No.188

A family of the information criteria using the phi-divergence for categorical data

Haruhiko Ogasawara

August 2017

Otaru University of Commerce

Discussion Paper, No.188, August, 2017, Center for Glocal Strategy,
Otaru University of Commerce, Otaru, Japan. ("glocal": a coined word from "global" and "local")

A family of the information criteria using the phi-divergence for categorical data

Haruhiko Ogasawara
Otaru University of Commerce

This work was partially supported by a Grant-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology (JSPS KAKENHI, Grant No.17K00042). Author's address: Otaru University of Commerce, 3-5-21, Midori, Otaru 047-8501, Japan; Email: hogasa@res.otaru-uc.ac.jp, hogasa.otaru@gmail.com.

A family of the information criteria using the phi-divergence for categorical data

The risk of the phi-divergence of a statistical model for categorical data is defined using two independent sets of data. The asymptotic bias of the phi-divergence based on current data as an estimator of the risk is shown to be equal to the negative penalty term of the Akaike information criterion (AIC). Though the higher-order asymptotic bias is derived, the higher-order bias depends on the form of the phi-divergence and the estimation method of parameters using a possible different form of the phi-divergence. An approximation to the higher-order bias is obtained based on the simple result of the saturated model. The information criteria using this approximation yield improved results in simulations for model selection. Some cases of the power divergences forming a subfamily of the phi-divergence show advantages over the AIC in simulations.

Keywords: power divergence; risk; model selection; asymptotic bias; Akaike information criterion.

1. Introduction

Various information criteria have been proposed. Many of them are based on the likelihood of parameters in a statistical model, whose typical cases are the Akaike information criterion (AIC, Akaike, 1973), the Takeuchi information criterion (TIC, Takeuchi, 1973; for the TIC see e.g., Burnham & Anderson, 2010, pp. 65-66) and the Bayes information criterion (BIC, Schwarz, 1978), The Mallows (1973) C_p for model selection in linear regression takes a least squares (LS) form, which is also seen as the Gauss discrepancy (Linhart & Zucchini, 1986, p.18) based on a likelihood. It is known that the C_p is asymptotically equivalent to the AIC.

In covariance structure analysis, the normal-theory (NT) or asymptotically-distribution-free (ADF) generalized LS (GLS) criteria for model selection have also been proposed (Browne & Cudeck, 1989; Yanagihara, Himeno & Yuan, 2010; Ogasawara, 2017), which are asymptotically equal to the AIC and TIC under some conditions. The criteria based on cross validation (cross validation criteria, CVCs; see Allen, 1971; Stone, 1974; Yanagihara, Yuan, Fujisawa & Hayashi, 2013) are similarly used for model selection, and can be shown to be asymptotically equal to the AIC or TIC (Stone, 1977).

In this paper, models for categorical or multinomial data are dealt with. For these models, among the criteria shown above, the AIC, TIC, BIC and CVC can be used, where the likelihood based on multinomial or categorical distributions are used for the criteria except the CVC. The ϕ-divergence statistic (see e.g., Cressie & Pardo, 2002a; Pardo, 2006) is a generalization of the log-likelihood ratio statistic for evaluating the goodness-of-fit of a model. While the original definition of the AIC is based on the likelihood rather than the likelihood ratio, the latter can also be used for model selection in essentially the same way, since an added term in the log-likelihood ratio common to candidate models is irrelevant to model selection.

The ϕ-divergences are also used for estimation of parameters as well as criteria for the badness of a model, whose estimators are the minimizing ϕ-divergence estimators (MϕEs; Morales, Pardo & Vajda, 1995, p.350; Pardo, 2006, Chapter 5). The ϕ-divergences and MϕEs are used in log-linear models (Cressie & Pardo, 2000, 2002b;

Cressie, Pardo & Pardo, 2003), logistic regression for grouped data (Pardo, Pardo & Pardo, 2005) and latent class models (Felipe, Miranda & Pardo, 2015).

Since the main term of the AIC can be replaced by the -2 times the log-likelihood ratio, it is natural to consider the information criterion using the ϕ-divergence, whose special case is the AIC using the ratio. The so-called penalty term in the AIC i.e., 2 times the number of independent parameters in a model is given by the negative asymptotic bias of -2 times the log-likelihood ratio using the maximum likelihood estimator (MLE) as an estimator of the corresponding risk. Note that the risk is defined by the two-fold expectation of -2 times the log-likelihood ratio i.e., one for the expectation of data independent of current data, and the other for the expectation of the current data yielding the MLE. Note that the CVC is a numerical evaluation of the risk.

It will be shown that the asymptotic bias of order $O(1)$ for the ϕ-divergence as an estimator of the risk is equal to -2 times the number of independent parameters, which is the negative penalty term in the AIC. Note that the asymptotic bias is common to different ϕ-divergences using different MϕEs. The corresponding higher-order asymptotic bias of a ϕ-divergence based on MϕEs using different ϕ-divergences will be shown, where the results depend on types of ϕ-divergences and MϕEs.

The ϕ information criterion denoted by ϕIC or PIC will be defined similarly to the AIC. It will be shown that the AIC does not necessarily give best results in model selection among typical ϕICs. The higher-order bias term can be used for correction of the remaining bias yielding the modified ϕIC (MϕIC or MPIC). Since the higher order term is complicated, a simple approximation (M*ϕIC or M*PIC) will be developed. It will be shown that the M*ϕIC performs better than the ϕIC in simulations for model selection.

2. The bias of the ϕ-divergence

The ϕ-divergence statistic for K-category multinomial data is defined by

$$C_\phi = \frac{2n}{\phi''(1)} D_\phi = \frac{2n}{\phi''(1)} D_\phi(\mathbf{p}, \boldsymbol{\pi}) \quad \text{with} \quad D_\phi = \sum_{k=1}^{K} \pi_k \phi(p_k / \pi_k), \tag{2.1}$$

where $\mathbf{p} = (p_1, ..., p_K)'$ is a $K \times 1$ vector of sample proportions for *K* categories based on *n* observations; $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta}) = (\pi_1, ..., \pi_K)'$ with $\pi_k = \pi_k(\boldsymbol{\theta})$ $(k = 1, ..., K)$ is a $K \times 1$ vector of model-based probabilities, which are functions of a $q \times 1$ vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_q)'$ of parameters ($q \le K - 1; \boldsymbol{\theta} \in \Theta,\ \Theta \subset R^q$); the convex function $\phi(x)$ is assumed to have the following properties:

$$\begin{aligned} &x > 0,\ \phi(1) = 0,\ \phi'(1) \text{ (the first derivative at } x = 1\text{)}, \\ &\phi''(1) \text{ (the second derivative at } x = 1\text{)} > 0, \\ &0\phi(0/0) = 0,\ 0\phi(v/0) = \lim_{u \to +\infty} \{\phi(u)/u\} \end{aligned} \tag{2.2}$$

(see e.g., Cressie & Pardo, 2002a; Pardo, 2006, Section 1.2), where D_ϕ was introduced by Csiszár (1963) and Ali and Silvey (1966).

When $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}(\hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}}$ is the $q \times 1$ vector of MϕEs with $\phi(\cdot)$ being possibly different from $\phi(\cdot)$ in $D_\phi(\cdot)$ of (2.1), we have the ϕ-divergence statistic

$$\hat{C}_\phi = \frac{2n}{\phi''(1)} \hat{D}_\phi = \frac{2n}{\phi''(1)} D_\phi(\mathbf{p}, \hat{\boldsymbol{\pi}}). \tag{2.3}$$

Probably, the most important sub-family of the ϕ-divergence is that of the power divergences (Cressie & Read, 1984; Read & Cressie, 1988), where

$$\phi(x) = \frac{x^{\lambda+1} - x}{\lambda(\lambda+1)} - \frac{x-1}{\lambda+1} \quad (-\infty < \lambda < +\infty, \lambda \ne 0, -1). \tag{2.4}$$

The cases of $\lambda = 0, -1$ are defined as the limiting values of $\phi(x)$ when $\lambda \to 0$ and $\lambda \to -1$, respectively. Equations (2.1) and (2.4) give an alternative expression of the power divergence

$$C_\phi = \frac{2n}{\phi''(1)} \sum_{k=1}^{K} \pi_k \left\{ \frac{(p_k / \pi_k)^{\lambda+1} - (p_k / \pi_k)}{\lambda(\lambda+1)} - \frac{(p_k / \pi_k) - 1}{\lambda+1} \right\}$$
$$= \frac{2n}{\phi''(1)\lambda(\lambda+1)} \left(\sum_{k=1}^{K} \frac{p_k^{\ \lambda+1}}{\pi_k^{\ \lambda}} - 1 \right). \tag{2.5}$$

Typical cases of the power divergences are as follows.

When $\lambda = 0$, $\hat{C}_\phi$ is the log-likelihood ratio statistic or the Kullback-Leibler (1951) divergence

$$G^2 = 2np_k \sum_{k=1}^{K} \ln(p_k / \hat{\pi}_k), \tag{2.6}$$

when $\lambda = -1$, $\hat{C}_\phi$ is the modified log-likelihood ratio statistic (Kullback, 1985)

$$GM^2 = 2n\hat{\pi}_k \sum_{k=1}^{K} \ln(\hat{\pi}_k / p_k), \tag{2.7}$$

when $\lambda = -2$, $\hat{C}_\phi$ is the Neyman (1949) modified chi-square statistic

$$NM^2 = n \sum_{k=1}^{K} (p_k - \hat{\pi}_k)^2 / p_k, \tag{2.8}$$

when $\lambda = -0.5$, $\hat{C}_\phi$ is the Freeman-Tukey (1950) chi-square statistic

$$T^2 = 4n \sum_{k=1}^{K} (\sqrt{p_k} - \sqrt{\hat{\pi}_k})^2, \tag{2.9}$$

and when $\lambda = 1$, $\hat{C}_\phi$ is Pearson's (1900) chi-square statistic

$$X^2 = n \sum_{k=1}^{K} (p_k - \hat{\pi}_k)^2 / \hat{\pi}_k. \tag{2.10}$$

For generality, restore the definition of $\hat{C}_\phi$ in (2.4) using a general $\phi(\cdot)$. Then, the risk of the model with $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$ is defined by

$$R_\phi \equiv \mathrm{E}\,\mathrm{E}^* \left\{ \frac{2n}{\phi''(1)} \sum_{k=1}^{K} \hat{\pi}_k \phi(p_k^* / \hat{\pi}_k) \right\}, \tag{2.11}$$

where $\mathbf{p}^* = (p_1^*, ..., p_K^*)'$ is a vector of sample proportions based on n observations independent of $\mathbf{p}$. $\mathrm{E}(\cdot)$ and $\mathrm{E}^*(\cdot)$ are the expectations using the distribution of $\mathbf{p}$ and $\mathbf{p}^*$, respectively. Recall that $\hat{\boldsymbol{\pi}}$ is obtained by a $\phi(\cdot)$, say $\phi_1(\cdot)$, which is possibly different from another $\phi(\cdot)$, say $\phi_2(\cdot)$, in (2.11). The bias of $\hat{C}_\phi$ for estimation of R_ϕ is given by $\mathrm{E}(\hat{C}_\phi) - R_\phi$.

Theorem 1. *Assume that the MϕE of $\boldsymbol{\theta}_0$ is given by minimizing C_ϕ when $\phi(\cdot) = \phi_1(\cdot)$ and that the ϕ-divergence statistic is constructed by C_ϕ when $\phi(\cdot) = \phi_2(\cdot)$ using the MϕE with $\phi_1(\cdot)$, where $\phi_2(\cdot)$ is possibly different from $\phi_1(\cdot)$. Suppose that $\hat{\boldsymbol{\pi}}$ can be expanded about $\boldsymbol{\pi}_0$ in the Taylor series up to the third powers of the elements of $\mathbf{p} - \boldsymbol{\pi}_0$ with the residual of order $O_p(n^{-2})$. Then, under correct model specification, the bias of $\hat{C}_\phi$ is given by*

$$\begin{aligned}
\mathrm{E}(\hat{C}_\phi) - R_\phi = \frac{2}{\phi_2''} \sum_{k=1}^{K} \underset{\text{(A)}}{\Bigg[} & -\frac{\phi_2''}{\pi_{0k}} \frac{\partial \pi_{0k}}{\partial \boldsymbol{\theta}_0'} \frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0'} n\mathrm{E}\{(\mathbf{p} - \boldsymbol{\pi}_0)(p_k - \pi_{0k})\} \\
& + n^{-1} \underset{\text{(B)}}{\Bigg[} -\frac{\phi_2''}{\pi_{0k}} \underset{\text{(C)}}{\Bigg[} \frac{1}{2} \left\{ \frac{\partial^2 \pi_{0k}}{(\partial \boldsymbol{\theta}_0')^{<2>}} \left(\frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0'} \right)^{<2>} + \frac{\partial \pi_{0k}}{\partial \boldsymbol{\theta}_0'} \frac{\partial^2 \boldsymbol{\theta}_0}{(\partial \boldsymbol{\pi}_0')^{<2>}} \right\} \qquad (2.12) \\
& \times n^2 \mathrm{E}\{(\mathbf{p} - \boldsymbol{\pi}_0)^{<2>} (p_k - \pi_{0k})\} \\
& + \frac{1}{6} \left\{ \frac{\partial^3 \pi_{0k}}{(\partial \boldsymbol{\theta}_0')^{<3>}} \left(\frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0'} \right)^{<3>} + 3 \frac{\partial^2 \pi_{0k}}{(\partial \boldsymbol{\theta}_0')^{<2>}} \left(\frac{\partial^2 \boldsymbol{\theta}_0}{(\partial \boldsymbol{\pi}_0')^{<2>}} \otimes \frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0'} \right) + \frac{\partial \pi_{0k}}{\partial \boldsymbol{\theta}_0'} \frac{\partial^3 \boldsymbol{\theta}_0}{(\partial \boldsymbol{\pi}_0')^{<3>}} \right\} \\
& \times n^2 \mathrm{E}\{(\mathbf{p} - \boldsymbol{\pi}_0)^{<3>} (p_k - \pi_{0k})\}_{\to O(n^{-2})} \underset{\text{(C)}}{\Bigg]}
\end{aligned}$$

$$-\frac{1}{2\pi_{0k}^{2}}(\phi_2^{(3)}+\phi_2'')\Bigg[_{\text{(D)}}\ \frac{\partial\pi_{0k}}{\partial\boldsymbol{\theta}_0'}\frac{\partial\boldsymbol{\theta}_0}{\partial\boldsymbol{\pi}_0'}n^2\mathrm{E}\{(\mathbf{p}-\boldsymbol{\pi}_0)(p_k-\pi_{0k})^2\}$$

$$+\frac{1}{2}\left\{\frac{\partial^2\pi_{0k}}{(\partial\boldsymbol{\theta}_0')^{<2>}}\left(\frac{\partial\boldsymbol{\theta}_0}{\partial\boldsymbol{\pi}_0'}\right)^{<2>}+\frac{\partial\pi_{0k}}{\partial\boldsymbol{\theta}_0'}\frac{\partial^2\boldsymbol{\theta}_0}{(\partial\boldsymbol{\pi}_0')^{<2>}}\right\}n^2\mathrm{E}\{(\mathbf{p}-\boldsymbol{\pi}_0)^{<2>}(p_k-\pi_{0k})^2\}_{\to O(n^{-2})}\ \Bigg]_{\text{(D)}}$$

$$+\frac{1}{2\pi_{0k}^{2}}(\phi_2^{(3)}+2\phi_2'')\Bigg[_{\text{(E)}}\ \left(\frac{\partial\pi_{0k}}{\partial\boldsymbol{\theta}_0'}\frac{\partial\boldsymbol{\theta}_0}{\partial\boldsymbol{\pi}_0'}\right)^{<2>}n^2\mathrm{E}\{(\mathbf{p}-\boldsymbol{\pi}_0)^{<2>}(p_k-\pi_{0k})\}$$

$$+2\times\frac{1}{2}\left(\frac{\partial\pi_{0k}}{\partial\boldsymbol{\theta}_0'}\frac{\partial\boldsymbol{\theta}_0}{\partial\boldsymbol{\pi}_0'}\right)\otimes\left\{\frac{\partial^2\pi_{0k}}{(\partial\boldsymbol{\theta}_0')^{<2>}}\left(\frac{\partial\boldsymbol{\theta}_0}{\partial\boldsymbol{\pi}_0'}\right)^{<2>}+\frac{\partial\pi_{0k}}{\partial\boldsymbol{\theta}_0'}\frac{\partial^2\boldsymbol{\theta}_0}{(\partial\boldsymbol{\pi}_0')^{<2>}}\right\}$$

$$\times n^2\mathrm{E}\{(\mathbf{p}-\boldsymbol{\pi}_0)^{<3>}(p_k-\pi_{0k})\}_{\to O(n^{-2})}\ \Bigg]_{\text{(E)}}$$

$$-\frac{1}{6\pi_{0k}^{3}}(\phi_2^{(4)}+2\phi_2^{(3)})\frac{\partial\pi_{0k}}{\partial\boldsymbol{\theta}_0'}\frac{\partial\boldsymbol{\theta}_0}{\partial\boldsymbol{\pi}_0'}n^2\mathrm{E}\{(\mathbf{p}-\boldsymbol{\pi}_0)(p_k-\pi_{0k})^3\}_{\to O(n^{-2})}$$

$$+\frac{1}{4\pi_{0k}^{3}}(\phi_2^{(4)}+4\phi_2^{(3)}+2\phi_2'')\left(\frac{\partial\pi_{0k}}{\partial\boldsymbol{\theta}_0'}\frac{\partial\boldsymbol{\theta}_0}{\partial\boldsymbol{\pi}_0'}\right)^{<2>}n^2\mathrm{E}\{(\mathbf{p}-\boldsymbol{\pi}_0)^{<2>}(p_k-\pi_{0k})^2\}_{\to O(n^{-2})}$$

$$-\frac{1}{6\pi_{0k}^{3}}(\phi_2^{(4)}+6\phi_2^{(3)}+6\phi_2'')\left(\frac{\partial\pi_{0k}}{\partial\boldsymbol{\theta}_0'}\frac{\partial\boldsymbol{\theta}_0}{\partial\boldsymbol{\pi}_0'}\right)^{<3>}n^2\mathrm{E}\{(\mathbf{p}-\boldsymbol{\pi}_0)^{<3>}(p_k-\pi_{0k})\}_{\to O(n^{-2})}\ \Bigg]_{\text{(B)}}\Bigg]_{\text{(A)}}$$

$$\begin{aligned}&+O(n^{-2})\\&\equiv b+n^{-1}b_\Delta+O(n^{-2})\\&=-2q+n^{-1}b_\Delta+O(n^{-2}),\end{aligned}$$

where $\mathbf{A}^{<k>}=\mathbf{A}\otimes\cdots\otimes\mathbf{A}$ (*k times of* $\mathbf{A}$) *is the k-fold Kronecker product of* $\mathbf{A}$ (*a vector or matrix*); $\dfrac{\partial^j\pi_{0k}}{(\partial\boldsymbol{\theta}_0)^{<j>}}=\dfrac{\partial^j\pi_k(\boldsymbol{\theta})}{(\partial\boldsymbol{\theta})^{<j>}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ $(j=1,\ 2,\ 3)$;

$\dfrac{\partial^j\boldsymbol{\theta}_0}{(\partial\boldsymbol{\pi}_0')^{<j>}}=\dfrac{\partial^j\hat{\boldsymbol{\theta}}}{(\partial\mathbf{p}')^{<j>}}\Big|_{\substack{\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}_0\\ \mathbf{p}=\boldsymbol{\pi}_0}}$ $(j=1,\ 2,\ 3)$ *are the j-th order partial derivatives of* $\hat{\boldsymbol{\theta}}$*, a* *possibly implicit function with respect to* **p** *at* $\mathbf{p} = \boldsymbol{\pi}_0$ *and* $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$ *;* $\mathrm{E}(\cdot)_{\to O(n^{-2})}$ *is the expectation up to order* $O(n^{-2})$ *;* $\phi_2\,'' = \phi_2\,''(1),\ \phi_2^{(j)} = \phi_2^{(j)}(1)\ (j = 3, 4)$ *; and e.g.,* $\underset{(\mathrm{A})}{[}\ \cdot\ \underset{(\mathrm{A})}{]}$ *is for ease of finding correspondence.*

The proof with the expression of $\partial^j \boldsymbol{\theta}_0 / (\partial \boldsymbol{\pi}_0\,')^{<j>} (j = 1,\ 2,\ 3)$ will be shown in the appendix. In Theorem 1. it is found that the asymptotic bias of order $O(1)$ i.e., $-2q$ does not depend on $\phi_1(\cdot)$ or $\phi_2(\cdot)$. The partial derivatives $\partial^j \pi_{0k} / (\partial \boldsymbol{\theta}_0)^{<j>}$ $= \partial^j \pi_k(\boldsymbol{\theta}) / (\partial \boldsymbol{\theta})^{<j>} |_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\ (j = 1,\ 2,\ 3)$ are determined only by the model $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$. It is known that

$$\frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0\,'} = \mathbf{I}_0^{-1} \frac{\partial \boldsymbol{\pi}_0\,'}{\partial \boldsymbol{\theta}_0} \mathrm{diag}^{-1}(\boldsymbol{\pi}_0) \tag{2.13}$$

(Morales et al., 1995, Theorem 3), where $\mathbf{I}_0 = \frac{\partial \boldsymbol{\pi}_0\,'}{\partial \boldsymbol{\theta}_0} \mathrm{diag}^{-1}(\boldsymbol{\pi}_0) \frac{\partial \boldsymbol{\pi}_0}{\partial \boldsymbol{\theta}_0\,'}$ is the Fisher information matrix per observation and $\mathrm{diag}^{-1}(\boldsymbol{\pi}_0) = \{\mathrm{diag}(\boldsymbol{\pi}_0)\}^{-1}$. Equation (2.13) does not depend on $\phi_1(\cdot)$ used for estimation of $\boldsymbol{\theta}_0$. However, Lemma 2 in the appendix shows that $\partial^j \boldsymbol{\theta}_0 / (\partial \boldsymbol{\pi}_0\,')^{<j>} (j = 2,\ 3)$ generally depend on $\phi_1(\cdot)$. Since there are $\phi_2^{(3)}$ and $\phi_2^{(4)}$ in b_Δ, we find that b_Δ depend on $\phi_2(\cdot)$ as well as $\phi_1(\cdot)$.

In the case of the power divergence, using (2.4), ϕ'', $\phi^{(3)}$ and $\phi^{(4)}$ in b_Δ become

$$\phi'' = 1,\ \phi^{(3)} = \lambda - 1 \text{ and } \phi^{(4)} = (\lambda - 1)(\lambda - 2), \tag{2.14}$$

respectively. From (2.14), it is found that for Pearson's $X^2 (\lambda = 1)$, $\phi^{(j)} (j = 2, 3, \ldots)$ vanish.

The ϕ information criterion (ϕIC or PIC) and the modified ϕIC (MϕIC or MPIC) are defined as

$$\phi\mathrm{IC} = \hat{C}_\phi + 2q \text{ and } \mathrm{M}\phi\mathrm{IC} = \hat{C}_\phi + 2q - n^{-1}\hat{b}_\Delta, \tag{2.15}$$

respectively, where $\hat{b}_\Delta$ is the sample counterpart of b_Δ. In $\hat{b}_\Delta$, $\boldsymbol{\theta}_0$ and $\boldsymbol{\pi}_0$ are replaced by $\hat{\boldsymbol{\theta}}$ and $\mathbf{p}$, respectively. Since $\mathrm{E}(\hat{b}_\Delta) = b_\Delta + O(n^{-1})$ under correct model specification, Theorem 1 with the known asymptotic distribution of $\hat{C}_\phi$ (Morales et al., 1995,) gives

Corollary 1. *Under the same conditions as in Theorem 1,*

$$\mathrm{E}(\phi\mathrm{IC}) - R_\phi = O(n^{-1}) \quad \textit{and} \quad \mathrm{E}(\mathrm{M}\phi\mathrm{IC}) - R_\phi = O(n^{-2}), \tag{2.16}$$

when $q \le K-1$. *The statistics* $\phi\mathrm{IC} - 2q$ *and* $\mathrm{M}\phi\mathrm{IC} - 2q$ *are asymptotically chi-square distributed with* $K-q-1$ *degrees of freedom when* $q \le K-2$.

Corollary 1 shows that the bias of $\hat{C}_\phi$ is removed up to orders $O(1)$ and $O(n^{-1})$ in the ϕIC and MϕIC, respectively and that the ϕIC and MϕIC are asymptotically equivalent to the AIC using the log-likelihood ratio in that they have a common asymptotic distribution. The quantity $2q$ in the added term in the ϕIC and MϕIC is seen as a upward correction of $\hat{C}_\phi$ which tends to be reduced as an estimator of R_ϕ. The term is also interpreted as a penalty for a model with q independent parameters as is done in the AIC.

3. An approximation to the higher-order asymptotic bias using the saturated model

While the added computation of the ϕIC over $\hat{C}_\phi$ is trivial, that of the MϕIC tends to be excessive. However, in the case of the saturated model, we obtain a simplified result.

Theorem 2. *In the case of the saturated mode with* $\hat{\boldsymbol{\theta}} = (p_1, \ldots, p_{K-1})' \equiv \mathbf{p}_{(K-1)}$ *without loss of generality and* $\hat{\boldsymbol{\pi}} = \mathbf{p}$, b_Δ *becomes*

$$b_\Delta = \frac{1}{\phi_2''}\left\{-\left(\frac{\phi_2^{(4)}}{2} + 2\phi_2^{(3)} + 2\phi_2''\right)\sum_{k=1}^{K}\frac{1}{\pi_{0k}} + \phi_2^{(4)}\left(K - \frac{1}{2}\right) + \phi_2^{(3)}(4K-2) + \phi_2''(3K-1)\right\}. \tag{3.1}$$

The proof will be given in the appendix. In the case of the saturated model, $\hat{\boldsymbol{\pi}} = \mathbf{p}$ with $\hat{C}_\phi = 0$ does not depends on $\phi_1(\cdot)$. Equation (3.1) depends on $\phi_2(\cdot)$ as well as

$\sum_{k=1}^{K} 1/\pi_{0k}$. In the case of the power divergence, recalling that $\phi_2'' = 1,\ \phi_2^{(3)} = \lambda - 1$ and $\phi_2^{(4)} = (\lambda-1)(\lambda-2)$, b_Δ becomes

$$
\begin{aligned}
b_\Delta &= -\frac{1}{2}\{(\lambda-1)(\lambda-2)+4(\lambda-1)+4\}\sum_{k=1}^{K}\frac{1}{\pi_{0k}} \\
&\quad +\frac{1}{2}(\lambda-1)(\lambda-2)(2K-1)+(\lambda-1)(4K-2)+3K-1 \qquad (3.2) \\
&= -\frac{1}{2}(\lambda^2+\lambda+2)\left(\sum_{k=1}^{K}\frac{1}{\pi_{0k}}-2K+1\right)-K+1 \\
&\le -(K-1)\left\{\frac{1}{2}(\lambda^2+\lambda+2)(K-1)+1\right\} \\
&\le -\frac{1}{8}(K-1)(7K+1).
\end{aligned}
$$

The second last inequality stems from $\sum_{k=1}^{K} 1/\pi_{0k} \ge K^2$. The last upper bound is attained when $\pi_{0k} = 1/K \ (k=1,\ldots,K)$ and $\lambda = -0.5$.

Note that the saturated model can be one of competing models especially when K is relatively small. In the case of K = 2, the saturated model with $q = K-1 = 1$ is the only parametric model for the binomial data. An approximation to b_Δ is defined as

$$b_\Delta^* = \{q/(K-1)\}b_\Delta, \qquad (3.3)$$

where b_Δ is given by (3.1) for the saturated model with K categories, which will be numerically illustrated in the next section. This approximation is based on heuristic findings that in numerical results, b_Δ is, in a crude sense, proportional to q.

It is known that the AIC tends to choose relatively complicated models in e.g., regression (Hurvich & Tsai, 1989; Fujikoshi & Satoh, 1997) and factor analysis (Akaike, 1987; Ichikawa, 1989; Ogasawara, 2016a). This corresponds to the fact that the correction term $2q$ in the AIC is not sufficiently large, which can be corrected by the higher-order correction term for the bias. For instance, Sugiura (1978) derived the exact bias of the AIC in linear regression under normality, which gives improved results of selecting correct models by the modified AIC. It is to be noted that Sugiura's (1978) result is useful in models other than regression (Burnham & Anderson, 2010, p.66). These findings with the asymptotic equivalence of the ϕ IC to the AIC give some justification for the approximation of (3.3).

Define

$$\begin{aligned} \mathrm{M}^*\phi\mathrm{IC} &= \hat{C}_\phi + 2q - n^{-1}\hat{b}_\Delta^* \\ &= \hat{C}_\phi + 2q - n^{-1}\{q/(K-1)\}\hat{b}_\Delta \\ &= \phi\mathrm{IC} - n^{-1}\{q/(K-1)\}\hat{b}_\Delta, \end{aligned} \tag{3.4}$$

where $\hat{b}_\Delta$ is the sample counterpart of b_Δ in (3.1) for the saturated model. The improved behavior of the $\mathrm{M}^*\phi\mathrm{IC}$ over the $\phi\mathrm{IC}$ will be numerically illustrated in the next section.

4. Numerical illustrations with simulations

In this section, we give numerical illustrations using simulations for model selection and biases of $\hat{C}_\phi$. The power divergences with $\lambda = 0$ (G^2, the log-likelihood ratio statistic), $\lambda = -1$ (GM^2, the modified log-likelihood ratio statistic), $\lambda = -2$ (the Neyman modified statistic), $\lambda = -0.5$ (the Freeman-Tukey statistic), $\lambda = 0.5$, and $\lambda = 2/3$ (the Cressie-Read statistic), $\lambda = 1$ (X^2, Pearson's statistic) and $\lambda = 2$ are used. The statistic with $\lambda = 2/3$ was proposed by Cressie and Read (1984) showing a good behavior as a goodness-of-fit statistic, which seems to give advantages in other respects (see e.g., Cressie et al., 2003; Pardo et al., 2005). The above statistics are employed as typical ones used in practice among the ϕ-divergence statistics. The $\phi\mathrm{IC}$ and $\mathrm{M}^*\phi\mathrm{IC}$ are used, rather than $\mathrm{M}\phi\mathrm{IC}$, using some or all of the above statistics.

Three sets of models are used. The first one with 4 categories was used by Fisher (1970, p.305) for the genetics of plants:

$$\boldsymbol{\pi} = \left(\frac{2+\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4}\right)' \ (0 < \theta < 1), \tag{4.1}$$

where $\theta_0 = 0.4$ is employed in this section. Note that the MLE is algebraically given by

$$\hat{\theta}_{\mathrm{ML}} = \frac{1}{2}[p_1 - 2p_2 - 2p_3 - p_4 + \{(p_1 - 2p_2 - 2p_3 - p_4)^2 + 8p_4\}^{1/2}]. \qquad (4.2)$$

For competing models,

$$\boldsymbol{\pi} = (\theta_1, \theta_2, \theta_2, 1 - \theta_1 - 2\theta_2) \ (0 < \theta_1 < 1, \ 0 < \theta_2 < 1, \ 0 < 1 - \theta_1 - 2\theta_2 < 1) \qquad (4.3)$$

and the saturated model are used. The model of (4.3) can also be seen as a saturated model when the second and third categories are combined. Note that all of the three candidate models are not incorrect ones though the first model is best in that it is most parsimonious among the candidates. Consequently, the best model is called a correct model in this paper. The second and third models are similar to overspecified models with extra covariates in regression analysis.

The second set of models with three categories was used by Bishop, Fienberg and Holland (1975, Example 14.7.1) for truncated Poisson variates:

$$\boldsymbol{\pi} = \{e^{-\theta}, \ \theta e^{-\theta}, \ 1 - (1 + \theta)e^{-\theta}\}' \ (0 < \theta < +\infty), \qquad (4.4)$$

where $\theta_0 = 1$ is used in this section. Note that the model more redundant than (4.4) is the saturated one. We use these two models.

The third set of models with four categories is an extended version of (4.4):

$$\boldsymbol{\pi} = \{e^{-\theta}, \ \theta e^{-\theta}, \ \theta^2 e^{-\theta}/2, 1 - (1 + \theta + 0.5\theta^2)e^{-\theta}\}' \ (0 < \theta < +\infty). \qquad (4.5)$$

A redundant model with

$$\begin{aligned} &\boldsymbol{\pi} = \{e^{-\theta_1}, \ \theta_1 e^{-\theta_1}, \ \theta_2, 1 - (1 + \theta_1)e^{-\theta_1} - \theta_2\}' \\ &(0 < \theta_1 < +\infty, \ 0 < \theta_2 < 1, \ 0 < 1 - (1 + \theta_1)e^{-\theta_1} - \theta_2 < 1) \end{aligned} \qquad (4.6)$$

is employed as a candidate model with the saturated one, where $\theta_{01} = 1.5$ and $\theta_{02} = \theta_{01}^2 \exp(-\theta_{01})/2$ are used for population values of θ_1 and θ_2, respectively. That is, in this set of models, we have three models.

For the $\mathrm{M}\phi\mathrm{E}$, we use three estimation methods. The first is to match $\phi_1(\cdot)$ to $\phi_2(\cdot)$. That is, e.g., when $\lambda = 0.5$ is used to have the $\mathrm{M}\phi\mathrm{E}$, $\lambda = 0.5$ is also used to form a power divergence statistic. The second method is that the MLEs are always used irrespective of the types of $\phi_2(\cdot)$ to have power divergences. This is employed due to the popularity of the MLE. The third method is similar to the second though the $\mathrm{M}\phi\mathrm{E}$s using $\lambda = 1$ (Pearson's statistic) are always used in place of the MLEs. The third method is employed considering the good behavior in simulations for model selection using the first method.

In the first half of this section, simulations of model selection are performed while in the second half, simulations for biases of $\hat{C}_\phi$ are carried out. In the three sets of models, the three models in each of the second and third sets with four categories are labelled as Models 1 to 3, where the number indicates the number of independent parameters with Model 3 being the saturated model. In the second set of models with three categories, the two models are similarly labelled by Models 1 and 2, where Model 2 is the saturated one.

Table 1 shows the proportions of the models selected by the $\phi\mathrm{IC}$ in 10,000 replications with three samples sizes *n* = 50, 200 and 800 when the matching method of estimation of the parameters and evaluation by power divergences is used. The results using the $\mathrm{M}^*\phi\mathrm{IC}$ will be shown later. When sampling zero(es) or empty cell(s) occurred , the observation was discarded due to the associated difficulty in numerical computation. The values of *Z* (zeroes) show the numbers of these observations until 10,000 regular observations were obtained. When at least one non-convergent case for estimation of parameters occurred, the observation was deleted. The values of *NC* (no-convergence) indicate the numbers of these observations.

In Table 1, we find that the proportions of the correct (simplest) models selected by the $\phi\mathrm{IC}$ are approximately 80 to 85%. While the differences of these proportions over different λ's are relatively small, the proportions using λ = 0.5 to 2 are larger than those using $\lambda = -2$ to -0.5. It is of interest to find that the proportion of the correct models when *n* = 800 are not necessarily the largest among those when *n* = 50, 200 and 800. This corresponds to the known property that the proportions of correct models selected by the AIC are not consistent.

Table 2 gives similar results for correct models when the MLEs (λ = 0) and estimators using λ =1 (Pearson's statistic) are always used. The results seem comparable to those in Table 1. Table 3 shows results when the $\mathrm{M}^*\phi\mathrm{IC}$ are used. Note that in a $\mathrm{M}^*\phi\mathrm{IC}$, $\sum_{k=1}^{K} 1/\pi_{0k}$ in b_Δ in (3.2) is replaced by its sample counterpart i.e., $\sum_{k=1}^{K} 1/p_k$. The values in the table show added percentages of the correct models selected by the $\mathrm{M}^*\phi\mathrm{IC}$ s over those by the $\phi\mathrm{IC}$ s. For instance, the first entry 1.96 in the first column for these percentages corresponds to the proportion 0.8372 (not shown in the tables) using the $\mathrm{M}^*\phi\mathrm{IC}$ minus the proportion 0.8176 in Table 1. That is, 0.8372 − 0.8176 = 0.0196. All of the values in the table are non-negative with only one zero (MLE, 3-category truncated Poisson, *n* = 800, $\lambda = -1$). The effect of $\hat{b}_\Delta$ in the $\mathrm{M}^*\phi\mathrm{IC}$ decreases as *n* increases since the associated term added to $O_p(1)$ is $-n^{-1}\hat{b}_\Delta$.

From Tables 1 to 3, the $\phi_2(\cdot)$ s with λ = 0.5 to 2 seem to give relatively good results in model selection as long as the models and data are similar to those in the tables. Considering these results, in the second half of this section, $\phi_2(\cdot)$ s with λ = 0 (G^2), λ = 2/3 (the Cressie-Read statistic), λ = 1 (X^2 Pearson's statistic) and λ = 2 are used to show simulated biases. Tables 4 to 8 give the results (Tables 5 to 8 are shown in the supplement to this paper (Ogasawara, 2018)). The regular observations used in Tables 4 to 8 are the same as those in Tables 1 to 3.

In Tables 4 to 8, S.B. (simulated bias) is the mean of the simulated $\hat{C}_\phi$ s minus the simulated mean of R_ϕ s over 10,000 replications; A.B. (asymptotic bias) = $-2q$; H.A.B (higher-order asymptotic bias) = $-2q + n^{-1}b_\Delta$; S. b_Δ (simulated b_Δ) = $n \times$ (the simulated bias +2 q); and b_Δ is given by Theorem 1. In Tables 4 to 8, the b_Δ^* s are shown in parentheses, which are given by (3.3). Note that in the case of the saturated model $b_\Delta^* = b_\Delta$. For instance, in Model 3 of Table 4, $b_\Delta^* = b_\Delta$ are $-21, -31, -39$ and -75 when λ = 0, 2/3, 1 and 2, respectively, which are actually computed by Theorem 1 but are found to be equal to those by Theorem 2. For example, when the MLE is used for $\phi_2(\cdot)$, (3.1) gives $b_\Delta^* = b_\Delta = -\sum_{k=1}^{K} 1/\pi_{0k} + K = -1/0.6 - 2/0.15 - 1/0.1 + 4 = -21$ using $\phi_2^{(4)} = 2,\ \phi_2^{(3)} = -1$ and $\phi_2'' = 1$. In Table 4, the S.B.s are more similar to the H.A.B.s than the fixed A.B. i.e., $-2q$ over λ s. The S. b_Δ s are in a crude sense similar to the b_Δ s. It is also found that the b_Δ s are similar to the b_Δ^* s. In Tables 5 to 8, similar tendencies are found.

5. Discussion

It is natural to ask an appropriate value of λ to be used in the $\phi\mathrm{IC}$ s and $\mathrm{M}^*\phi\mathrm{IC}$ s when the power divergence is used. In the numerical illustrations, it was found that relatively large λ s for $\phi_2(\cdot)$ are appropriate. The familiar MLE giving the AIC does not show the best results. Considering good behaviors in other respects, the $\phi\mathrm{IC}$ and $\mathrm{M}^*\phi\mathrm{IC}$ with λ = 2/3 may be reasonable. The criteria with λ = 1 may also be reasonable considering good results in the numerical illustrations though they are limited ones. The $\phi_1(\cdot)$ s yielding the $\mathrm{M}\phi\mathrm{E}$ s do not necessarily have to match the $\phi_2(\cdot)$ s in the $\phi\mathrm{IC}$ and $\mathrm{M}^*\phi\mathrm{IC}$. Tables 2 using the MLEs, when relatively large λ s for $\phi_2(\cdot)$ are employed, show reasonable results. The MLEs of parameters are of practical use in that "since software for computing MLEs is readily available" (Cressie & Read, 2002a, p.1554). Recall also that the MLE of the single parameter in the model for the genetics of plants (Fisher, 1970) was algebraically available with no *NC*s (no-convergences).

While the main results of Theorems 1 and 2 in this paper are for the general ϕ-divergence, the numerical results are given by the power divergences since the latter subfamily includes important statistics used in practice. However, there are many ϕ-divergences other than the power divergences (see e.g., Pardo, 2006, p.6). The behavior of the $\phi\mathrm{IC}$ and $\mathrm{M}^*\phi\mathrm{IC}$ based on these ϕ-divergences is to be investigated as a remaining issue.

Appendix

A1. Proof of Theorem 1

We give two lemmas.

Lemma 1.

$$\frac{\partial^2 \hat{\pi}_k \phi_1(p_k/\hat{\pi}_k)}{\partial \hat{\pi}_k^2}\Big|_{\substack{\hat{\pi}_k=\pi_{0k}\\ p_k=\pi_{0k}}} = \frac{\phi_1''}{\pi_{0k}},\ \frac{\partial^3 \hat{\pi}_k \phi_1(p_k/\hat{\pi}_k)}{\partial \hat{\pi}_k^3}\Big|_{\substack{\hat{\pi}_k=\pi_{0k}\\ p_k=\pi_{0k}}} = -\frac{1}{\pi_{0k}^2}(\phi_1^{(3)}+3\phi_1''),$$

$$\frac{\partial^4 \hat{\pi}_k \phi_1(p_k/\hat{\pi}_k)}{\partial \hat{\pi}_k^4}\Big|_{\substack{\hat{\pi}_k=\pi_{0k}\\ p_k=\pi_{0k}}} = \frac{1}{\pi_{0k}^3}(\phi_1^{(4)}+8\phi_1^{(3)}+12\phi_1''),$$

$$\frac{\partial^2 \hat{\pi}_k \phi_1(p_k/\hat{\pi}_k)}{\partial \hat{\pi}_k \partial p_k}\Big|_{\substack{\hat{\pi}_k=\pi_{0k}\\ p_k=\pi_{0k}}} = -\frac{\phi_1''}{\pi_{0k}},\ \frac{\partial^3 \hat{\pi}_k \phi_1(p_k/\hat{\pi}_k)}{\partial \hat{\pi}_k \partial p_k^2}\Big|_{\substack{\hat{\pi}_k=\pi_{0k}\\ p_k=\pi_{0k}}} = -\frac{1}{\pi_{0k}^2}(\phi_1^{(3)}+\phi_1''), \tag{A.1}$$

$$\frac{\partial^3 \hat{\pi}_k \phi_1(p_k/\hat{\pi}_k)}{\partial \hat{\pi}_k^2 \partial p_k}\Big|_{\substack{\hat{\pi}_k=\pi_{0k}\\ p_k=\pi_{0k}}} = \frac{1}{\pi_{0k}^2}(\phi_1^{(3)}+2\phi_1''),$$

$$\frac{\partial^4 \hat{\pi}_k \phi_1(p_k/\hat{\pi}_k)}{\partial \hat{\pi}_k \partial p_k^3}\Big|_{\substack{\hat{\pi}_k=\pi_{0k}\\ p_k=\pi_{0k}}} = -\frac{1}{\pi_{0k}^3}(\phi_1^{(4)}+2\phi_1^{(3)}),$$

$$\frac{\partial^4 \hat{\pi}_k \phi_1(p_k/\hat{\pi}_k)}{\partial \hat{\pi}_k^2 \partial p_k^2}\Big|_{\substack{\hat{\pi}_k=\pi_{0k}\\ p_k=\pi_{0k}}} = \frac{1}{\pi_{0k}^3}(\phi_1^{(4)}+4\phi_1^{(3)}+2\phi_1''),$$

$$\frac{\partial^4 \hat{\pi}_k \phi_1(p_k/\hat{\pi}_k)}{\partial \hat{\pi}_k^3 \partial p_k}\Big|_{\substack{\hat{\pi}_k=\pi_{0k}\\ p_k=\pi_{0k}}} = -\frac{1}{\pi_{0k}^3}(\phi_1^{(4)}+6\phi_1^{(3)}+6\phi_1'')\ \ (k=1,...,K)$$

Proof. Using $\phi_1'=0$, the direct derivation gives the results. Q.E.D.

Lemma 2. *The partial derivatives of* $\hat{\boldsymbol{\theta}}$ *using* $\phi_1(\cdot)$ *with respect to* $\mathbf{p}$ *at* $\mathbf{p}=\boldsymbol{\pi}_0$ *with the assumption of their existence are given as follows.*

$$
\begin{aligned}
\frac{\partial^2 \boldsymbol{\theta}_0}{\partial \pi_{0i}\partial \pi_{0j}} &= -\left(\frac{\partial^2 D_{\phi 0}}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'}\right)^{-1}\left(\frac{\partial^3 D_{\phi 0}}{\partial \boldsymbol{\theta}_0 (\partial \boldsymbol{\theta}_0')^{<2>}}\frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0i}} \otimes \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0j}}\right. \\
&\qquad \left. + \sum_{(i,j)}^{2} \frac{\partial^3 D_{\phi 0}}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0' \partial \pi_{0i}}\frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0j}} + \delta_{ij}\frac{\partial^3 D_{\phi 0}}{\partial \boldsymbol{\theta}_0 \partial \pi_{0i}^2}\right) \\
&= -\frac{1}{\phi_1''}\mathbf{I}_0^{-1}\left[\sum_{a=1}^{K}\left\{-\frac{1}{\pi_{0a}^2}(\phi_1^{(3)} + 3\phi_1'')\frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0}\left(\frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0'}\right)^{<2>}\right.\right. \\
&\qquad \left. + \frac{\phi_1''}{\pi_{0a}}\left(\frac{\partial^2 \pi_{0a}}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'} \otimes \frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0'} + \frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0} \otimes \frac{\partial^2 \pi_{0a}}{(\partial \boldsymbol{\theta}_0')^{<2>}}\right)\right\}\frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0i}} \otimes \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0j}} \\
&+ \sum_{(i,j)}^{2}\left\{\frac{1}{\pi_{0i}^2}(\phi_1^{(3)} + 2\phi_1'')\frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0}\frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0'} - \frac{\phi_1''}{\pi_{0i}}\frac{\partial^2 \pi_{0i}}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'}\right\}\frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0j}} \\
&\left. - \frac{\delta_{ij}}{\pi_{0i}^2}(\phi_1^{(3)} + \phi_1'')\frac{\partial \pi_{0i}}{\partial \boldsymbol{\theta}_0}\right],
\end{aligned}
\tag{A.2}
$$

$$
\begin{aligned}
\frac{\partial^3 \boldsymbol{\theta}_0}{\partial \pi_{0i}\partial \pi_{0j}\partial \pi_{0k}} &= -\left(\frac{\partial^2 D_{\phi 0}}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'}\right)^{-1}\left\{\frac{\partial^4 D_{\phi 0}}{\partial \boldsymbol{\theta}_0 (\partial \boldsymbol{\theta}_0')^{<3>}}\frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0i}} \otimes \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0j}} \otimes \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0k}}\right. \\
&+ \sum_{(i,j,k)}^{3}\left(\frac{\partial^3 D_{\phi 0}}{\partial \boldsymbol{\theta}_0 (\partial \boldsymbol{\theta}_0')^{<2>}}\frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0i}} \otimes \frac{\partial^2 \boldsymbol{\theta}_0}{\partial \pi_{0j}\partial \pi_{0k}} + \frac{\partial^4 D_{\phi 0}}{\partial \boldsymbol{\theta}_0 (\partial \boldsymbol{\theta}_0')^{<2>}\partial \pi_{0i}}\frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0j}} \otimes \frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0k}}\right. \\
&\qquad \left.\left. + \frac{\partial^3 D_{\phi 0}}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0' \partial \pi_{0i}}\frac{\partial^2 \boldsymbol{\theta}_0}{\partial \pi_{0j}\partial \pi_{0k}} + \delta_{ij}\frac{\partial^4 D_{\phi 0}}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0' \partial \pi_{0i}^2}\frac{\partial \boldsymbol{\theta}_0}{\partial \pi_{0k}}\right) + \delta_{ijk}\frac{\partial^4 D_{\phi 0}}{\partial \boldsymbol{\theta}_0 \partial \pi_{0i}^3}\right\}
\end{aligned}
$$

$$
= -\frac{1}{\phi_1''}\mathbf{I}_0^{-1}\Big[_{(A)}\sum_{a=1}^{K}\Big[_{(B)}\Big[_{(C)}\ \frac{1}{\pi_{0a}^3}(\phi_1^{(4)}+8\phi_1^{(3)}+12\phi_1'')\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0}\left(\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0'}\right)^{<3>}
$$

$$
-\frac{1}{\pi_{0a}^2}(\phi_1^{(3)}+3\phi_1'')\left\{\frac{\partial^2\pi_{0a}}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0'}\otimes\left(\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0'}\right)^{<2>}+\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0}\frac{\partial^2\pi_{0a}}{(\partial\boldsymbol{\theta}_0')^{<2>}}\otimes\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0'}\right.
$$

$$
\left.+\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0}\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0'}\otimes\frac{\partial^2\pi_{0a}}{(\partial\boldsymbol{\theta}_0')^{<2>}}\right\}
$$

$$
+\frac{\phi_1''}{\pi_{0a}}\left(\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0}\otimes\frac{\partial^3\pi_{0a}}{(\partial\boldsymbol{\theta}_0')^{<3>}}+2\frac{\partial^2\pi_{0a}}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0'}\otimes\frac{\partial^2\pi_{0a}}{(\partial\boldsymbol{\theta}_0')^{<2>}}+\frac{\partial^3\pi_{0a}}{\partial\boldsymbol{\theta}_0(\partial\boldsymbol{\theta}_0')^{<2>}}\otimes\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0'}\right)\ \Big]_{(C)}
$$

$$
\times\frac{\partial\boldsymbol{\theta}_0}{\partial\pi_{0i}}\otimes\frac{\partial\boldsymbol{\theta}_0}{\partial\pi_{0j}}\otimes\frac{\partial\boldsymbol{\theta}_0}{\partial\pi_{0k}}
$$

$$
+\sum_{(i,j,k)}^{3}\left\{-\frac{1}{\pi_{0a}^2}(\phi_1^{(3)}+3\phi_1'')\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0}\left(\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0'}\right)^{<2>}\right.
$$

$$
\left.+\frac{\phi_1''}{\pi_{0a}}\left(\frac{\partial^2\pi_{0a}}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0'}\otimes\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0'}+\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0}\frac{\partial^2\pi_{0a}}{(\partial\boldsymbol{\theta}_0')^{<2>}}\right)\right\}\frac{\partial\boldsymbol{\theta}_0}{\partial\pi_{0i}}\otimes\frac{\partial^2\boldsymbol{\theta}_0}{\partial\pi_{0j}\partial\pi_{0k}}\ \Big]_{(B)}
$$

$$
+\sum_{(i,j,k)}^{3}\Big[_{(D)}\ \left\{-\frac{1}{\pi_{0i}^3}(\phi_1^{(4)}+6\phi_1^{(3)}+6\phi_1'')\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0}\left(\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0'}\right)^{<2>}\right.
$$

$$
+\frac{\phi_1^{(3)}}{\pi_{0i}^2}\left(\frac{\partial^2\pi_{0i}}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0'}\otimes\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0'}+\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0}\frac{\partial^2\pi_{0i}}{(\partial\boldsymbol{\theta}_0')^{<2>}}\right)
$$

$$
\left.-\frac{\phi_1''}{\pi_{0i}}\frac{\partial^3\pi_{0i}}{\partial\boldsymbol{\theta}_0'(\partial\boldsymbol{\theta}_0')^{<2>}}+2\frac{\phi_1''}{\pi_{0i}^2}\left(\frac{\partial^2\pi_{0i}}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0'}\otimes\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0'}+\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0}\frac{\partial^2\pi_{0i}}{(\partial\boldsymbol{\theta}_0')^{<2>}}\right)\right\}\frac{\partial\boldsymbol{\theta}_0}{\partial\pi_{0j}}\otimes\frac{\partial\boldsymbol{\theta}_0}{\partial\pi_{0k}}
$$

$$
+\left\{\frac{1}{\pi_{0i}^2}(\phi_1^{(3)}+2\phi_1'')\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0}\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0'}-\frac{\phi_1''}{\pi_{0i}}\frac{\partial^2\pi_{0i}}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0'}\right\}\frac{\partial^2\boldsymbol{\theta}_0}{\partial\pi_{0j}\partial\pi_{0k}}
$$

$$
+\delta_{ij}\left\{\frac{1}{\pi_{0i}^3}(\phi_1^{(4)}+4\phi_1^{(3)}+2\phi_1'')\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0}\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0'}-\frac{1}{\pi_{0i}^2}(\phi_1^{(3)}+\phi_1'')\frac{\partial^2\pi_{0i}}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0'}\right\}\frac{\partial\boldsymbol{\theta}_0}{\partial\pi_{0k}}\ \Big]_{(D)}
$$

$$
-\delta_{ijk}\frac{1}{\pi_{0i}^3}(\phi_1^{(4)}+2\phi_1^{(3)})\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0}\ \Big]_{(A)}\quad (i,j,k=1,\dots,K),
$$

where D_ϕ *in place of* C_ϕ *is used since minimizing* D_ϕ *is equivalent to that of* C_ϕ*;* $\mathbf{I}_0$ *is the Fisher information matrix per observation;* $\sum_{(i,j)}^{2}(\cdot)$ *is the sum of two terms exchanging i and j with* $\sum_{(i,j,k)}^{3}(\cdot)$ *defined similarly; and* δ_{ij} *is the Kronecker delta with* $\delta_{ijk} = \delta_{ij}\delta_{jk}$.

Proof. The formulas of the partial derivatives in implicit functions (see e.g., Ogasawara, 2009, Subsection A.1) are used. Note, however, that while in Ogasawara (2009), the log-likelihood is used, which is linear in terms of **p** (temporally $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}(\mathbf{p})$ is not seen as a function of **p**), D_ϕ in Lemma 2 is generally non-linear in terms of **p**. Q.E.D.

For the proof of Theorem 1, using the assumption in the theorem, we have

$$
\begin{aligned}
&\hat{\pi}_k - \pi_{0k} \\
&= \frac{\partial \pi_{0k}}{\partial \boldsymbol{\theta}_0{}'} \frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0} (\mathbf{p} - \boldsymbol{\pi}_0) + \frac{1}{2} \left\{ \frac{\partial^2 \pi_{0k}}{(\partial \boldsymbol{\theta}_0{}')^{<2>}} \left(\frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0{}'} \right)^{<2>} + \frac{\partial \pi_{0k}}{\partial \boldsymbol{\theta}_0{}'} \frac{\partial^2 \boldsymbol{\theta}_0}{(\partial \boldsymbol{\pi}_0{}')^{<2>}} \right\} (\mathbf{p} - \boldsymbol{\pi}_0)^{<2>} \\
&+ \frac{1}{6} \left\{ \frac{\partial^3 \pi_{0k}}{(\partial \boldsymbol{\theta}_0{}')^{<3>}} \left(\frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0{}'} \right)^{<3>} + 3 \frac{\partial^2 \pi_{0k}}{(\partial \boldsymbol{\theta}_0{}')^{<2>}} \left(\frac{\partial^2 \boldsymbol{\theta}_0}{(\partial \boldsymbol{\pi}_0{}')^{<2>}} \otimes \frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0{}'} \right) + \frac{\partial \pi_{0k}}{\partial \boldsymbol{\theta}_0{}'} \frac{\partial^3 \boldsymbol{\theta}_0}{(\partial \boldsymbol{\pi}_0{}')^{<3>}} \right\} \\
&\quad \times (\mathbf{p} - \boldsymbol{\pi}_0)^{<3>} \\
&+ O_p(n^{-2}) \ (k = 1, ..., K).
\end{aligned}
\tag{A.3}
$$

Noting the common term in the expansions of $\mathrm{E}\{\hat{\pi}_k \phi_2(p_k / \hat{\pi}_k)\}$ and $\mathrm{EE}^*\{\hat{\pi}_k \phi_2(p_k^* / \hat{\pi}_k)\}$, we obtain

$$
\mathrm{E}(\hat{C}_\phi) - R_\phi = \frac{2n}{\phi_2{}''} \mathrm{E} \left[\sum_{k=1}^{K} [\, \hat{\pi}_k \phi_2(p_k / \hat{\pi}_k) - \mathrm{E}^*\{\phi_2(p_k^* / \hat{\pi}_k)\} \,] \right]_{(\mathrm{A})}^{(\mathrm{A})} \tag{A.4}
$$

$$
\begin{aligned}
&= \frac{2n}{\phi_2''} \sum_{k=1}^{K} \Bigg[_{(A)} \ \frac{1}{2} \sum_{\substack{i \neq 0 \\ i \neq 2}}^{2} \binom{2}{i} \frac{\partial^2 \hat{\pi}_k \phi_2 (p_k / \hat{\pi}_k)}{\partial \hat{\pi}_k^i \partial p_k^{2-i}} \Big|_{\substack{\hat{\pi}_k = \pi_{0k} \\ p_k = \pi_{0k}}} \ n \mathrm{E} \{ (\hat{\pi}_k - \pi_{0k})^i (p_k - \pi_{0k})^{2-i} \}_{\to O(n^{-1}) + O(n^{-2})} \\
&+ \frac{n^{-1}}{6} \sum_{\substack{i \neq 0 \\ i \neq 3}}^{3} \binom{3}{i} \frac{\partial^3 \hat{\pi}_k \phi_2 (p_k / \hat{\pi}_k)}{\partial \hat{\pi}_k^i \partial p_k^{3-i}} \Big|_{\substack{\hat{\pi}_k = \pi_{0k} \\ p_k = \pi_{0k}}} \ n^2 \mathrm{E} \{ (\hat{\pi}_k - \pi_{0k})^i (p_k - \pi_{0k})^{3-i} \}_{\to O(n^{-2})} \\
&+ \frac{n^{-1}}{24} \sum_{\substack{i \neq 0 \\ i \neq 4}}^{4} \binom{4}{i} \frac{\partial^4 \hat{\pi}_k \phi_2 (p_k / \hat{\pi}_k)}{\partial \hat{\pi}_k^i \partial p_k^{4-i}} \Big|_{\substack{\hat{\pi}_k = \pi_{0k} \\ p_k = \pi_{0k}}} \ n^2 \mathrm{E} \{ (\hat{\pi}_k - \pi_{0k})^i (p_k - \pi_{0k})^{4-i} \}_{\to O(n^{-2})} \ \Bigg]_{(A)} + O(n^{-2}),
\end{aligned}
$$

where $\mathrm{E}\{\cdot\}_{\to O(n^{-1}) + O(n^{-2})}$ is the expectation up to order $O(n^{-2})$ with the main term being of order $O(n^{-1})$ for clarity..

Using Lemma 1 and (A.3) in (A.4) gives Theorem 1, where the expressions of $\partial \boldsymbol{\theta}_0 / \partial \boldsymbol{\pi}_0'$ and $\partial^j \boldsymbol{\theta}_0 / (\partial \boldsymbol{\pi}_0')^{<j>} \, (j = 2, 3)$ are given by (2.13) and Lemma 2, respectively. The result of $b = -2q$ is given as follows. From (2.13), it follows that

$$
\begin{aligned}
b &= \frac{2}{\phi_2''} \left[\sum_{k=1}^{K} - \frac{\phi_2''}{\pi_{0k}} \frac{\partial \pi_{0k}}{\partial \boldsymbol{\theta}_0'} \frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0'} \mathrm{E} \{ (\mathbf{p} - \boldsymbol{\pi}_0)(p_k - \pi_{0k}) \} \right] \\
&= -2 \sum_{k=1}^{K} \frac{1}{\pi_{0k}} \frac{\partial \pi_{0k}}{\partial \boldsymbol{\theta}_0'} \mathbf{I}_0^{-1} \frac{\partial \boldsymbol{\pi}_0'}{\partial \boldsymbol{\theta}_0} \mathrm{diag}^{-1}(\boldsymbol{\pi}_0) \{ \mathrm{diag}(\boldsymbol{\pi}_0) - \boldsymbol{\pi}_0 \boldsymbol{\pi}_0' \}_{\cdot k} \\
&= -2 \sum_{k=1}^{K} \frac{1}{\pi_{0k}} \frac{\partial \pi_{0k}}{\partial \boldsymbol{\theta}_0'} \mathbf{I}_0^{-1} \frac{\partial \boldsymbol{\pi}_0'}{\partial \boldsymbol{\theta}_0} (\mathbf{I}_{(K)})_{\cdot k} \\
&= -2 \mathrm{tr} \left\{ \mathbf{I}_0^{-1} \frac{\partial \boldsymbol{\pi}_0'}{\partial \boldsymbol{\theta}_0} \mathrm{diag}^{-1}(\boldsymbol{\pi}_0) \frac{\partial \boldsymbol{\pi}_0}{\partial \boldsymbol{\theta}_0'} \right\} \\
&= -2q,
\end{aligned}
\qquad \text{(A.5)}
$$

where $\mathbf{I}_0 = \frac{\partial \boldsymbol{\pi}_0'}{\partial \boldsymbol{\theta}_0} \mathrm{diag}^{-1}(\boldsymbol{\pi}_0) \frac{\partial \boldsymbol{\pi}_0}{\partial \boldsymbol{\theta}_0'}$ with $\mathrm{diag}^{-1}(\boldsymbol{\pi}_0) = \{ \mathrm{diag}(\boldsymbol{\pi}_0) \}^{-1}$; $\mathbf{I}_{(K)}$ is the $K \times K$ identity matrix; and $(\cdot)_{\cdot k}$ is the *k*-th column of a matrix.

For b_Δ in (2.12), we require the following

$$\kappa_3(a,b,c) \equiv n^2\{(p_a - \pi_{0a})(p_b - \pi_{0b})(p_c - \pi_{0c})\}$$
$$= \delta_{abc}\pi_{0a} - \sum_{(a,b,c)}^{3} \delta_{ab}\pi_{0a}\pi_{0c} + 2\pi_{0a}\pi_{0b}\pi_{0c} \qquad \text{(A.6)}$$

(Stuart & Ort, 1994, Equation (7.18)) and

$$\begin{aligned} m_4(a,b,c,d) &= n^2\{(p_a - \pi_{0a})(p_b - \pi_{0b})(p_c - \pi_{0c})(p_d - \pi_{0d})\}_{\to O(n^{-2})} \\ &= (\delta_{ab}\pi_{0a} - \pi_{0a}\pi_{0b})(\delta_{cd}\pi_{0c} - \pi_{0c}\pi_{0d}) \\ &\quad + (\delta_{ac}\pi_{0a} - \pi_{0a}\pi_{0c})(\delta_{bd}\pi_{0b} - \pi_{0b}\pi_{0d}) \\ &\quad + (\delta_{ad}\pi_{0a} - \pi_{0a}\pi_{0d})(\delta c\pi_{0b} - \pi_{0b}\pi_{0c}) \end{aligned} \qquad \text{(A.7)}$$
$$(a,b,c,d = 1,...,K).$$

A.2 Proof of Theorem 2

In the special case of the saturated model, we employ the reparametrization $\boldsymbol{\theta} = \boldsymbol{\pi}$ and $\hat{\boldsymbol{\theta}} = \mathbf{p}$ with $q = K$ dependent parameters rather than $\hat{\boldsymbol{\theta}} = \mathbf{p}_{(K-1)}$ with $K-1$ independent parameters. The restriction of $\mathbf{1}_{(K)}'\hat{\boldsymbol{\pi}} = 1$ is automatically satisfied, where $\mathbf{1}_{(K)}$ is the $K \times 1$ vector of 1's, since $\hat{\boldsymbol{\pi}} = \mathbf{p}$ with $\mathbf{1}_{(K)}'\mathbf{p} = 1$

Note that under the parametrization with $\hat{\boldsymbol{\theta}} = \mathbf{p}_{(K-1)}$ for the saturated model, the $K \times K$ matrix $\frac{\partial \boldsymbol{\pi}_0}{\partial \boldsymbol{\theta}_0'} \frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\pi}_0'}$ becomes singular, whose rank is $K-1$ and $\partial^j \boldsymbol{\theta}_0 / (\partial \boldsymbol{\pi}_0')^{<j>} (j = 2, 3)$ do not vanish. For the results under $\hat{\boldsymbol{\theta}} = \mathbf{p}_{(K-1)}$, see Ogasawara (2018).

Let

$$\kappa_3(k) \equiv n^2\{(p_k - \pi_{0k})^3\} = \pi_{0k}(1-\pi_{0k})(1-2\pi_{0k}) \quad \text{and} \qquad \text{(A.8)}$$
$$m_4(k) \equiv n^2\{(p_k - \pi_{0k})^4\}_{\to O(n^{-2})} = 3\pi_{0k}^2(1-\pi_{0k})^2 \ (k = 1,...,K).$$

Then, b_Δ in Theorem 1 becomes

$$
\begin{aligned}
b_\Delta = \frac{2}{\phi_2''}\sum_{k=1}^{K}\Bigg\{ & -\frac{1}{2\pi_{0k}^2}(\phi_2^{(3)} + \phi_2'')\kappa_3(k) + \frac{1}{2\pi_{0k}^2}(\phi_2^{(3)} + 2\phi_2'')\kappa_3(k) \\
& -\frac{1}{6\pi_{0k}^3}(\phi_2^{(4)} + 2\phi_2^{(3)})m_4(k) + \frac{1}{4\pi_{0k}^3}(\phi_2^{(4)} + 4\phi_2^{(3)} + 2\phi_2'')m_4(k) \\
& -\frac{1}{6\pi_{0k}^3}(\phi_2^{(4)} + 6\phi_2^{(3)} + 6\phi_2'')m_4(k)\Bigg\}
\end{aligned} \tag{A.9}
$$

$$
\begin{aligned}
&= \frac{1}{\phi_2''}\sum_{k=1}^{K}\Bigg[\{-(\phi_2^{(3)} + \phi_2'') + \phi_2^{(3)} + 2\phi_2''\}\left(\frac{1}{\pi_{0k}} - 3 + 2\pi_{0k}\right) \\
&\quad + \{-(\phi_2^{(4)} + 2\phi_2^{(3)}) + (3/2)(\phi_2^{(4)} + 4\phi_2^{(3)} + 2\phi_2'') - (\phi_2^{(4)} + 6\phi_2^{(3)} + 6\phi_2'')\} \\
&\qquad \times\left(\frac{1}{\pi_{0k}} - 2 + \pi_{0k}\right)\Bigg] \\
&= \frac{1}{\phi_2''}\left\{\sum_{k=1}^{K}\phi_2''\left(\frac{1}{\pi_{0k}} - 3 + 2\pi_{0k}\right) - \left(\frac{\phi_2^{(4)}}{2} + 2\phi_2^{(3)} + 3\phi_2''\right)\left(\frac{1}{\pi_{0k}} - 2 + \pi_{0k}\right)\right\},
\end{aligned}
$$

which gives (3.1).

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (Eds.), *Proceedings of the 2nd international symposium on information theory* (pp. 267-281). Budapest: Académiai Kiado.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317-332.

Ali, S. M., & Silvey, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society, B, 28*, 131-142.

Allen, D. M. (1971). Mean square error of prediction as a criterion for selecting variables. *Technometrics, 13*, 469-475.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice.* Cambridge, MA: MIT Press.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research, 24*, 445-455.

Burnham, K. P., & Anderson, D. R. (2010). *Model selection and multimodel inference – A* *practical information-theoretic approach* (2nd ed.). New York: Springer.

Cressie, N., & Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, B, 46*, 440-464.

Cressie, N., & Pardo, L. (2000). Minimum ϕ-divergence estimator and hierarchical testing in loglinear models. *Statistica Sinica, 10*, 867-884.

Cressie, N., & Pardo, L. (2002a). Phi-divergence statistic. In A. H. El-Shaarawi & W. W. Piegorsch (Eds.), *Encyclopedia of environmetrics* (*Vol. 3*, pp 1551-1555). New York: Wiley.

Cressie, N., & Pardo, L. (2002b). Model checking in loglinear models using ϕ-divergences and MLEs. *Journal of Statistical Planning and Inference, 103*, 437-453.

Cressie, N., Pardo, L., & Pardo, M. C. (2003). Size and power considerations for testing loglinear models using ϕ-divergence test statistics. *Statistica Sinica, 13*, 555-570.

Csiszár, I. (1963). Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität on Markhoffschen Ketten. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences, Series A, 8*, 85-108.

Felipe, P., Miranda, P., & Pardo, L. (2015). Minimum ϕ-divergence estimation in constrained latent class models for binary data. *Psychometrika, 80*, 1020-1042.

Fisher, R. A. (1970). *Statistical methods for research workers* (14th ed.). Edinburgh: Oliver & Boyd.

Freeman, M. F., & Tukey, J. W. (1950). Transformations related to the angular and the square root. *The Annals of Mathematical Statistics, 21*, 607-611.

Fujikoshi, Y., & Satoh, K. (1997). Modified AIC and C_p in multivariate linear regression. *Biometrika, 84*, 707-716.

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika, 76*, 297-307.

Ichikawa, M. (1988). Empirical assessment of AIC procedure for model selection in factor analysis. *Behaviormetrika*, *No.24*, 33-40.

Kullback, S. (1985). Kullback information. In S. Kotz & N L. Johnson (Eds.), *Encyclopedia of statistical sciences Vol. 4* (pp. 421-425). New York: Wiley.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of*

Mathematical Statistics, 22, 79-86.

Linhart, H., & Zucchini, W. (1986). *Model selection*. New York: Wiley.

Mallows, C. L. (1973). Some comments on C_P. *Technometrics, 15*, 661-675.

Morales, D., Pardo, L., & Vajda, I. (1995). Asymptotic divergence of estimates of discrete distributions. *Journal of Statistical Planning and Inference, 48*, 347-369.

Neyman, J. (1949). Contribution to the theory of the χ^2 test. *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*, 239-273.

Ogasawara, H. (2009). Asymptotic cumulants of the parameter estimators in item response theory. *Computational Statistics, 24,* 313-331.

Ogasawara, H. (2016a). Bias correction of the Akaike information criterion in factor analysis. *Journal of Multivariate Analysis, 149*, 144-159.

Ogasawara, H. (2016b). Optimal information criteria minimizing their asymptotic mean square errors. *Sankhyā, B, 78,* 152-182.

Ogasawara, H. (2017). Expected predictive least squares for model selection in covariance structures. *Journal of Multivariate Analysis, 155,* 151-164.

Ogasawara, H. (2018). An expository supplement to the paper "A family of the information criteria using the phi-divergence for categorical data". To appear in *Economic Review* (*Otaru University of Commerce*).

Pardo, L. (2006). *Statistical inference based on divergence measures*. Boca Raton, FL: Chapman & Hall.

Pardo, J. A., Pardo, L., & Pardo, M. C. (2005). Minimum ϕ-divergence estimator in logistic regression models. *Statistical papers, 47*, 91-108.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophy Magazine Series (5), 50*, 157-172.

Read, T. R. C., & Cressie, N. (1988). *Goodness-of-fit statistics for discrete multivariate data*. New York: Springer.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461-464.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal* *of the Royal Statistical Society, B, 36*, 111-147.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society, B, 39*, 44-47.

Stuart, A., & Ord, J. K. (1994). *Kendall's advanced theory of statistics: Distribution theory* (6th ed., Vol.1). London: Arnold.

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

Takeuchi, K. (1976). Distributions of information statistics and criteria of the goodness of models. *Mathematical Science, 153*, 12-18. (in Japanese)

Yanagihara, H., Himeno, T., & Yuan, K.-H. (2010). GLS discrepancy based information criteria for selecting covariance stricture models. *Behaviormetrika, 37*, 71-86.

Yanagihara, H., Yuan K.-H., Fujisawa, H., & Hayashi, K. (2013). A class of model selection criteria based on cross-validation method. *Hiroshima Mathematical Journal, 43*, 149-177.

Table 1. 10,000 times the proportions of models selected by the ϕIC s (the number of replications = 10,000)

	n = 50			n = 200			n = 800		
The ϕ_2 in a ϕIC matches the ϕ_1 for estimation of parameters.									
The genetics of plants (Fisher, 1970; 4 categories)									
	(Z = 135, NC = 1,383)			(Z = 0, NC = 144)			(Z = 0, NC = 0)		
Model:	1	2	3	1	2	3	1	2	3
λ = 0 (G^2)	8176	1061	763	8022	1180	798	7853	1342	805
λ = -1 (GM^2)	8139	1066	795	8009	1178	813	7853	1344	803
λ = -2 (Neyman)	8078	1147	775	7968	1212	820	7837	1358	805
λ = -0.5 (T^2)	8166	1062	772	8022	1168	810	7860	1340	800
λ = 0.5	8194	1069	737	8014	1189	797	7860	1343	797
λ = 2/3 (C-R)	8203	1112	685	8023	1193	784	7859	1348	793
λ = 1 (X^2)	8205	1122	673	8046	1189	765	7863	1344	793
λ = 2	**8241**	1150	609	**8069**	1183	748	**7872**	1335	793
3-category truncated Poisson variate (Bishop et al., 1975, p.503)									
	(Z = 0, NC = 2)			(Z = 0, NC = 0)			(Z = 0, NC = 0)		
Model:	1	2		1	2		1	2	
λ = 0 (G^2)	8399	1601		8537	1463		8394	1606	
λ = -1 (GM^2)	8389	1611		8484	1516		8374	1626	
λ = -2 (Neyman)	8389	1611		8452	1548		8367	1633	
λ = -0.5 (T^2)	8398	1602		8501	1499		8379	1621	
λ = 0.5	**8402**	1598		8572	1428		8411	1589	
λ = 2/3 (C-R)	**8402**	1598		**8573**	1427		8417	1583	
λ = 1 (X^2)	8356	1644		**8573**	1427		8434	1566	
λ = 2	8306	1694		8549	1451		**8452**	1548	

Note. n = the number of observations, Z = the number of deleted cases with zero frequenc(ies), NC = the number of deleted cases due to non-convergence, G^2 = the log-likelihood ratio statistic, GM^2 = the modified log-likelihood ratio statistic, Neyman = Neyman's statistic, T^2 = the Freeman-Tukey statistic, C-R = the Cressie-Read statistic, X^2 = Pearson's statistic. The number for model identification is the number of independent parameters. The boldfaced value indicates the largest proportion in the correct (simplest) model among 8 power divergences.

Table 1. (continued)

	n = 50			n = 200			n = 800		
The ϕ_2 in a ϕIC matches the ϕ_1 for estimation of parameters.									
4-category truncated Poisson variate									
	(Z = 0, NC = 334)			(Z = 0, NC = 0)			(Z = 0, NC = 0)		
Model:	1	2	3	1	2	3	1	2	3
λ = 0 (G^2)	7900	1287	813	7877	1318	805	7851	1367	782
λ = -1 (GM^2)	7786	1310	904	7843	1338	819	7836	1383	781
λ = -2 (Neyman)	7696	1365	939	7817	1353	830	7840	1377	783
λ = -0.5 (T^2)	7813	1324	863	7859	1329	812	7846	1373	781
λ = 0.5	7912	1293	795	7880	1319	801	7854	1364	782
λ = 2/3 (C-R)	7934	1276	790	7884	1316	800	7855	1363	782
λ = 1 (X^2)	7955	1265	780	7891	1311	798	**7858**	1361	781
λ = 2	**7962**	1278	760	**7905**	1309	786	7853	1371	776

Table 2. 10,000 times the proportions of the correct (simplest) models selected by the ϕICs (the number of replications = 10,000)

	The genetics of plants (Fisher, 1970; 4 categories)			3-category truncated Poisson (Bishop et al., 1975)			4-category truncated Poisson		
n:	50	200	800	50	200	800	50	200	800
The MLEs (λ = 0) are used for all power divergences.									
$\lambda = 0$ (G^2)	7753	7806	7867	8481	8488	8416	7875	7909	7886
$\lambda = -1$ (GM^2)	7491	7769	7851	8475	8449	8399	7772	7882	7884
$\lambda = -2$ (Neyman)	7333	7698	7846	8475	8413	8383	7675	7834	7875
$\lambda = -0.5$ (T^2)	7589	7787	7864	8475	8451	8406	7810	7888	7882
$\lambda = 0.5$	7798	7808	**7874**	**8487**	8548	8429	7899	7914	7894
$\lambda = 2/3$ (C-R)	7794	7814	7872	**8487**	8548	8437	7907	7914	7895
$\lambda = 1$ (X^2)	7833	**7829**	7867	8454	**8552**	8448	7932	7916	**7899**
$\lambda = 2$	**7874**	7816	7860	8409	8531	**8464**	**7948**	**7919**	7892
The parameter estimators by $\lambda = 1$ (X^2) are used for all power divergences.									
$\lambda = 0$ (G^2)	8077	7912	7861	8423	8450	8386	7863	7868	7803
$\lambda = -1$ (GM^2)	7948	7865	7858	8415	8406	8369	7729	7840	7804
$\lambda = -2$ (Neyman)	7751	7797	7852	8411	8370	8355	7589	7813	7784
$\lambda = -0.5$ (T^2)	8037	7904	7854	8415	8412	8374	7785	7850	7804
$\lambda = 0.5$	8159	7932	7865	**8436**	8493	8398	7913	7881	7810
$\lambda = 2/3$ (C-R)	8168	7937	**7869**	**8436**	**8495**	8402	7932	7885	7808
$\lambda = 1$ (X^2)	8180	7948	7863	8390	**8495**	8416	7947	7891	7811
$\lambda = 2$	**8248**	**7984**	7858	8341	8473	**8433**	**7975**	**7897**	7**813**

Note. n = the number of observations, G^2 = the log-likelihood ratio statistic, GM^2 = the modified log-likelihood ratio statistic, Neyman = Neyman's statistic, T^2 = the Freeman-Tukey statistic, C-R = the Cressie-Read statistic, X^2 = Pearson's statistic. The boldfaced value indicates the largest proportion in the correct (simplest) model among 8 power divergences.

Table 3. Added percentages of the correct (simplest) models selected by the $\mathrm{M}^*\phi\mathrm{IC}$ s over those by the $\phi\mathrm{IC}$ s (the number of replications = 10,000)

	The genetics of plants (Fisher, 1970; 4 categories)			3-category truncated Poisson (Bishop et al., 1975)			4-category truncated Poisson		
n:	50	200	800	50	200	800	50	200	800
The ϕ_2 in a $\phi\mathrm{IC}$ matches the ϕ_1 for estimation of parameters.									
$\lambda = 0$ (G^2)	1.96	.57	.21	.35	.59	.10	1.23	.27	.10
$\lambda = -1$ (GM^2)	1.90	.47	.16	.42	.12	.02	1.63	.41	.09
$\lambda = -2$ (Neyman)	4.18	.99	.26	.65	.23	.02	2.26	.50	.12
$\lambda = -0.5$ (T^2)	1.81	.45	.08	.33	.36	.08	1.70	.39	.08
$\lambda = 0.5$	2.96	1.05	.10	.33	.23	.07	2.05	.44	.10
$\lambda = 2/3$ (C-R)	2.97	1.10	.14	.33	.22	.19	2.09	.47	.10
$\lambda = 1$ (X^2)	4.09	1.08	.23	.91	.22	.23	2.20	.56	.15
$\lambda = 2$	6.11	1.79	.44	1.99	.26	.03	4.66	1.10	.26
The MLEs ($\lambda = 0$) are used for all power divergences.									
$\lambda = 0$ (G^2)	2.30	.38	.08	.41	.77	.08	1.36	.28	.14
$\lambda = -1$ (GM^2)	3.02	.43	.10	.41	.13	.00	1.55	.21	.08
$\lambda = -2$ (Neyman)	4.23	1.23	.24	.61	.27	.06	1.96	.63	.18
$\lambda = -0.5$ (T^2)	3.07	.57	.06	.41	.42	.06	1.54	.33	.08
$\lambda = 0.5$	3.30	.81	.11	.35	.20	.08	1.99	.47	.06
$\lambda = 2/3$ (C-R)	3.72	.79	.20	.36	.20	.11	2.22	.58	.10
$\lambda = 1$ (X^2)	5.30	.88	.34	.80	.16	.18	2.39	.60	.12
$\lambda = 2$	8.56	2.16	.49	1.71	.24	.20	4.11	.99	.35

Note. n = the number of observations, G^2 = the log-likelihood ratio statistic, GM^2 = the modified log-likelihood ratio statistic, Neyman = Neyman's statistic, T^2 = the Freeman-Tukey statistic, C-R = the Cressie-Read statistic, X^2 = Pearson's statistic.

Table 3. (continued)

	The genetics of plants (Fisher, 1970; 4 categories)			3-category truncated Poisson (Bishop et al., 1975)			4-category truncated Poisson		
n:	50	200	800	50	200	800	50	200	800
The parameter estimators by $\lambda = 1$ (X^2) are used for all power divergences.									
$\lambda = 0$ (G^2)	2.84	.56	.09	.53	.58	.07	1.27	.35	.12
$\lambda = -1$ (GM^2)	2.36	.64	.16	.48	.08	.01	1.48	.27	.10
$\lambda = -2$ (Neyman)	4.15	1.01	.20	.73	.33	.10	2.30	.50	.20
$\lambda = -0.5$ (T^2)	2.17	.38	.11	.48	.43	.06	1.63	.40	.10
$\lambda = 0.5$	3.19	.74	.19	.41	.17	.04	1.92	.31	.05
$\lambda = 2/3$ (C-R)	3.33	.80	.16	.43	.15	.14	2.00	.38	.10
$\lambda = 1$ (X^2)	4.43	1.00	.30	1.11	.15	.18	2.28	.45	.09
$\lambda = 2$	6.32	1.69	.54	2.25	.26	.01	4.53	1.01	.19

Table 4. Simulated and asymptotic biases of the power divergences (the number of replications = 10,000)

The ϕ_2 in a ϕIC matches the ϕ_1 for estimation of parameters.

The genetics of plants (Fisher, 1970; 4 categories)

Model 1	S.B.	A.B.	H.A.B.	S. b_Δ	b_Δ	S.B.	A.B.	H.A.B.	S. b_Δ	b_Δ
			n = 50					n = 200		
λ = 0 (G^2)	-2.10	-2	-2.13	-4.8	-6.6	-2.05	-2	-2.03	-10.1	-6.6
λ = 2/3 (C-R)	-2.14	-2	-2.14	-6.8	-7.1	-2.06	-2	-2.04	-12.7	-7.1
λ = 1 (X^2)	-2.22	-2	-2.19	-10.8	-9.3	-2.09	-2	-2.05	-17.2	-9.3
λ = 2	-2.72	-2	-2.48	-36.2	-24.0	-2.21	-2	-2.12	-42.9	-24.0
			n = 800			(b_Δ^*)				
λ = 0 (G^2)	-2.01	-2	-2.01	-5.4	-6.6	(-7.0)				
λ = 2/3 (C-R)	-2.01	-2	-2.01	-10.8	-7.1	(-10.3)				
λ = 1 (X^2)	-2.02	-2	-2.01	-16.7	-9.3	(-13.0)				
λ = 2	-2.06	-2	-2.03	-47.1	-24.0	(-25.0)				
Model 2			n = 50					n = 200		
λ = 0 (G^2)	-3.79	-4	-4.24	10.5	-12.0	-3.98	-4	-4.06	4.8	-12.0
λ = 2/3 (C-R)	-3.82	-4	-4.34	8.8	-17.2	-4.00	-4	-4.09	-.3	-17.2
λ = 1 (X^2)	-3.92	-4	-4.46	4.2	-23.0	-4.04	-4	-4.12	-7.4	-23.0
λ = 2	-4.50	-4	-5.06	-25.1	-53.0	-4.24	-4	-4.27	-47.7	-53.0
			n = 800			(b_Δ^*)				
λ = 0 (G^2)	-3.99	-4	-4.02	4.5	-12.0	(-14.0)				
λ = 2/3 (C-R)	-4.01	-4	-4.02	-6.9	-17.2	(-20.7)				
λ = 1 (X^2)	-4.02	-4	-4.03	-17.7	-23.0	(-26.0)				
λ = 2	-4.09	-4	-4.07	-70.2	-53.0	(-50.0)				

Note. n = the number of observations, S.B. = simulated bias, A.B. = asymptotic bias = $-2q$, H.A.B. = $b + n^{-1}b_\Delta = -2q + n^{-1}b_\Delta$, S. b_Δ = simulated $b_\Delta = n(\text{S.B.} + 2q)$, G^2 = the log-likelihood ratio statistic, C-R = the Cressie-Read statistic, X^2 = Pearson's statistic. The number for model identification is the number of independent parameters.

Table 4. (continued)

The ϕ_2 in a ϕIC matches the ϕ_1 for estimation of parameters.

The genetics of plants (Fisher, 1970; 4 categories)

Model 3	S.B.	A.B.	H.A.B.	S. b_Δ	b_Δ	S.B.	A.B.	H.A.B.	S. b_Δ	b_Δ
			n = 50					n = 200		
$\lambda = 0$ (G^2)	-6.00	-6	-6.42	.1	-21.0	-6.03	-6	-6.11	-6.1	-21.0
$\lambda = 2/3$ (C-R)	-6.35	-6	-6.62	-17.5	-31.0	-6.11	-6	-6.16	-21.1	-31.0
$\lambda = 1$ (X^2)	-6.70	-6	-6.78	-35.2	-39.0	-6.18	-6	-6.20	-35.1	-39.0
$\lambda = 2$	-9.01	-6	-7.50	-150.5	-75.0	-6.53	-6	-6.38	-105.7	-75.0
			n = 800			($b^*_\Delta = b_\Delta$)				
$\lambda = 0$ (G^2)	-5.99	-6	-6.03	6.1	-21.0	(-21.0)				
$\lambda = 2/3$ (C-R)	-6.02	-6	-6.04	-13.6	-31.0	(-31.0)				
$\lambda = 1$ (X^2)	-6.04	-6	-6.05	-29.6	-39.0	(-39.0)				
$\lambda = 2$	-6.13	-6	-6.09	-103.7	-75.0	(-75.0)				

This **Discussion Paper Series** is published by the **Center for Glocal Strategy**(changed from the **Center for Business Creation** on April 2016) .

Discussion Paper Series
Institute of Economic Research
Otaru University of Commerce

No.	Title	Author/s	Date
1.	ﾎｰｷﾝｽﾞ=ｻｲﾓﾝの条件に関する諸説の統合について	ﾀﾞｽｸﾞﾌﾟﾀ,ﾃﾞｨﾊﾟﾝｶｰ	Jul.1992
2.	Motivation and Causal Inferences in the Budgetary Control	Yoshihiro Naka	Aug.1992
3.	Проблемы управления рабочей силой на предприятиях Далънего Востока(социологические аспекты)	Анатолий Михайловичн Шкурки	Nov.1992
4.	Dynamic Tax Incidence in a Finite Horizon Model	Jun-ichi Itaya	Jan.1993
5.	Business Cycles with Asset Price Bubbles and the Role of Monetary Policy	Hiroshi Shibuya	Jun.1993
6.	Continuous Double-Sided Auctions in Foreign Exchange Markets	Ryosuke Wada	Aug.1993
7.	The Existence of Ramsey Equilibrium with Consumption Externality	Sadao Kanaya & Tomoichi Shinotsuka	Sep.1993
8.	Money, Neutrality of Consumption Taxes, and Growth in Intertemporal Optimizing Models	Jun-ichi Itaya	Nov.1993
9.	Product Returns in the Japanese Distribution System:A Case Study of a Japanese Wholesaler's Return Reduction Efforts	Jeffery Alan Brunson	Mar.1994
10.	Dynamics, Consistent Conjectures and Heterogeneous Agents in the Private Provision of Public Goods	Jun-ichi Itaya & Dipankar Dasgupta	Jun.1994
11.	Intra-industry Investment and Imperfect Markets A Geometric approach in Simple General Equilibrium	Laixun Zhao	Oct.1994
12.	Sit-Down to Split:Flint GM Workers in 1937-1939	Satoshi Takata	Dec.1994
13.	The Complementarity between Endogenous Protection and Direct foreign Investment	Laixun Zhao	Feb.1995
14.	Consumption Taxation and Tax Prepayment approach in Dynamic General equilibrium Models with Consumer Durables	Jun-ichi Itaya	Mar.1995
15.	Regulatory System and Supervision of the Financial Institutions in Japan	Osamu Ito	May 1995
16.	Financial Restructuring and the U. S. Regulatory Framework	Jane W. D'Arista	May 1995
17.	The Legacy of the Bubble Economy in Japan:Declining cross Shareholding and Capital Formation	Hiroo Hojo	May 1995
18.	Stockownership in the U. S.:Capital Formation and Regulation	Marshall E. Blume	May 1995
19.	International Joint Ventures and Endogenous Protection a Political-Economy Approach	Laixun Zhao	Nov.1995
20.	GM社をめぐるｱﾒﾘｶ労働史研究：ﾌｧｲﾝとｴｯﾂﾞﾌｫｰｽの現場像の吟味	高田聡	Feb.1996
21.	卸売業の経営と戦略ーー卸売流通研究会ﾋｱﾘﾝｸﾞ調査録(1):日用雑貨卸売企業	卸売流通研究会（代表　高宮城朝則）	Apr.1996
22.	卸売業の経営と戦略ーー卸売流通研究会ﾋｱﾘﾝｸﾞ調査録(2):食品・酒類卸売企業	卸売流通研究会（代表　高宮城朝則）	Apr.1996
23.	A Note on the Impacts of Price Shocks on Wage in Unionized Economies	Laixun Zhao	May 1996
24.	Transfer Pricing and the Nature of the subsidiary firm	Laixun Zhao	Jun.1996
25.	The Incidence of a Tax on Pure in an Altruistic Overlapping Generations Economy	Jun-ichi Itaya	Sep.1996
26.	'Small Government' in the 21st Century	Hiroshi Shibuya	Sep.1996
27.	Characteristics and Reforms of Public Health Insurance System in Japan	Takashi Nakahama	Sep.1996
28.	The Role of Local Governments in Urban Development Policy	Yoshinori Akiyama	Sep.1996
29.	Optimal Taxation and the Private Provision of Public Goods	Jun-ichi Itaya & David de Meza & Gareth D. Myles	Oct.1996
30.	Comparison of Agricultural Policy in the U. S. and the Japan	Toshikazu Tateiwa	Oct.1996
31.	US Health Insurance:Types, Patterns of Coverage and Constraints to Reform	Dwayne A. Banks	Oct.1996
32.	International Capital Flows and National Macroeconomic Policies	Jane W. D'Arista	Oct.1996
33.	Financial Liberalization and Securitization in Housing Finance and the Changing Roles of the Government	Syn-ya Imura	Oct.1996
34.	Social Efficiency and the 'Market Revolution' in US Housing Finance	Gary Dymski & Dorene Isenberg	Oct.1996
35.	Government Expenditure and the Balance of Payments:Budget Deficit, Financial Integration, and Economic Diplomacy	Hiroshi Shibuya	Nov.1996
36.	A History of PBGC and Its Roles	C. David Gustafson	Nov.1996
37.	Dynamic Provision of Public Goods as Environmental Externalities	Toshihiro Ihori & Jun-ichi Itaya	Mar.1997

38.	A Comparative Static Analysis of the Balanced Budget Incidence in the Presence of Sector-Specific Unemployment	Koh Sumino	Mar.1997
39.	An Econometric Study of Trade Creation and Trade Diversion in the EEC,LAFTA and CMEA:A Simple Application of the Gravity Model	Masahiro Endoh	Apr.1997
40.	A Dynamic Model of Fiscal Reconstruction	Toshihiro Ihori & Jun-ichi Itaya	Apr.1997
41.	The Japanese Way of Solving Financial Institution Failures	Osamu Ito	Jul.1997
42.	The Federal Role in Community Development in the U.S. :Evolution vs. Devolution	Jane Knodell	Oct.1997
43.	Rent-Seeking Behavior in the War of Attrition	Jun-ichi Itaya & Hiroyuki Sano	Oct.1997
44.	ｻﾊﾘﾝ石油・ｶﾞｽ開発ﾌﾟﾛｼﾞｪｸﾄと北海道経済の活性化 第1号	北東ｱｼﾞｱｰｻﾊﾘﾝ研究会	May 1998
45.	購買部門の戦略性と企業間連携について	伊藤 一	Jun.1998
46.	The Formation of Customs Unions and the Effect on Government Policy Objectives	Masahiro Endoh	Jul.1998
47.	The Transition of Postwar Asia-Pacific Trade Relations	Masahiro Endoh	Jul.1998
48.	地域型ﾍﾞﾝﾁｬｰ支援ｼｽﾃﾑの研究 Ⅰ－道内製造業系ﾍﾞﾝﾁｬｰ企業のｹｰｽｽﾀﾃﾞｨｰ	地域経済社会ｼｽﾃﾑ研究会 日本開発銀行札幌支店	Jul.1998
49.	Fiscal Reconstruction Policy and Free Riding Behavior of Interest Groups	Toshihiro Ihori & Jun-ichi Itaya	Aug.1998
50.	Quellen zum Markwesen des Osnabrücker Landes im Niedersächsischen Staatsarchiv Osnabrück(mit Schwerpunkt :Verfassung,Hölting,Siedlung und Konflikten im 17.und 18.Jahrhundert)	Susumu Hirai	Sep.1998
51.	Equity and Continuity with a Continuum of Generations	Tomoichi Shinotsuka	Dec.1998
52.	Public Resources Allocation and Election System	Akihiko Kawaura	Mar.1999

Discussion Paper Series
Center for Business Creation
Otaru University of Commerce

53.	消費者の価格ﾌﾟﾛﾓｰｼｮﾝ反応への影響を考慮した広告効果測定結果ﾓﾃﾞﾙの構築	奥瀬喜之	Jun.1999
54.	地域型ﾍﾞﾝﾁｬｰ支援ｼｽﾃﾑの研究 Ⅱ－地域型ﾍﾞﾝﾁｬｰ･ｲﾝｷｭﾍﾞｰｼｮﾝの設計－	小樽商科大学ﾋﾞｼﾞﾈｽ創造ｾﾝﾀｰ & 日本開発銀行札幌支店	Jul.1999
55.	ｻﾊﾘﾝ石油・ｶﾞｽ開発ﾌﾟﾛｼﾞｪｸﾄと北海道経済の活性化 第2号	北東ｱｼﾞｱｰｻﾊﾘﾝ研究会	May 1999
56.	石鹸洗剤ﾒｰｶｰにおけるﾏｰｹﾃｨﾝｸﾞ･ﾁｬﾈﾙ行動の変遷	高宮城朝則	Dec.1999
57.	長期的取引関係における資源蓄積と展開	近藤公彦&坂川裕司	Dec.1999
58.	Exernalities:A Pigovian Tax vs. A Labor Tax	Ko Sumino	Dec.1999
59.	A New Dimension of Service Quality:An Empirical Study in Japan.	Makoto Matsuo & Carolus Praet & Yoshiyuki Okuse	Dec.1999
60.	Aftermath of the Flint Sit-Down Strike:Grass-Roots Unionism and African-American Workers, 1937-1939	Satoshi Takata	Mar.2000
61.	Tariff induced dumping in the intermediate-good market	Chisato Shibayama	Apr.2000
62.	Deregulation, Monitoring and Ownership structure:A Case Study of Japanese Banks	Akihiko Kawaura	Apr.2000
63.	ｻﾊﾘﾝ石油・ｶﾞｽ開発ﾌﾟﾛｼﾞｪｸﾄと北海道経済の活性化 第3号	北東ｱｼﾞｱｰｻﾊﾘﾝ研究会	Apr.2000
64.	A Cooperative and Competitive Organizational Culture, Innovation, and Performance: An Empirical Study of Japanese Sales Departments	Makoto Matsuo	May 2000
65.	Foreign Exchange Market Maker's Optimal Spread with Heterogeneous Expectations	Ryosuke Wada	Jun.2000
66.	ダンピングとダンピング防止法の起源 歴史的文脈における「不公正貿易」概念の成立	柴山千里	Oct.2000
67.	The Organizational Learning Process: A Review	Makoto Matsuo	Dec.2000
68.	The Weak Core of Simple Games with Ordinal Preferences: Implementation in Nash Equilibrium	Tomoichi Shinotsuka & Koji Takamiya	Jan.2001
69.	業態開発におけるイノベーションと競争－ビブレのケース－	近藤公彦	Jan.2001
70.	Budget Distribution Problem	Tomoichi Shinotsuka	Feb.2001
71.	小売バイヤー組織の機能と顧客対応	伊藤 一	May 2001
72.	The Effect of Intra-Organizational Competition on Knowledge Creation:Case Study of a Japanese Financial Company	Makoto Matsuo	May 2001
73.	ｻﾊﾘﾝ石油・ｶﾞｽ開発ﾌﾟﾛｼﾞｪｸﾄと北海道経済の活性化 第4号	北東ｱｼﾞｱｰｻﾊﾘﾝ研究会	Mar.2001
74.	The Weak Core of Simple Games with Ordinal Preferences: Implementation in Nash Equilibrium	Tomoichi Shinotsuka & Koji Takamiya	Oct.2001
75.	環境保全型河川計画と景観構築に係る計画技術の研究	地域環境問題研究会（代表 八木宏樹）	Oct.2001
76.	Additivity, Bounds, and Continuity in Budget Distribution Problem	Tomoichi Shinotsuka	Dec.2001

77. Monetary Policy in Bhutan: Implications of Indian Rupee Circulation	Akihiko Kawaura	Dec.2001
78. Optimal Multiobject Auctions with Correlated Types	Tomoichi Shinotsuka & Simon Wilkie	Feb.2002
79. ｻﾊﾘﾝ石油・ｶﾞｽ開発ﾌﾟﾛｼﾞｪｸﾄと北海道経済の活性化 第5号	北東ｱｼﾞｱｰｻﾊﾘﾝ研究会	Mar.2002
80. The Case Study of Retail Buying Organization in Japanese Context	Hajime Itoh	Mar.2002
81. 宿泊業のサービスのサービス構成要素に関する重要度調査法に関しての一考察 北海道への台湾人観光客の事例を中心に	稲葉由之&沈潔如&伊藤 一	Feb.2003
82. ブティック経営における販売要素の分析 -AHPによる経営者・販売員間における重要度認識比較に関する一考察-	伊藤 一&橋詰敦樹	Mar.2003
83. 温泉地に対するイメージギャップに関する調査	伊藤 一	Mar.2003
84. Literature Review on Retail Buyer from Research on Industrial Purchasing	Hajime Itoh	
85. The Comparison Study on Retail Buyer Behaviour between UK, Australia and Japan	Hajime Itoh	
86. 社会科学研究の基礎－大学院生のための研究法－	ダン・レメニイ他著 抄訳 稲葉由之&奥瀬善之&近藤公彦&玉井健一&高宮城朝則&松尾 睦	Mar.2002
87. マーケティング行為からみた小売業による需要創造 －明治期呉服店の経営行為を考察対象として－	坂川裕司	May 2002
88. Interdependent Utility Functions in an Intergenerational Context	Tomoichi Shinotsuka	May 2002
89. Internal and External Views of the Corporate Reputation in the Japanese Hotel Industry	Hajime Itoh	Feb.2003
90. ｻﾊﾘﾝ石油・ｶﾞｽ開発ﾌﾟﾛｼﾞｪｸﾄと北海道経済の活性化 第6号	北東ｱｼﾞｱｰｻﾊﾘﾝ研究会	Mar.2003
91.小売購買行動研究に関する展望 －「買い手視角」での小売購買行動研究に向けて－	坂川裕司	May 2003
92.商品購買における「情報システムの逆機能」 －リスク回避的バイヤーにみる合理性とその弊害－	坂川裕司	Sep.2003
93.An Experiment of Round-Robin Tournament by Excel's Macro -Using 160 Students' Data from Cournot Duopoly Game-	Masaru Uzawa	Apr.2004
94.Earnings Management through Deferred Tax Aseets -In Case of Banking Company-	Hiroshi Onuma	Jun.2004
97.Competition between Matching Markets	Koji Yokota	May 2005
98.On the role of asymmetric information in the aggregate matching function	Koji Yokota	Apr.2006
99.A note on Optimal Taxation in the Presence of Externalities	Tomoichi Shinotsuka & Ko Sumino	Feb.2005
100. A Note on Jones' Model of Growth	Mutsuhiro Kato	Mar.2005
101. 整数ナップサック問題が多項式時間で解ける特殊な場合を定める条件について	飯田浩志	Jul.2005
102. ＩＴ技術者の熟達化と経験学習	松尾 睦	Sep.2005
103. Product De-listing by Retail Buyers: Relational Antecedents and Consequences	Gary Davies & Hajime Itoh	Dec.2005
104. 米国地域経営史における多文化主義的発展－１９３０年代ミシガン州フリントにおけるアフリカ系コミュニティの起業基盤を中心に－	高田 聡	May 2006
105. 環境便益を反映させた環境指標の開発 Developing an environmental indicator including environmental benefits	山本 充	Apr.2006
106. A Critical Investigation of Long-run Properties of Endogenous Growth Models	Mutsuhiro Kato	May 2006
107. What is National Income in Jones' Model of Growth? :An Expository Annotation	Mutsuhiro Kato	Jun.2006
108. A Further Analysis of the Consumer Behavior in Jones' R&D-Based Model of Economic Growth	Mutsuhiro Kato	Aug.2006
109. 看護師の経験学習プロセス	松尾 睦 & 正岡経子 & 吉田真奈美 & 丸山知子 & 荒木奈緒	Feb.2007
110. Comments on knapsack problems with a penalty	Iida Hiroshi	Mar.2007
111. 看護師の経験学習に関する記述的分析	松尾 睦 & 正岡経子 & 吉田真奈美 & 丸山知子 & 荒木奈緒	Jul.2007
112. 頂点被覆へのリスト減少法の解析に関する一考察	飯田浩志	Dec.2007
113. 小中学校における校長の経営観－探索的分析－	松尾 睦	Dec.2007
114. インタビュー調査：戦後復興期大阪における自転車部品製造業者・問屋の経営活動	田中幹大	Apr.2008
115. Partitionのある風景	飯田浩志	Jun.2008
116. Multiproduct Firms and Dumping	Chisato Shibayama & Yasunori Ishii	Jul.2008
117. モスクワの低層住宅団地開発―2つのケース－	小田福男	Mar.2009
118. 整数ナップサックの周期性について	飯田浩志	Mar.2009

119.	Discussion paper series no.118への補遺	飯田浩志	Jul.2009
120.	環境フィードバック効果を考慮したSandmoモデルによる二重配当仮説の再考察	角野　浩	Jul.2009
121.	部分線形モデルの差分推定量の漸近理論	劉　慶豊	Oct.2009
122.	モデル平均理論の新展開	劉　慶豊	Oct.2009
123.	Production Theory with Convex Labor Friction: Foundation of an Optimal Non-market-clearing Economy	Koji Yokota	Dec.2009
124.	19世紀ドイツの農村ゲマインデ制と政治参加資格 ---北西ドイツ・ハノーファーを中心に-----	平井　進	Feb.2010
125.	環境経営と企業業績に関する実証研究（再検討:2003-2008)	加賀田和弘	Apr.2010
126.	「北海道ブランド」の仕入れに関する研究 ―台湾小売バイヤーの視点から―	沈　潔如	Apr.2010
127.	Generalized Cp Model Averaging for Heteroskedastic models	Qingfeng Liu	Oct.2010
128.	How to solve the collapsing subset-sum problem revisited	Hiroshi Iida	Jan.2011
129.	顧客関係のマネジメントの系譜	近藤公彦	Feb.2011
130.	An Application of Forecast Combination Methods to Default Risk Prediction	Qingfeng Liu	Feb.2011
131.	An effect of consumer's earlier decision to purchase a discount ticket	Ryosuke Ishii & Kuninori Nakagawa	Feb.2011
132.	On the Behavior of money flows on the real side and the financial side in Hokkaido prefecture	Toshiaki Kanzaki	Mar.2011
133.	星野リゾート― 顧客志向の組織マネジメント ―	乙政佐吉 & 近藤隆史	Mar.2011
134.	（ケース）札幌ビズカフェ―地域企業家ネットワークにおける中間主導型組織の役割―	加藤敬太	Mar.2011
135.	二重配当効果の最適課税ルールにしたがった再評価	角野　浩	Mar.2011
136.	18・19世紀前半北海沿岸農村社会の地域役職者：Landschaft Eiderstedt	平井　進	Mar.2011
137.	Tax Collecting Efforts and Local Allocation Tax Grants in Japan: The Effect of Administrative Reform Incentive Assessment on Local Tax Collection Rates	Mitsunari Ishida	Mar.2011
138.	The bargaining family with strategic interaction	Atsue Mizushima & Koichi Futagami	Mar.2011
139.	Generalized Cp Model Averaging for Heteroskedastic Models (Revised Version)	Qingfeng Liu	Apr.2011
140.	Exclusion of agents, virtual surplus and a transversality condition in adverse selection	Naoki Kojima	May 2011
141.	Implementability by a canonical indirect mechanism of an optimal two-dimensional direct mechanism	Naoki Kojima	Jun.2011
142.	18・19世紀前半北西ドイツ北海沿岸地方の領邦官吏と自治組織役職者：Landschaft S?derdithmarschen	平井　進	Jun.2011
143.	CRMにおける顧客関係のマネジメント	近藤公彦	Aug.2011
144.	企業家ネットワーキングによる地域企業のビジネスシステム・イノベーション―サムライ日本プロジェクトの事例分析―	加藤敬太	Oct.2011
145.	Observable Actions	Ryosuke Ishii	Nov.2012
146.	Dumping in Transition Economies and the Effects of Anti-Dumping Policy	Chisato Shibayama & Yasunori Ishii	Mar.2012
147.	Time Discount and Convex Hiring Cost	Koji Yokota	May 2012
148.	Two-dimensional Mechanism Design and Implementability by an Indirect Mechanism	Naoki Kojima	Jun.2012
149.	北海道経済と開発のプロセス	神崎稔章	Dec.2012
150.	道内における地域経済の現状について	渡久地朝央 & Baljinnyam Maitsetseg	Dec.2012
151.	モンゴルにおける資本主義転換後の地域間経済格差に関するパネルデータ分析	渡久地朝央 & Baljinnyam Maitsetseg	Dec.2012
152.	モンゴルにおける食肉価格の動向に関するパネルデータ分析	Baljinnyam Maitsetseg & 渡久地朝央	Dec.2012
153.	付加価値率の動向と地方自治体による政策効果の関係について－北海道における製造産業を対象としたパネルデータ分析－	渡久地朝央	Dec.2012
154.	CRMにおける組織能力	近藤公彦	Feb.2013
155.	19世紀北西ドイツの農村ゲマインデ制の変革---自治参加資格と家屋・土地保有要件---	平井　進	Feb.2013
156.	北海道における産業クラスターに関する文献資料目録	加藤敬太	Mar.2013
157.	トヨタ自動車北海道のマネジメント	乙政佐吉	Mar.2013
158.	Mechanism design to the budget constrained buyer: a canonical mechanism approach	Naoki Kojima	May 2013
159.	First Price Package Auction with Many Traders	Yasuhiro Shirata	Jun.2013
160.	整数ナップサックの周期性についてあれこれ	飯田浩志	Jul.2013

161.	Non-cooperative versus Cooperative Family	Atsue Mizushima & Koichi Futagami	Oct.2013
162.	Perverse effects of a ban on child labour in an overlapping generations model	Kouki Sugawara & Atsue Mizushima & Koichi Futagami	Oct.2013
163.	Human Infrastructure, Child Labor, and Growth	Atsue Mizushima	Oct.2013
164.	18・19世紀前半北海沿岸農村社会の指導的地域役職者・領邦地方官吏と土地所有：Landschaft Norderdithmarschen	平井　進	Mar.2014
165.	ビジネスシステムの形成から見る6次産業化―パイオニアジャパングループの事例分析―	笹本香菜 & 加藤敬太	Mar.2014
166.	ナップサック問題への2近似算法について雑感	飯田浩志	Jul.2014
167.	A further addendum to "Some thoughts on the 2-approximation algorithm for knapsack problems: A survey"	Hiroshi Iida	Nov.2014
168.	ワイン産業における企業家コミュニティー北海道と山梨県の比較事例研究ー	長村知幸	Jan.2015
169.	サハリン住宅建設企業の動向―2013年9月調査に基づいて―	小田福男	Mar.2015
170.	Asymptotic cumulants of some information criteria	Haruhiko Ogasawara	Apr.2015
171.	*k*KPをめぐる雑感	飯田浩志	Jun.2015
172.	Should income inequality be praised? Multiple public goods Provision, income distribution and social welfare	Jun-ichi Itaya & Atsue Mizushima	Jun.2015
173.	三種の*k*KP→□→0-1KPまとめ	飯田浩志	Nov.2015
174.	Asymptotic cumulants of some information criteria (2nd version)	Haruhiko Ogasawara	Nov.2015
175.	A family of the adjusted estimators maximizing the asymptotic mean expected log-likelihood	Haruhiko Ogasawara	Nov.2015
176.	Distribution-free properties of some asymptotic cumulants for the Mallows C*p* and its modifications	Haruhiko Ogasawara	Dec.2015
177.	A family of the adjusted estimators maximizing the asymptotic predictive expected log-likelihood (2nd version)	Haruhiko Ogasawara	Feb.2016
178.	サハリン低層住宅調査(2015年秋)	小田福男	Mar.2016

Discussion Paper Series
Center for Glocal Strategy
Otaru University of Commerce

179.	又別のE-*k*KP→0-1KP	飯田浩志	Apr.2016
180.	企業の配当行動が利益平準化行動に及ぼす影響	市原啓善	May 2016
181.	General screening contracts to budget-constrained agents	Naoki Kojima	Nov.2016
182.	The revelation principle and regularity conditions	Naoki Kojima	Nov.2016
183.	Durable goods, upgrades and price-quality competition in network externalities	Naoki Kojima	Nov.2016
184.	Predictive estimation of a covariance matrix and its structural parameters	Haruhiko Ogasawara	Nov.2016
185.	直接*k*KP→0-1KP模索中	飯田浩志	Feb.2017
186.	ナップサック問題あたりの最近の話題	飯田浩志	Apr.2017
187.	Endogenous Timing and Income Inequality in the Voluntary Provision of Public Goods	Jun-ichi Itaya & Atsue Mizushima	May 2017
188.	A family of the information criteria using the phi-divergence for categorical data	Haruhiko Ogasawara	Aug.2017

Discussion Paper Series Center for Business Creation No.53-178	Jun.1999-Mar.2016
Discussion Paper Series Institute of Economic Research Otaru University of Commerce No.1-52	Jul.1992-Mar.1999
Discussion Paper Series Department of Economics, Otaru University of Commerce No.1-16	Feb.1985-Oct.1991
Discussion Paper Series Department of Commerce, Otaru University of Commerce No.1-2	Apr.1985-May 1989

Center for Glocal Strategy, National University Corporation Otaru University of Commerce
3-5-21, Midori, Otaru, Hokkaido 047-8501, Japan　Tel +81-134-27-5290　Fax +81-134-27-5293
E-mail:cbcjimu@office.otaru-uc.ac.jp

国立大学法人小樽商科大学グローカル戦略推進センター研究支援部門
〒047-8501　北海道小樽市緑３丁目５番２１号　Tel 0134-27-5482　Fax 0134-27-5483
E-mail:skonno@res.otaru-uc.ac.jp