4.96	4.96	**3.43**	*3.49*	3.59	3.83	4.96	4.97	4.96
			0.5	4.96	4.96	**3.36**	*3.44*	3.54	3.74	4.96	4.96	4.96
			0.95	3.84	3.80	**2.90**	*3.12*	3.22	3.37	3.84	3.84	3.84
	5	1	0.25	4.42	4.41	**3.88**	**3.88**	4.16	4.70	4.42	4.42	4.42
			0.5	4.38	4.37	**3.80**	*3.82*	3.93	4.56	4.38	4.38	4.38
			0.95	3.34	3.33	3.19	**3.06**	3.12	3.12	3.34	3.34	3.34
		2	0.25	4.16	4.15	**3.77**	*3.79*	3.90	3.98	4.16	4.16	4.16
			0.5	4.01	3.99	**3.59**	*3.59*	3.65	3.80	4.01	4.01	4.01
			0.95	2.51	2.48	**2.05**	*2.05*	2.05	2.07	2.51	2.51	2.51

Table 8 Probability (%) of selecting correct number of clusters with each criterion ($k^* = 2$; 1 / 2)

(n_1, n_2)	p	δ	ρ	AIC($\hat{G}(k) \mid \alpha$) The values of α: 0	0.1	0.5	1	1.5	2	BIC	CAIC	AIC^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

Table 8 Probability (%) of selecting correct number of clusters with each criterion ($k^* = 2$; 1 / 2) (continued)

				$\mathrm{AIC}(\hat{G}(k) \mid \alpha)$								
				The values of α								
(n_1, n_2)	p	δ	ρ	*0*	*0.1*	*0.5*	*1*	*1.5*	*2*	*BIC*	*CAIC*	AIC^C
(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 ($k^* = 2$; 2 / 2)

(n_1, n_2)	p	δ	ρ	AIC($\hat{G}(k)\mid\alpha$) The values of α: *0*	*0.1*	*0.5*	*1*	*1.5*	*2*	*BIC*	*CAIC*	*AIC^C*
(100, 100)	2	1	0.25 0.5	0.00	0.00	11.75	**100.0**	**100.0**	0.00	0.00	0.00	0.00
				0.00	0.00	1.34	**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	**100.0**	**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	60.82	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
	5	1	0.25	0.00	0.00	58.98	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.5	0.00	0.00	29.66	**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
		2	0.25	0.00	0.00	90.17	**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	**100.0**	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
(500, 500)	2	1	0.25	0.00	0.00	3.63	**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	0.00	**100.0**	**100.0**	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	61.31	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.5	0.00	0.00	80.93	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.95	0.00	0.00	80.95	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
	5	1	0.25	0.00	0.00	38.07	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.5	0.00	0.00	26.35	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.95	0.00	0.00	29.96	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
		2	0.25	0.00	0.00	60.01	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.5	0.00	0.00	59.39	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00
			0.95	0.00	0.00	78.29	**100.0**	**100.0**	**100.0**	0.00	0.00	0.00

Table 10 Probability (%) of selecting correct number of clusters with each criterion (k^* = 4; 1 / 5)

				$\mathrm{AIC}(\hat{G}(k)\,|\,\alpha)$								
				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*	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

Table 10 Probability (%) of selecting correct number of clusters with each criterion (k^* = 4; 1 / 5) (continued)

(n_1, n_2, n_3, n_4)	p	δ	ρ	AIC$(\hat{G}(k) \mid \alpha)$ — *The values of* α: 0	0.1	0.5	1	1.5	2	*BIC*	*CAIC*	AIC^C
(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 11 Probability (%) of selecting correct number of clusters with each criterion (k^* = 4; 2 / 5)

				$\mathrm{AIC}(\hat{G}(k) \mid \alpha)$								
				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^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

Table 11 Probability (%) of selecting correct number of clusters with each criterion ($k^* = 4$; 2 / 5) (continued)

				$\mathrm{AIC}(\hat{G}(k) \mid \alpha)$								
				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^C
(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 12 Probability (%) of selecting correct number of clusters with each criterion ($k^* = 4$; 3 / 5)

				$\mathrm{AIC}(\hat{G}(k) \mid \alpha)$								
				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^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

Table 12 Probability (%) of selecting correct number of clusters with each criterion (k^* = 4; 3 / 5) (continued)

				$\mathrm{AIC}(\hat{G}(k)\,|\,\alpha)$								
				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*	AIC^C
(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 13 Probability (%) of selecting correct number of clusters with each criterion ($k^* = 4$; 4 / 5)

				$\mathrm{AIC}(\hat{G}(k) \mid \alpha)$								
				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*^{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

Table 13 Probability (%) of selecting correct number of clusters with each criterion ($k^* = 4$; 4 / 5) (continued)

(n_1, n_2, n_3, n_4)	p	δ	ρ	AIC$(\hat{G}(k) \mid \alpha)$ The values of α: *0*	*0.1*	*0.5*	*1*	*1.5*	*2*	*BIC*	*CAIC*	*AIC^C*
(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 14 Probability (%) of selecting correct number of clusters with each criterion ($k^* = 4$; 5 / 5)

				$\mathrm{AIC}(\hat{G}(k) \mid \alpha)$								
				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^C
(100, 100, 100, 100)	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*	**43.83**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	**20.46**	*18.88*	0.00	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	**56.67**	*8.82*	5.09	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**59.70**	*16.31*	8.11	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	*58.50*	**58.90**	7.29	3.28	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**92.08**	**92.08**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	*0.00*	*0.00*	**99.87**	**99.87**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.00	0.02	**42.39**	*35.19*	0.00	0.00	0.00	0.00
		2	0.25	0.00	0.00	57.25	57.46	**57.77**	57.25	0.00	0.00	0.00
			0.5	0.00	0.00	61.40	61.41	**62.20**	61.40	0.00	0.00	0.00
			0.95	0.07	*0.91*	**59.93**	**59.93**	**59.93**	**59.93**	0.35	0.43	0.07
(500, 500, 500, 500)	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*	**50.00**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.95	*0.00*	*0.00*	**82.11**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
		2	0.25	0.00	0.00	**59.31**	*15.83*	7.85	0.00	0.00	0.00	0.00
			0.5	0.00	0.00	**59.42**	*15.64*	7.81	0.00	0.00	0.00	0.00
			0.95	0.00	0.00	*30.20*	**35.51**	12.72	5.44	0.00	0.00	0.00
	5	1	0.25	*0.00*	*0.00*	**100.0**	**100.0**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
			0.5	0.00	0.00	**89.94**	**89.94**	*45.95*	0.00	0.00	0.00	0.00
			0.95	*0.00*	*0.00*	*0.00*	**42.12**	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
		2	0.25	0.00	0.00	59.65	**59.74**	59.66	59.65	0.00	0.00	0.00
			0.5	*0.00*	*0.00*	**60.98**	**60.98**	**60.98**	**60.98**	*0.00*	*0.00*	*0.00*
			0.95	0.00	0.01	*63.41*	**69.48**	**69.48**	**69.48**	0.00	0.00	0.00

Furthermore, $\mathrm{AIC}(\hat{G}(k)\,|\,0.5)$, $\mathrm{AIC}(\hat{G}(k)\,|\,1)$, and $\mathrm{AIC}(\hat{G}(k)\,|\,1.5)$ work well when not only sample size is small but also it is large that are shown in Tables 2 and 7.

Next, we consider the probability (%) of correctly selecting the number of clusters; this is shown in Tables 8 to 14. In the results in Tables 8 and 9, when $k^* = 2$, $\mathrm{AIC}(\hat{G}(k)\,|\,1)$ and $\mathrm{AIC}(\hat{G}(k)\,|\,1.5)$ almost always select the correct number of clusters, and $\mathrm{AIC}(\hat{G}(k)\,|\,0.5)$ is always either the best or second best method, overall. Tables 10 to 14 show the results when $k^* = 4$. Based on these results, we see that $\mathrm{AIC}(\hat{G}(k)\,|\,1)$ is almost always the best method, followed by $\mathrm{AIC}(\hat{G}(k)\,|\,0.5)$ through all situations.

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 $\mathrm{AIC}(\hat{G}(k)\,|\,0.5)$, $\mathrm{AIC}(\hat{G}(k)\,|\,1)$, or $\mathrm{AIC}(\hat{G}(k)\,|\,1.5)$ 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 analysis. 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 minimising 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 $\mathrm{AIC}(\hat{G}(k)\,|\,\alpha)$, given in (10), we set α to be 0, 0.1, 0.5, 1, 1.5, 2.

Table 15 The number of times each number of clusters was selected by each criterion for the 'iris' data

	$\mathrm{AIC}(\hat{G}(k)\,|\,\alpha)$						*BIC*	*CAIC*	*AIC^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	1,648	10,000	10,000	0	0	0
3 (= k^*)	0	0	*1,042*	**7,987**	0	0	0	0	0
4	973	1,516	3,858	365	0	0	1,091	1,577	973
5	9,027	8,484	5,100	0	0	0	8,909	8,423	9,027

In Table 15, we list how many times each candidate number of clusters was selected. Furthermore, as in Tables 1 to 14, for simplicity, $\mathrm{BIC}(\hat{G}(k))$, $\mathrm{CAIC}(\hat{G}(k))$, and $\mathrm{AIC^C}(\hat{G}(k))$ are written as BIC, CAIC, and AIC$^\mathrm{C}$, respectively.

$\mathrm{AIC}(\hat{G}(k))$, $\mathrm{BIC}(\hat{G}(k))$, $\mathrm{CAIC}(\hat{G}(k))$, $\mathrm{AIC^C}(\hat{G}(k))$, and $\mathrm{AIC}(\hat{G}(k) \mid 0.1)$ 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 $\mathrm{AIC}(\hat{G}(k))$. Thus, overall, $\mathrm{AIC}(\hat{G}(k) \mid 1.5)$ and $\mathrm{AIC}(\hat{G}(k) \mid 2)$ tend to select fewer clusters. Moreover, $\mathrm{AIC}(\hat{G}(k) \mid 1)$ selected the true number of clusters the most frequently. Thus, $\mathrm{AIC}(\hat{G}(k) \mid 1)$ is the best method for selecting the number of clusters, and, as can be seen in the table, the second best methods is $\mathrm{AIC}(\hat{G}(k) \mid 0.5)$.

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

5 Conclusions

In practice, prior to formally analysing 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 non-hierarchical. We briefly discussed these in Section 2. Compared to hierarchical clustering, non-hierarchical clustering can be applied to more types of data (Saito and Yadohisa, 2006), and so, in this paper, we have focused on non-hierarchical clustering.

We note that for some non-hierarchical 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), we showed the relationship between clustering analysis and maximum likelihood estimators in the multivariate linear regression model. Based on this relation, we guess that we can use the formal information criteria for selecting the number of clusters. However, when we use them, we often select the large number of clusters as shown in Table 15. Thus, adding the penalty parameter for the number of clusters, we proposed a new AIC-type criterion (10) 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 in (6).

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 $\mathrm{AIC}(\hat{G}(k) \mid 0.5)$, $\mathrm{AIC}(\hat{G}(k) \mid 1)$, or $\mathrm{AIC}(\hat{G}(k) \mid 1.5)$ to select the number of clusters. Furthermore, from the results of analysing the 'iris' data set (using built in data set in the R language), we note that $\mathrm{AIC}(\hat{G}(k) \mid 1)$ is the best criterion for selecting the number of clusters.

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

Acknowledgements

The authors wish to thank anonymous referees for their valuable comments.

References

Akaike, H. (1973) 'Information theory and an extension of the maximum likelihood principle', in Petrov, B.N. and Csáki, F. (Eds.): *The 2nd International Symposium on Information Theory*, Akadémiai Kiadó, Budapest, pp.267–281.

Bedrick, E.J. and Tsai, C-L. (1994) 'Model selection for multivariate regression in small samples', *Biometrics*, Vol. 50, No. 1, pp.226–231.

Bijnen, E.J. (1973) *Cluster Analysis*, Tilburg University Press, Tilburg, Netherlands.

Bozdogan, H. (1986) 'Multi-sample cluster analysis as an alternative to multiple comparison procedures', *Bulletin of Informatics and Cybernetics*, Vol. 22, Nos. 1–2, pp.95–130.

Bozdogan, H. (1987) 'Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions', *Psychometrika*, Vol. 52, No. 3, pp.345–370.

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

Fisher, R.A. (1936) 'The use of multiple measurements in taxonomic problems', *Annals of Eugenics*, Vol. 7, No. 2, pp.179–188.

Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The Elements of Statistical Learning*, 2nd ed., Springer Science+Business Media, Springer, New York.

Konishi, S. and Kitagawa, G. (2008) *Information Criteria and Statistical Modeling*, Springer Science+Business Media, Springer, New York.

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

Lütkepohl, H. (1996) *Handbook of Matrices*, JohnWiley and Sons, New York.

MacQueen, J.B. (1967) 'Some methods for classification and analysis of multivariate observations', in Neyman, J. (Ed.): *The Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, Vol. 1, No. 3, pp.281–298.

Manning, C.D., Raghavan, P. and Schütze, H. (2009) *Introduction to Information Retrieval*, Cambridge University Press.

Marriott, F.H.C. (1982) 'Optimization methods of cluster analysis', *Biometrika*, Vol. 69, No. 1, pp.417–421.

Muirhead, R.J. (1982) *Aspects of Multivariate Statistical Theory*, John Wiley and Sons, New York.

Rao, C.R., Shalabh, T.H. and Heumann, C. (2008) *Linear Models and Generalization, Least Squares and Alternatives*, 3rd extended ed., Springer-Verlag Berlin Heidelberg.

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

Saito, T. and Yadohisa, H. (2006) *Analysis of Relevant Data - Multidimensional Scaling and Cluster Analysis*, Kyoritsu shuppan (in Japanese).

Schott, J.R. (2005) *Matrix Analysis for Statistics*, John Wiley and Sons, New York.

Schwarz, G. (1978) 'Estimating the dimension of a model', *Annals of Statistics*, Vol. 6, No. 2, pp.461–464.

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

Yanagihara, H. and Yoshimoto, A. (2005) 'Clustering individual growth patterns in even-aged forest stand', *FORMATH*, in Japanese, Vol. 4, No. 1, pp.49–70.

Appendix

Prove of the renewal condition for the cluster criterion in non-hierarchical clustering

In Section 2, we illustrated the cluster criterion for non-hierarchical clustering, and we then considered the cluster criteria that were listed in the following Table.

Table A1 Cluster criteria for renewing clustered data

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

We now consider whether y_r in the s^{th} cluster moves to the t^{th} 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 i^{th} 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 y_r in the s^{th} cluster to the t^{th} cluster is $CC(W(G'(k))) < CC(W(G(k)))$ in each cluster criterion. In this section, letting $a = \{n_t / (n_t + 1)\}^{1/2} (y_r - \bar{y}_t)$ and $b = \{n_s / (n_s - 1)\}^{1/2} (y_r - \bar{y}_s)$, we prove that, for each cluster criterion, this condition coincides with the renewal conditions listed in Table A2.

The condition (11) is directly proved based on some simple result. Moreover, Yanagihara and Yoshimoto (2005) has already proposed the equivalent condition (12) is same as $CC(W(G'(k))) < CC(W(G(k)))$ for (ii). In order to prepare some notations and relations, we touch these proved conditions. Further, using the corresponding equivalent condition for each cluster criterion, we can easily check $CC(W(G'(k))) < CC(W(G(k)))$ is satisfied or not.

Table A2 The conditions for renewal of each cluster based on the corresponding $CC(W(G(k)))$

CC(W(G(k)))	*Renewal condition*	
(i)	$a'a < b'b$	(11)
(ii)	$(a-b)'W\left(G(k)\right)^{-1}(a+b) < a'W\left(G(k)\right)^{-1}(ab'-ba')W\left(G(k)\right)^{-1}b$	(12)
(iii)	$a'W_t^{-1}a\det\left(W_t\right) < b'W_s^{-1}b\det\left(W_s\right)$	(13)
(iv)	$\left(1+a'W_t^{-1}a\right)^{n_t+1}\det\left(W_t\right) < \left(1-b'W_s^{-1}b\right)^{-(n_s-1)}\det\left(W_s\right)$	(14)
(v)	$1+(a+b)'W\left(G(k)\right)^{-1}(a+b) - a'W\left(G(k)\right)^{-1}(ab'-ba')W\left(G(k)\right)^{-1}b < \left\{\frac{n_t}{n_s}\left(1-\frac{1}{n_s}\right)^{n_s-1}\left(1+\frac{1}{n_t}\right)^{n_t+1}\right\}^{2/n}$	(15)
(vi)	$\left(1-b'W_s^{-1}b\right)^{n_s-1}\left(1+a'W_t^{-1}a\right)^{n_t+1}\det\left(W_tW_s^{-1}\right) < \left\{\frac{n_t}{n_s}\left(1-\frac{1}{n_s}\right)^{n_s-1}\left(1+\frac{1}{n_t}\right)^{n_t+1}\right\}^{2}$	(16)

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

$$W\left(G'(k)\right) = \sum_{\substack{j=1,\ldots,k\\ j\neq s,t}}\sum_{i\in C_j}\left(y_i-\bar{y}_j\right)\left(y_i-\bar{y}_j\right)' + W_{t'} + W_{s'}, \tag{17}$$

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

$$\bar{y}_{t'} = \frac{n_t \bar{y}_t + y_r}{n_t + 1} \tag{18}$$

$$\bar{y}_{s'} = \frac{n_s \bar{y}_s - y_r}{n_s - 1}, \tag{19}$$

since $\bar{y}_s = \sum_{i \in C_s} y_i / n_s$ and $\bar{y}_t = \sum_{i \in C_t} y_i / n_t$. Hence, we can calculate $W_{t'}$ and $W_{s'}$ in equation (17) by using these results.

Using (18), we derive $y_i - \bar{y}_{t'} = y_i - \bar{y}_t - (y_r - \bar{y}_t)/(n_t + 1)$. Then, the following result is obtained:

$$\begin{aligned} W_{t'} &= \sum_{i \in C_t} (y_i - \bar{y}_{t'})(y_i - \bar{y}_{t'})' + (y_r - \bar{y}_{t'})(y_r - \bar{y}_{t'})' \\ &= \sum_{i \in C_t} \left\{ y_i - \bar{y}_t - \frac{y_r - \bar{y}_t}{n_t + 1} \right\} \left\{ y_i - \bar{y}_t - \frac{y_r - \bar{y}_t}{n_t + 1} \right\}' \\ &\quad + \left(\frac{n_t}{n_t + 1} \right)^2 (y_r - \bar{y}_t)(y_r - \bar{y}_t)' \\ &= W_t + aa', \end{aligned}$$

since $a = \{n_t / (n_t + 1)\}^{1/2} (y_r - \bar{y}_t)$, $\sum_{i \in C_t} (y_i - \bar{y}_t)(y_r - \bar{y}_t)' = O_p$, and $\sum_{i \in C_t} (y_r - \bar{y}_t)(y_i - \bar{y}_t)' = O_p$, where O_p is a $p \times p$ zero matrix. By a similar calculation, using (19), we derive:

$$\begin{aligned} W_{s'} &= \sum_{i \in C_s} (y_i - \bar{y}_{s'})(y_i - \bar{y}_{s'})' - (y_r - \bar{y}_{s'})(y_r - \bar{y}_{s'}) \\ &= \sum_{i \in C_s} \left\{ y_i - \bar{y}_s + \frac{y_r - \bar{y}_s}{n_s - 1} \right\} \left\{ y_i - \bar{y}_s + \frac{y_r - \bar{y}_s}{n_s - 1} \right\}' \\ &\quad - \left(\frac{n_s}{n_s - 1} \right)^2 (y_r - \bar{y}_s)(y_r - \bar{y}_s)' \\ &= W_s - bb', \end{aligned}$$

since $b = \{n_s / (n_s - 1)\}^{1/2} (y_r - \bar{y}_s)$, $\sum_{i \in C_s} (y_i - \bar{y}_s)(y_r - \bar{y}_s)' = O_p$, and $\sum_{i \in C_s} (y_r - \bar{y}_s)(y_i - \bar{y}_s)' = O_p$.

From (17), these results imply:

$$W(G'(k)) = W(G(k)) + aa' - bb'.$$

This directly shows that $CC(W(G'(k))) < CC(W(G(k)))$, which is based on (i), coincides with the renewal condition (11).

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

$$\det\left(W\left(G'(k)\right)\right) = \det\left(W\left(G(k)\right)\right)\det\left(I_p + E\right) \tag{20}$$

Let λ_i be the i^{th} eigenvalue of E, then $\det(I_p + E) = \prod_{i=1}^{p}(1+\lambda_i)$. From Lütkepuhl (1996, p. 65), the nonzero eigenvalues in $\lambda_1, \ldots, \lambda_p$ are equal to υ_1 and υ_2, which are the eigenvalues of a 2 × 2 matrix $F = \{(1,0)'(1,0) + (0,1)'(0,-1)\}(a,b)'W(G(k))^{-1}(a,b)$. Thus, $\det(I_p + E) = 1 + \upsilon_1 + \upsilon_2 + \upsilon_1\upsilon_2$. Since υ_1 and υ_2 are eigenvalues of F, from the Cayley-Hamilton theorem, υ_1 and υ_2 are the solutions to the following quadratic equation:

$$\upsilon^2 - \operatorname{tr}(F)\upsilon + \det(F) = 0.$$

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

$$\begin{aligned}\upsilon_1 + \upsilon_2 &= \operatorname{tr}(F) = (a-b)'W\left(G(k)\right)^{-1}(a+b),\\ \upsilon_1\upsilon_2 &= \det(F) = -a'W\left(G(k)\right)^{-1}(ab' - ba')W\left(G(k)\right)^{-1}b.\end{aligned}$$

Hence, we obtain:

$$\begin{aligned}&\det\left(I_p + E\right)\\ &= 1 + (a-b)'W\left(G(k)\right)^{-1}(a+b) - a'W\left(G(k)\right)^{-1}(ab' - ba')W\left(G(k)\right)^{-1}b.\end{aligned} \tag{21}$$

We directly obtain that the renewal condition (12) coincides with $CC(W(G'(k))) < CC(W(G(k)))$ based on (ii) (this was previously shown by Yanagihara and Yoshimoto (2005)).

Using these results, we can prove $CC(W(G'(k))) < CC(W(G(k)))$ based on the cluster criterion (iii) is equivalent to the renewal condition (13). When we use the cluster criterion (iii), since it is organised as a summation of $\det(W_i)^{1/p}$, satisfying $CC(W(G'(k))) < CC(W(G(k)))$ coincides with satisfying $\det(W_{s'})^{1/p} + \det(W_{t'})^{1/p} < \det(W_s)^{1/p} + \det(W_t)^{1/p}$. This condition is equivalent to $1 < \exp\{\det(W_s)^{1/p} - \det(W_{s'})^{1/p} + \det(W_t)^{1/p} - \det(W_{t'})^{1/p}\} = \exp\{(\det(W_s) - \det(W_{s'}) + \det(W_t) - \det(W_{t'}))/p\}$. Thus, we consider the condition that $0 < \det(W_s) - \det(W_{s'}) + \det(W_t) - \det(W_{t'})$.

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

$$\det(P+Q) = \left\{1 + \operatorname{tr}\left(P^{-1}Q\right)\right\}\det(P),$$

for any non-singular matrix P and some appropriate dimensional matrix Q with rank $(Q) = 1$. Since $W_{t'} = W_t + aa', W_s = W_{s'} + bb'$, rank $(aa') = 1$, rank $(bb') = 1$, and we assume W_i is always a non-singular matrix, we obtain:

$$\det\left(W_{t'}\right) = \left(1 + a'W_t^{-1}a\right)\det\left(W_t\right). \tag{22}$$

$$\det\left(W_s\right) = \left(1 + b'W_{s'}^{-1}b\right)\det\left(W_{s'}\right). \tag{23}$$

Hence, we obtain:

$$\det(W_s) - \det(W_{s'}) + \det(W_t) - \det(W_{t'})$$
$$= b'W_{s'}^{-1}b\det(W_{s'}) - a'W_t^{-1}a\det(W_t).$$

From Schott (2005, Corollary 1.7.2), the following equation holds for any non-singular matrix R:

$$(R - cc')^{-1} = R^{-1} + \frac{R^{-1}cc'R^{-1}}{1 - c'R^{-1}c}, \tag{24}$$

for some vector c that is of suitable dimension, and such that $R - cc'$ is a non-singular matrix. We note that $\det(R - cc') \neq 0$ is equivalent to $1 - c'Rc \neq 0$ (see, e.g., Siotani et al., 1985, A.1.3).

Here, since $W_{s'} = W_s - bb'$ is assumed to be a non-singular matrix, we obtain $1 - b'W_s^{-1}b' \neq 0$ (see, e.g., Siotani et al., 1985, A.1.3). Thus, from (24) and $\det(W_s) \neq 0$, we have:

$$W_{s'}^{-1} = W_s^{-1} + \frac{W_s^{-1}bb'W_s^{-1}}{1 - b'W_s^{-1}b}.$$

Hence, we obtain $b'W_{s'}^{-1}b = b'W_s^{-1}b/(1 - b'W_s^{-1}b)$ and $1 + b'W_{s'}^{-1}b = (1 - b'W_s^{-1}b)^{-1}(> 0)$. Using this and (23), we see that the renewal condition $0 < b'W_{s'}^{-1}b\det(W_{s'}) - a'W_t^{-1}a\det(W_t)$ is equivalent to:

$$a'W_t^{-1}a\det(W_t) < \frac{b'W_s^{-1}b}{1 - b'W_s^{-1}b}\det(W_{s'}).$$

From (23) and $1 + b'W_{s'}^{-1}b = (1 - b'W_s^{-1}b)^{-1}$, we can prove the renewal condition (13) is equivalent to $CC(W(G'(k))) < CC(W(G(k)))$, based on the cluster criterion (iii).

Next, we prove that the renewal condition (14) is equivalent to $CC(W(G'(k)))$ $< CC(W(G(k)))$ 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 s^{th} and t^{th} clusters, we obtain $CC(W(G(k)))$ $= \det(W_s)^{n_s}\det(W_t)^{n_t} z$ and $CC(W(G'(k))) = \det(W_{s'})^{n_s - 1}\det(W_{t'})^{n_t + 1} z$, where $z = \prod_{\substack{i=1,\dots,k \\ i \neq s,t}} \det(W_i)^{n_i}$ can be considered to be a constant. Note that $z > 0$, since $\det(W_i)$ > 0 for any i and $G(k)$. Thus, the condition for $CC(W(G'(k))) < CC(W(G(k)))$ based on (iv) coincides with $\det(W_{s'})^{n_s-1}\det(W_{t'})^{n_t+1} < \det(W_s)^{n_s}\det(W_t)^{n_t}$. Since $\det(W_i) > 0$ for any i and $G(k)$, from (22) and (23), this inequality is equivalent to:

$$\left(1 + a'W_t^{-1}a\right)^{n_t+1}\det(W_t) < \left(1 + b'W_{s'}^{-1}b\right)^{n_s}\det(W_{s'}).$$

Hence, since $1 + b'W_{s'}^{-1}b = (1 - b'W_s^{-1}b)^{-1}$ and using (23), we can prove that (14) is coincides with $CC(W(G'(k))) < CC(W(G(k)))$, based on the cluster criterion (iv).

Next, we prove that the renewal condition (15) coincides with $CC(W(G'(k)))$ < $CC(W(G(k)))$ when we use the cluster criterion (v). Based on (v), since $G(k)$ and $G'(k)$ differ in the s^{th} and t^{th} clusters, we can express this condition as:

$$n\left\{\log\left(\det\left(W\left(G'(k)\right)\right)\right)-\log\left(\det\left(W\left(G(k)\right)\right)\right)\right\}$$
$$<2\left\{(n_s-1)\log(n_s-1)+(n_t+1)\log(n_t+1)-\left(n_s\log(n_s)+n_t\log(n_t)\right)\right\}.$$

The left-hand side in the above inequation is $n\log\{\det(I_p+E)\}$, from (20). On the other hand, the right-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:

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

Hence, we obtain that the renewal condition (15) is equivalent to $CC(W(G'(k)))$ < $CC(W(G(k)))$, based on (v) from (21).

Finally, we prove that $CC(W(G'(k)))<CC(W(G(k)))$, based on the cluster criterion (vi), is equivalent to (16). Considering the difference between $G(k)$ and $G'(k)$, this inequation is equivalent to:

$$(n_s-1)\log\left(\det\left(W_{s'}\right)\right)+(n_t+1)\log\left(\det\left(W_{t'}\right)\right)$$
$$-2\left\{(n_s-1)\log(n_s-1)+(n_t+1)\log(n_t+1)\right\}$$
$$<n_s\log\left(\det\left(W_s\right)\right)+n_t\log\left(\det\left(W_t\right)\right)-2\left\{n_s\log(n_s)+n_t\log(n_t)\right\}.$$

Using (23), we obtain $\det(W_{s'})=(1-b'W_s^{-1}b)\det(W_s)$, since $1+b'W_{s'}^{-1}b=(1-bW_s^{-1}b)^{-1}$. Furthermore, using (22), the above inequation is equivalent to:

$$(n_s-1)\log\left(1-b'W_s^{-1}b\right)-\log\left(\det\left(W_s\right)\right)$$
$$+(n_t+1)\log\left(1+a'W_t^{-1}a\right)+\log\left(\det\left(W_t\right)\right)$$
$$<2\left\{n_s\log\left(1-\frac{1}{n_s}\right)-\log(n_s-1)+n_t\log\left(1+\frac{1}{n_t}\right)+\log(n_t+1)\right\}.$$

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

$$\log\left\{\left(1-b'W_s^{-1}b\right)^{n_s-1}\left(1+a'W_t^{-1}a\right)^{n_t+1}\det\left(W_tW_s^{-1}\right)\right\}$$
$$<\log\left[\left(1-\frac{1}{n_s}\right)^{n_s}\left(1+\frac{1}{n_t}\right)^{n_t}\frac{n_t+1}{n_s-1}\right]^2.$$

Hence, we obtain that the renewal condition (16) coincides with $CC(W(G'(k)))$ < $CC(W(G(k)))$, based on (vi).