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

Haruhiko Ogasawara

August 2017

Otaru University of Commerce
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 \( \phi \)-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 \( \phi \)-divergences are also used for estimation of parameters as well as criteria for the badness of a model, whose estimators are the minimizing \( \phi \)-divergence estimators (M\( \phi \)Es; Morales, Pardo & Vajda, 1995, p.350; Pardo, 2006, Chapter 5). The \( \phi \)-divergences and M\( \phi \)Es 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 $\phi$-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 $\phi$-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 $\phi$-divergences using different M$\phi$Es. The corresponding higher-order asymptotic bias of a $\phi$-divergence based on M$\phi$Es using different $\phi$-divergences will be shown, where the results depend on types of $\phi$-divergences and M$\phi$Es.

The $\phi$ information criterion denoted by $\phi$IC 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 $\phi$ICs. The higher-order bias term can be used for correction of the remaining bias yielding the modified $\phi$IC (M$\phi$IC or MPIC). Since the higher order term is complicated, a simple approximation (M*$\phi$IC or M*PIC) will be developed. It will be shown that the M*$\phi$IC performs better than the $\phi$IC in simulations for model selection.

2. The bias of the $\phi$-divergence

The $\phi$-divergence statistic for $K$-category multinomial data is defined by
\[ C_\phi = \frac{2n}{\phi''(1)} D_\phi = \frac{2n}{\phi''(1)} D_\phi (\mathbf{p}, \mathbf{\pi}) \quad \text{with} \quad D_\phi = \sum_{k=1}^{K} \pi_k \phi\left(p_k / \pi_k\right), \quad (2.1) \]

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

\[ x > 0, \quad \phi(1) = 0, \quad \phi'(1) \quad \text{(the first derivative at} \ x = 1), \]
\[ \phi''(1) \quad \text{(the second derivative at} \ x = 1) > 0, \quad (2.2) \]
\[ 0\phi(0 / 0) = 0, \quad 0\phi(v / 0) = \lim_{u \to \infty} \{\phi(u) / u\} \]

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

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

\[ \hat{C}_\phi = \frac{2n}{\phi''(1)} \hat{D}_\phi = \frac{2n}{\phi''(1)} D_\phi (\mathbf{p}, \hat{\mathbf{\pi}}). \quad (2.3) \]

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

\[ \phi(x) = \frac{x^{\lambda+1} - x}{\lambda(\lambda + 1)} - \frac{x - 1}{\lambda + 1} \quad (-\infty < \lambda < +\infty, \lambda \neq 0, -1). \quad (2.4) \]

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.
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

\[
G^2 = 2np_k \sum_{k=1}^{K} \ln\left(\frac{p_k}{\hat{\pi}_k}\right),
\]

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

\[
GM^2 = 2n\hat{\pi}_k \sum_{k=1}^{K} \ln\left(\frac{\hat{\pi}_k}{p_k}\right),
\]

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

\[
NM^2 = n \sum_{k=1}^{K} \frac{(p_k - \hat{\pi}_k)^2}{p_k},
\]

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

\[
T^2 = 4n \sum_{k=1}^{K} \left(\sqrt{p_k} - \sqrt{\hat{\pi}_k}\right)^2,
\]

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

\[
X^2 = n \sum_{k=1}^{K} \frac{(p_k - \hat{\pi}_k)^2}{\hat{\pi}_k}.
\]

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

\[
R_\phi = \mathbb{E} E^* \left\{ \frac{2n}{\phi''(1)} \sum_{k=1}^{K} \hat{\pi}_k \phi\left(\frac{p_k^*}{\hat{\pi}_k}\right) \right\},
\]

(2.11)
where $\mathbf{p}^* = (p_1^*, ..., p_K^*)'$ is a vector of sample proportions based on $n$ observations independent of $\mathbf{p}$. $E(\cdot)$ and $E^*(\cdot)$ are the expectations using the distribution of $\mathbf{p}$ and $\mathbf{p}^*$, respectively. Recall that $\hat{\pi}$ 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 $R_\phi$ is given by $E(\hat{C}_\phi) - R_\phi$.

**Theorem 1.** Assume that the $M\phi E$ of $\theta_0$ is given by minimizing $C_\phi$ when $\phi(\cdot) = \phi_1(\cdot)$ and that the $\phi$-divergence statistic is constructed by $C_\phi$ when $\phi(\cdot) = \phi_2(\cdot)$ using the $M\phi E$ with $\phi_1(\cdot)$, where $\phi_2(\cdot)$ is possibly different from $\phi_1(\cdot)$. Suppose that $\hat{\pi}$ can be expanded about $\pi_0$ in the Taylor series up to the third powers of the elements of $\mathbf{p} - \pi_0$ with the residual of order $O_p(n^{-2})$. Then, under correct model specification, the bias of $\hat{C}_\phi$ is given by

$$E(\hat{C}_\phi) - R_\phi = \frac{2}{\phi_2''} \sum_{k=1}^{K} \left[ -\frac{\phi''_2}{\pi_{0k}} \frac{\partial \pi_{0k}}{\partial \theta'_0} \frac{\partial \theta_0}{\partial \pi'_0} - nE\{(\mathbf{p} - \pi_0)(p_k - \pi_{0k})\} \right]$$

$$+ n^{-1} \left[ -\frac{\phi''_2}{\pi_{0k}} \frac{1}{2} \left( \frac{\partial^2 \pi_{0k}}{(\partial \theta'_0)^{<2>}} \left( \frac{\partial \theta_0}{\partial \pi'_0} \right)^{<2>} + \frac{\partial \pi_{0k}}{\partial \theta'_0} \left( \frac{\partial \theta_0}{\partial \pi'_0} \right)^{<2>} \right) \right]$$

$$\times n^2 E\{(\mathbf{p} - \pi_0)^{<2>}(p_k - \pi_{0k})\}$$

$$+ \frac{1}{6} \left[ \frac{\partial^3 \pi_{0k}}{(\partial \theta'_0)^{<3>}} \left( \frac{\partial \theta_0}{\partial \pi'_0} \right)^{<3>} + 3 \frac{\partial^2 \pi_{0k}}{(\partial \theta'_0)^{<2>}} \left( \frac{\partial^2 \theta_0}{(\partial \pi'_0)^{<2>}} \otimes \frac{\partial \theta_0}{\partial \pi'_0} \right) + \frac{\partial \pi_{0k}}{\partial \theta'_0} \left( \frac{\partial \theta_0}{\partial \pi'_0} \right)^{<3>} \right]$$

$$\times n^2 E\{(\mathbf{p} - \pi_0)^{<3>}(p_k - \pi_{0k})\} \rightarrow O(n^{-2})$$
\[- \frac{1}{2\pi_{0k}^2}(\phi_2^{(3)} + \phi_2^{''}) \left[ \begin{array}{c}
\frac{\partial \pi_{0k}}{\partial \theta_0} \\
\frac{\partial \theta_0}{\partial \pi_0'}
\end{array} \right] n^2 E\{(p - \pi_0)(p_k - \pi_{0k})^2\} \right]
\]
\[+ \frac{1}{2} \left\{ \frac{\partial^2 \pi_{0k}}{\partial \theta_0' (\partial \pi_0')}^{<2>} \right\} n^2 E\{(p - \pi_0)^{<2>}(p_k - \pi_{0k})^2\} \to O(n^{-2}) \right] \quad (D)
\]
\[+ \frac{1}{2\pi_{0k}^2}(\phi_2^{(3)} + 2\phi_2^{''}) \left[ \begin{array}{c}
\frac{\partial \pi_{0k}}{\partial \theta_0} \\
\frac{\partial \theta_0}{\partial \pi_0'}
\end{array} \right] n^2 E\{(p - \pi_0)^{<2>}(p_k - \pi_{0k})\} \right]
\]
\[+ 2 \times \frac{1}{2} \left\{ \frac{\partial \pi_{0k}}{\partial \theta_0' (\partial \pi_0')}^{<2>} \right\} \times \left\{ \begin{array}{c}
\frac{\partial^2 \pi_{0k}}{\partial \theta_0' (\partial \pi_0')}^{<2>} \\
\frac{\partial \theta_0}{\partial \pi_0'}^{<2>}
\end{array} \right\} \times n^2 E\{(p - \pi_0)^{<3>}(p_k - \pi_{0k})\} \to O(n^{-2}) \right] \quad (E)
\]
\[\frac{1}{6\pi_{0k}^3}(\phi_2^{(4)} + 2\phi_2^{(3)} + \phi_2^{''}) \left[ \begin{array}{c}
\frac{\partial \pi_{0k}}{\partial \theta_0} \\
\frac{\partial \theta_0}{\partial \pi_0'}
\end{array} \right] n^2 E\{(p - \pi_0)(p_k - \pi_{0k})^3\} \to O(n^{-2}) \right]
\]
\[+ \frac{1}{4\pi_{0k}^3}(\phi_2^{(4)} + 4\phi_2^{(3)} + 2\phi_2^{''}) \left[ \begin{array}{c}
\frac{\partial \pi_{0k}}{\partial \theta_0} \\
\frac{\partial \theta_0}{\partial \pi_0'}
\end{array} \right] n^2 E\{(p - \pi_0)^{<2>}(p_k - \pi_{0k})^2\} \to O(n^{-2}) \right]
\]
\[- \frac{1}{6\pi_{0k}^3}(\phi_2^{(4)} + 6\phi_2^{(3)} + 6\phi_2^{''}) \left[ \begin{array}{c}
\frac{\partial \pi_{0k}}{\partial \theta_0} \\
\frac{\partial \theta_0}{\partial \pi_0'}
\end{array} \right] n^2 E\{(p - \pi_0)^{<3>}(p_k - \pi_{0k})\} \to O(n^{-2}) \right] \quad (B)(A)
\]
\[+ O(n^{-2}) \equiv b + n^{-1}b_\Delta + O(n^{-2}) \]
\[= -2q + n^{-1}b_\Delta + O(n^{-2}), \]

where \( A^{<k>} = A \otimes \cdots \otimes A \) (k times of \( A \)) is the k-fold Kronecker product of \( A \) (a vector or matrix); \( \frac{\partial^j \pi_{0k}}{\partial \theta_0 (\partial \pi_0')}^{<j>} = \left. \frac{\partial^j \pi_k(\theta)}{\partial \theta_0^{<j>}} \right|_{\theta=\theta_0} \quad (j = 1, 2, 3) \); \( \frac{\partial^j \theta_0}{\partial \pi_0'}^{<j>} = \left. \frac{\partial^j \hat{\theta}}{\partial p^{<j>}} \right|_{p=\pi_0} \quad (j = 1, 2, 3) \) are the j-th order partial derivatives of \( \hat{\theta} \), a
possibly implicit function with respect to \( p \) at \( p = \pi_0 \) and \( \hat{\theta} = \theta_0 \); \( E(\cdot) \sim O(n^{-2}) \) 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., \[ \begin{bmatrix} \cdot \\ (A) \\ (A) \end{bmatrix} \] is for ease of finding correspondence.

The proof with the expression of \( \hat{\partial}^j(\theta) / (\hat{\partial}^j \pi_0) \) \((j = 1, 2, 3)\) 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 \( \partial^j \pi_{0k} / (\partial \theta_0)^{<j>} \)
\[= \partial^j \pi_k(\theta) / (\partial \theta)^{<j>} \bigg|_{\theta = \theta_0} \ (j = 1, 2, 3) \] are determined only by the model \( \pi = \pi(\theta) \). It is known that
\[ \frac{\partial \theta_0}{\partial \pi_0} = I_0^{-1} \frac{\partial \pi_0}{\partial \theta_0} \text{diag}^{-1}(\pi_0) \] (Morales et al., 1995, Theorem 3), where \( I_0 = \frac{\partial \pi_0}{\partial \theta_0} \text{diag}^{-1}(\pi_0) \frac{\partial \pi_0}{\partial \theta_0} \) is the Fisher information matrix per observation and \( \text{diag}^{-1}(\pi_0) = \{\text{diag}(\pi_0)\}^{-1} \). Equation (2.13) does not depend on \( \phi_1(\cdot) \) used for estimation of \( \theta_0 \). However, Lemma 2 in the appendix shows that \( \partial^j \theta_0 / (\partial \pi_0)^{<j>} \) \((j = 2, 3)\) generally depend on \( \phi_1(\cdot) \). Since there are \( \phi_2^{(3)} \) and \( \phi_2^{(4)} \) in \( b_\lambda \), we find that \( b_\lambda \) depend on \( \phi_2(\cdot) \) as well as \( \phi_1(\cdot) \).

In the case of the power divergence, using (2.4), \( \phi'' \), \( \phi^{(3)} \) and \( \phi^{(4)} \) in \( b_\lambda \) become
\[ \phi'' = 1, \ \phi^{(3)} = (\lambda - 1) \quad \text{and} \quad \phi^{(4)} = (\lambda - 1)(\lambda - 2), \] respectively. From (2.14), it is found that for Pearson’s \( X^2(\lambda = 1) \), \( \phi^{(j)}(j = 2, 3, \ldots) \) vanish.

The \( \phi \) information criterion (\( \phi \)IC or PIC) and the modified \( \phi \)IC (M\( \phi \)IC or MPIC) are defined as
\[ \phi \text{IC} = \hat{C}_\phi + 2q \quad \text{and} \quad \text{M}\phi \text{IC} = \hat{C}_\phi + 2q - n^{-1}\hat{b}_\lambda, \] respectively, where \( \hat{b}_\lambda \) is the sample counterpart of \( b_\lambda \). In \( \hat{b}_\lambda \), \( \theta_0 \) and \( \pi_0 \) are replaced
by $\hat{\theta}$ and $p$, respectively. Since $E(\hat{b}_\lambda) = b_\lambda + O(n^{-1})$ 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,

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

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

Corollary 1 shows that the bias of $\hat{C}_\phi$ is removed up to orders $O(1)$ and $O(n^{-1})$ in the $\phi IC$ and $M\phi IC$, respectively and that the $\phi IC$ and $M\phi IC$ 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 $\phi IC$ and $M\phi IC$ is seen as an upward correction of $\hat{C}_\phi$ which tends to be reduced as an estimator of $R_\phi$. 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 $\phi IC$ over $\hat{C}_\phi$ is trivial, that of the $M\phi IC$ 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 $\hat{\theta} = (p_1, \ldots, p_{K-1})'$ $\equiv p_{(K-1)}$

without loss of generality and $\hat{\pi} = p$, $b_\lambda$ becomes

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

\[
\begin{align*}
   b_{\lambda} &= -\frac{1}{2} \{(\lambda - 1)(\lambda - 2) + 4(\lambda - 1) + 4\} \sum_{k=1}^{K} \frac{1}{\pi_{0k}} \\sum_{k=1}^{K} 1/ \pi_{0k} \\
   &\quad + \frac{1}{2} (\lambda - 1)(\lambda - 2)(2K - 1) + (\lambda - 1)(4K - 2) + 3K - 1 \\
   &= -\frac{1}{2} (\lambda^2 + \lambda + 2) \left( \sum_{k=1}^{K} \frac{1}{\pi_{0k}} - 2K + 1 \right) - K + 1 \\
   &\leq -(K - 1) \left\{ \frac{1}{2} (\lambda^2 + \lambda + 2)(K - 1) + 1 \right\} \\
   &\leq -\frac{1}{8} (K - 1)(7K + 1).
\end{align*}
\]

The second last inequality stems from \( \sum_{k=1}^{K} 1/ \pi_{0k} \geq K^2 \). 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 \( b_{\lambda} \) is defined as

\[
b_{\lambda}^* = \{q / (K - 1)\} b_{\lambda},
\]

where \( b_{\lambda} \) 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, \( b_{\lambda} \) 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 \( \phi IC \) to the AIC give some justification for the approximation of (3.3).

Define

\[
\hat{M}^* \phi IC = \hat{C}_\phi + 2q - n^{-1}\hat{b}_\Delta^*
\]

\[
= \hat{C}_\phi + 2q - n^{-1}\{q / (K - 1)\}\hat{b}_\Delta,
\]

\[
= \phi IC - n^{-1}\{q / (K - 1)\}\hat{b}_\Delta,
\]

(3.4)

where \( \hat{b}_\Delta \) is the sample counterpart of \( b_\Delta \) in (3.1) for the saturated model. The improved behavior of the \( M^* \phi IC \) over the \( \phi IC \) 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 \( \phi \)-divergence statistics. The \( \phi IC \) and \( M^* \phi IC \) are used, rather than \( M \phi IC \), 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:

\[
\pi = \left( \frac{2 + \theta}{4}, \frac{1 - \theta}{4}, \frac{1 - \theta}{4}, \frac{\theta}{4} \right),
\]

(0 < \( \theta \) < 1),

(4.1)
where $\theta_0 = 0.4$ is employed in this section. Note that the MLE is algebraically given by
\[
\hat{\theta}_{ML} = \frac{1}{2} \left[ p_1 - 2p_2 - 2p_3 - p_4 + \left( (p_1 - 2p_2 - 2p_3 - p_4)^2 + 8p_4 \right)^{1/2} \right].
\] (4.2)

For competing models,
\[
\pi = (\theta_1, \theta_2, \theta_2, 1 - \theta_1 - 2\theta_2) \quad (0 < \theta_1 < 1, \ 0 < \theta_2 < 1, \ 0 < 1 - \theta_1 - 2\theta_2 < 1)
\] (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:
\[
\pi = \{e^{-\theta}, \ \theta e^{-\theta}, \ 1 - (1 + \theta)e^{-\theta}\}' \quad (0 < \theta < +\infty),
\] (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):
\[
\pi = \{e^{-\theta}, \ \theta e^{-\theta}, \ \theta^2 e^{-\theta}/2, 1 - (1 + \theta + 0.5\theta^2)e^{-\theta}\}' \quad (0 < \theta < +\infty).
\] (4.5)

A redundant model with
\[
\pi = \{e^{-\theta_1}, \ \theta_1 e^{-\theta_1}, \ \theta_2, 1 - (1 + \theta_1)e^{-\theta_1} - \theta_2\}'
\quad (0 < \theta_1 < +\infty, \ 0 < \theta_2 < 1, \ 0 < 1 - (1 + \theta_1)e^{-\theta_1} - \theta_2 < 1)
\] (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 $\theta_1$ and $\theta_2$, respectively. That is, in this set of models, we have three models.

For the $M\phi E$, 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 $M\phi E$, $\lambda = 0.5$ 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 $M\phi E$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 $\phi IC$ in 10,000 replications with three sample 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 $M^*\phi IC$ 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 $\phi IC$ are approximately 80 to 85%. While the differences of these proportions over different $\lambda$’s are relatively small, the proportions using $\lambda = 0.5$ 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 ($\hat{\lambda} = 0$) and estimators
using $\lambda = 1$ (Pearson’s statistic) are always used. The results seem comparable to those in Table 1. Table 3 shows results when the $M^*\phi IC$ are used. Note that in a $M^*\phi IC$, 
\[ \sum_{k=1}^{K} 1 / \pi_{0k} \text{ in } b_\lambda \text{ in (3.2) is replaced by its sample counterpart i.e., } \sum_{k=1}^{K} 1 / p_k. \]

The values in the table show added percentages of the correct models selected by the $M^*\phi IC$s over those by the $\phi IC$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 $M^*\phi IC$ 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 $\hat{b}_\lambda$ in the $M^*\phi IC$ decreases as $n$ increases since the associated term added to $O_p(1)$ is $-n^{-1}\hat{b}_\lambda$.

From Tables 1 to 3, the $\phi_2(\cdot)$s with $\lambda = 0.5$ 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 $\lambda = 0$ ($G^2$), $\lambda = 2/3$ (the Cressie-Read statistic), $\lambda = 1$ ($\chi^2$ Pearson’s statistic) and $\lambda = 2$ 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 $R_\phi$'s over 10,000 replications; A.B. (asymptotic bias) $= -2q$; H.A.B (higher-order asymptotic bias) $= -2q + n^{-1}b_\lambda$; S. $b_\lambda$ (simulated $b_\lambda$) = $n \times$ (the simulated bias +2 $q$); and $b_\lambda$ is given by Theorem 1. In Tables 4 to 8, the $b_\lambda^*$s are shown in parentheses, which are given by (3.3). Note that in the case of the saturated model $b_\lambda^* = b_\lambda$.

For instance, in Model 3 of Table 4, $b_\lambda^* = b_\lambda$ are $-21$, $-31$, $-39$ and $-75$ when $\lambda = 0, 2/3, 1$ 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

\[
b_\Delta^* = b_\Delta = -\sum_{k=1}^{K} 1/\pi_{0k} + K = -1/0.6 - 2/0.15 - 1/0.1 + 4 = -21 \quad \text{using}
\]

\[
\phi_2^{(4)} = 2, \quad \phi_2^{(3)} = -1 \quad \text{and} \quad \phi_2^{''} = 1. \]

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 \( \lambda \) s. The S. \( b_\Delta \) s are in a crude sense similar to the \( b_\Delta^* \) s. It is also found that the \( b_\Delta \) s are similar to the \( b_\Delta^* \) s. In Tables 5 to 8, similar tendencies are found.

5. Discussion

It is natural to ask an appropriate value of \( \lambda \) to be used in the \( \phi_{\text{IC}} \) s and \( \text{M}^*\phi_{\text{IC}} \) s when the power divergence is used. In the numerical illustrations, it was found that relatively large \( \lambda \) 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 \( \phi_{\text{IC}} \) and \( \text{M}^*\phi_{\text{IC}} \) with \( \lambda = 2/3 \) may be reasonable. The criteria with \( \lambda = 1 \) may also be reasonable considering good results in the numerical illustrations though they are limited ones. The \( \phi_1(\cdot) \) s yielding the \( \text{M}\phi_{\text{E}} \) s do not necessarily have to match the \( \phi_2(\cdot) \) s in the \( \phi_{\text{IC}} \) and \( \text{M}^*\phi_{\text{IC}} \). Tables 2 using the MLEs, when relatively large \( \lambda \) 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 NCs (no-convergences).

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

A1. Proof of Theorem 1

We give two lemmas.

**Lemma 1.**

\[
\frac{\partial^2 \hat{\pi}_k \phi(p_k / \hat{\pi}_k)}{\partial \hat{\pi}_k^2} \bigg|_{\hat{\pi}_k = \pi_{0k}, p_k = \pi_{0k}} = \frac{\phi^{''}}{\pi_{0k}}, \quad \frac{\partial^3 \hat{\pi}_k \phi(p_k / \hat{\pi}_k)}{\partial \hat{\pi}_k^3} \bigg|_{\hat{\pi}_k = \pi_{0k}, p_k = \pi_{0k}} = -\frac{1}{\pi_{0k}^2} (\phi^{(3)} + 3\phi^{''}),
\]

\[
\frac{\partial^4 \hat{\pi}_k \phi(p_k / \hat{\pi}_k)}{\partial \hat{\pi}_k^4} \bigg|_{\hat{\pi}_k = \pi_{0k}, p_k = \pi_{0k}} = \frac{1}{\pi_{0k}^3} (\phi^{(4)} + 8\phi^{(3)} + 12\phi^{''}),
\]

\[
\frac{\partial^2 \hat{\pi}_k \phi(p_k / \hat{\pi}_k)}{\partial \hat{\pi}_k \partial p_k} \bigg|_{\hat{\pi}_k = \pi_{0k}, p_k = \pi_{0k}} = -\frac{\phi^{''}}{\pi_{0k}}, \quad \frac{\partial^3 \hat{\pi}_k \phi(p_k / \hat{\pi}_k)}{\partial \hat{\pi}_k \partial p_k^2} \bigg|_{\hat{\pi}_k = \pi_{0k}, p_k = \pi_{0k}} = -\frac{1}{\pi_{0k}^2} (\phi^{(3)} + \phi^{''}),
\]

\[(A.1)\]

\[
\frac{\partial^3 \hat{\pi}_k \phi(p_k / \hat{\pi}_k)}{\partial \hat{\pi}_k \partial p_k^2} \bigg|_{\hat{\pi}_k = \pi_{0k}, p_k = \pi_{0k}} = \frac{1}{\pi_{0k}^2} (\phi^{(3)} + 2\phi^{''}),
\]

\[
\frac{\partial^4 \hat{\pi}_k \phi(p_k / \hat{\pi}_k)}{\partial \hat{\pi}_k \partial p_k^3} \bigg|_{\hat{\pi}_k = \pi_{0k}, p_k = \pi_{0k}} = -\frac{1}{\pi_{0k}^3} (\phi^{(4)} + 2\phi^{(3)}),
\]

\[
\frac{\partial^4 \hat{\pi}_k \phi(p_k / \hat{\pi}_k)}{\partial \hat{\pi}_k \partial p_k^2} \bigg|_{\hat{\pi}_k = \pi_{0k}, p_k = \pi_{0k}} = \frac{1}{\pi_{0k}^3} (\phi^{(4)} + 4\phi^{(3)} + 2\phi^{''}),
\]

\[
\frac{\partial^4 \hat{\pi}_k \phi(p_k / \hat{\pi}_k)}{\partial \hat{\pi}_k \partial p_k} \bigg|_{\hat{\pi}_k = \pi_{0k}, p_k = \pi_{0k}} = \frac{1}{\pi_{0k}^3} (\phi^{(4)} + 6\phi^{(3)} + 6\phi^{''}) \quad (k = 1, \ldots, K)
\]

Proof. Using \[\phi^{'}(1) = 0\], the direct derivation gives the results. Q.E.D.

**Lemma 2.** *The partial derivatives of \[\hat{\theta}\] using \[\phi(\cdot)\] with respect to \[p\] at \[p = \pi_{0}\] with the assumption of their existence are given as follows.*
\[
\frac{\delta^2 \theta_0}{\delta \pi_{0i} \delta \pi_{0j}} = - \left( \frac{\delta^2 D_{\theta_0}}{\delta \theta_0^2} \right)^{-1} \left( \frac{\delta^3 D_{\theta_0}}{\delta \theta_0 (\delta \theta_0^2)^{2>}} \frac{\partial \theta_0}{\partial \pi_{0i}} \otimes \frac{\partial \theta_0}{\partial \pi_{0j}} \right) + \sum_{(i,j)}^2 \frac{\delta^3 D_{\theta_0}}{\partial \theta_0 \partial \pi_{0i} \partial \pi_{0j}} \partial \theta_0 + \delta_j \frac{\delta^3 D_{\theta_0}}{\partial \theta_0 \partial \pi_{0i}^2} \]

\[
= - \frac{1}{\phi''} \Gamma_0 \left[ \sum_{a=1}^K \left\{ - \frac{1}{\pi_{0a}} (\phi_1^{(3)} + 3 \phi_1 '') \frac{\partial \pi_{0a}}{\partial \theta_0} \left( \frac{\partial \pi_{0a}}{\partial \theta_0} \right)^{2>} \right. \right. \\
\left. \left. + \phi'' \left( \frac{\partial^2 \pi_{0a}}{\partial \theta_0 \partial \theta_0} \otimes \frac{\partial \pi_{0a}}{\partial \theta_0} + \frac{\partial \pi_{0a}}{\partial \theta_0} \otimes \frac{\partial^2 \pi_{0a}}{\partial \theta_0 \left( \frac{\partial \theta_0}{\partial \theta_0} \right)^{2>}} \right) \right\} \frac{\partial \theta_0}{\partial \pi_{0i}} \otimes \frac{\partial \theta_0}{\partial \pi_{0j}} \right] \\
+ \sum_{(i,j)}^2 \left\{ \frac{1}{\pi_{0i}} (\phi_1^{(3)} + 2 \phi_1 '') \frac{\partial \pi_{0i}}{\partial \theta_0} \frac{\partial \pi_{0i}}{\partial \theta_0} - \phi'' \frac{\partial^2 \pi_{0i}}{\partial \theta_0 \partial \theta_0} \right\} \frac{\partial \theta_0}{\partial \pi_{0i}} \\
- \delta_j \left( \phi_1^{(3)} + \phi_1 '' \right) \frac{\partial \pi_{0i}}{\partial \theta_0} \right] \\
\]

\[
\frac{\delta^3 \theta_0}{\delta \pi_{0i} \delta \pi_{0j} \delta \pi_{0k}} = - \left( \frac{\delta^2 D_{\theta_0}}{\delta \theta_0^2} \right)^{-1} \left( \frac{\delta^4 D_{\theta_0}}{\delta \theta_0 (\delta \theta_0^2)^{3>}} \frac{\partial \theta_0}{\partial \pi_{0i}} \otimes \frac{\partial \theta_0}{\partial \pi_{0j}} \otimes \frac{\partial \theta_0}{\partial \pi_{0k}} \right) + \sum_{(i,j,k)}^3 \left( \frac{\delta^3 D_{\theta_0}}{\delta \theta_0 (\delta \theta_0^2)^{2>}} \frac{\partial \theta_0}{\partial \pi_{0i}} \otimes \frac{\partial^2 \theta_0}{\partial \pi_{0j} \partial \pi_{0k}} + \frac{\delta^4 D_{\theta_0}}{\partial \theta_0 (\delta \theta_0^2)^{3>}} \frac{\partial \theta_0}{\partial \pi_{0i}} \otimes \frac{\partial \theta_0}{\partial \pi_{0j}} \otimes \frac{\partial \theta_0}{\partial \pi_{0k}} \right) \\
+ \frac{\delta^3 D_{\theta_0}}{\delta \theta_0 \partial \pi_{0i} \partial \pi_{0j} \partial \pi_{0k}} + \delta_j \frac{\delta^4 D_{\theta_0}}{\partial \theta_0 \partial \pi_{0i} \partial \pi_{0j} \partial \pi_{0k}} + \delta_{jk} \frac{\delta^4 D_{\theta_0}}{\partial \theta_0 \partial \pi_{0i} \partial \pi_{0j} \partial \pi_{0k}} \right] \\
\]
\[-\frac{1}{\phi_1^n} \mathbf{I}_0 \left[ \sum_{a=1}^{\kappa} \left[ \frac{1}{\pi_{0a}^3} (\phi_1^{(4)} + 8 \phi_1^{(3)} + 12 \phi_1''') \frac{\partial \pi_{0a}}{\partial \theta_0} \left( \frac{\partial \pi_{0a}}{\partial \theta_0}' \right) \right] \right. \]
\[-\frac{1}{\pi_{0a}^2} (\phi_1^{(3)} + 3 \phi_1''') \left\{ \frac{\partial^2 \pi_{0a}}{\partial \theta_0 \partial \theta_0}' \otimes \left( \frac{\partial \pi_{0a}}{\partial \theta_0}' \right) \right\} + \frac{\partial \pi_{0a}}{\partial \theta_0} \frac{\partial^2 \pi_{0a}}{\partial \theta_0' \partial \theta_0}' \right\} + \frac{\partial \pi_{0a}}{\partial \theta_0} \frac{\partial^2 \pi_{0a}}{\partial \theta_0' \partial \theta_0}' \right\} \right] \]
\[+ \frac{\phi_1''}{\pi_{0a}} \left( \frac{\partial \pi_{0a}}{\partial \theta_0}' \right) \right\} \right] \right] \]
\[+ \sum_{(i,j,k)} \left[ \left\{ \frac{1}{\pi_{0i}^3} (\phi_1^{(4)} + 6 \phi_1^{(3)} + 6 \phi_1''') \frac{\partial \pi_{0i}}{\partial \theta_0} \left( \frac{\partial \pi_{0i}}{\partial \theta_0}' \right) \right\] \]
\[-\frac{\phi_1''}{\pi_{0i}^2} \frac{\partial^3 \pi_{0i}}{\partial \theta_0' \partial \theta_0' \partial \theta_0'} + 2 \frac{\phi_1''}{\pi_{0i}^2} \left( \frac{\partial^2 \pi_{0i}}{\partial \theta_0 \partial \theta_0'} \otimes \frac{\partial \pi_{0i}}{\partial \theta_0}' + \frac{\partial \pi_{0i}}{\partial \theta_0} \frac{\partial^2 \pi_{0i}}{\partial \theta_0' \partial \theta_0}' \right) \right\] \]
\[+ \delta_{ij} \left[ \frac{1}{\pi_{0i}^3} (\phi_1^{(4)} + 4 \phi_1^{(3)} + 2 \phi_1''') \frac{\partial \pi_{0i}}{\partial \theta_0} \frac{\partial \pi_{0i}}{\partial \theta_0}' - \frac{1}{\pi_{0i}^2} (\phi_1^{(3)} + \phi_1''') \frac{\partial^2 \pi_{0i}}{\partial \theta_0 \partial \theta_0'} \right] \]
\[-\delta_{jk} \left[ \frac{1}{\pi_{0i}^3} (\phi_1^{(4)} + 2 \phi_1^{(3)}) \frac{\partial \pi_{0i}}{\partial \theta_0} \right] \quad (i, j, k = 1, \ldots, K), \]
where \( D\phi \) in place of \( C\phi \) is used since minimizing \( D\phi \) is equivalent to that of \( C\phi \); \( I_0 \) 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 \( \delta_{ij} \) is the Kronecker delta with \( \delta_{jk} = \delta_{ij} \delta_{jk} \).

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 \( \hat{\pi} = \pi(p) \) is not seen as a function of \( p \)), \( D\phi \) 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

\[
\hat{\kappa}_k - \pi_{0k} = \frac{\partial \pi_{0k}}{\partial \theta_0} (p - \pi_0) + \frac{1}{2} \left\{ \frac{\partial^2 \pi_{0k}}{\partial \theta_0^2} \left( \frac{\partial \theta_0}{\partial \pi_0} \right)^{<2>} + \frac{\partial \pi_{0k}}{\partial \theta_0} \left( \frac{\partial^2 \theta_0}{\partial \pi_0} \right)^{<2>} \right\} (p - \pi_0)^{<2>}
+ \frac{1}{6} \left\{ \frac{\partial^3 \pi_{0k}}{\partial \theta_0^3} \left( \frac{\partial \theta_0}{\partial \pi_0} \right)^{<3>} + 3 \frac{\partial^2 \pi_{0k}}{\partial \theta_0^2} \left( \frac{\partial^2 \theta_0}{\partial \pi_0^2} \right)^{<3>} \otimes \left( \frac{\partial \theta_0}{\partial \pi_0} \right)^{<2>} + \frac{\partial \pi_{0k}}{\partial \theta_0} \left( \frac{\partial^3 \theta_0}{\partial \pi_0^3} \right)^{<3>} \right\} (p - \pi_0)^{<3>}
+ O_p(n^{-2}) \quad (k = 1,...,K).
\]

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

\[
E(C_{\phi}) - R_{\phi} = \frac{2n}{\phi_2} E \left[ \sum_{k=1}^K \left[ \hat{\kappa}_k \phi_2(p_k / \hat{\kappa}_k) - E^*\{\phi_2(p_k^* / \hat{\kappa}_k)\} \right] \right] \quad (A.4)\]
\[
\begin{align*}
&= \frac{2n}{\phi_2''} \sum_{k=1}^{K} \left[ \frac{1}{2} \sum_{i=0}^{2} \binom{2}{i} \frac{\partial^2 \hat{\pi}_k \phi_2(p_k / \hat{\pi}_k)}{\partial \hat{\pi}_k \partial p_k^{3-i}} \bigg|_{\hat{\pi}_k = \pi_{0k}, p_k = \pi_{0k}} n E\{ (\hat{\pi}_k - \pi_{0k})'(p_k - \pi_{0k})^{2-i} \} \right] + O(n^{-1}) + O(n^{-2}) \\
&+ \frac{n^{-1}}{6} \sum_{i=0}^{3} \binom{3}{i} \frac{\partial^3 \hat{\pi}_k \phi_2(p_k / \hat{\pi}_k)}{\partial \hat{\pi}_k \partial p_k^{3-i}} \bigg|_{\hat{\pi}_k = \pi_{0k}, p_k = \pi_{0k}} n^2 E\{ (\hat{\pi}_k - \pi_{0k})'(p_k - \pi_{0k})^{3-i} \} + O(n^{-2}) \\
&+ \frac{n^{-1}}{24} \sum_{i=4}^{4} \binom{4}{i} \frac{\partial^4 \hat{\pi}_k \phi_2(p_k / \hat{\pi}_k)}{\partial \hat{\pi}_k \partial p_k^{3-i}} \bigg|_{\hat{\pi}_k = \pi_{0k}, p_k = \pi_{0k}} n^2 E\{ (\hat{\pi}_k - \pi_{0k})'(p_k - \pi_{0k})^{4-i} \} + O(n^{-2}) \\
\end{align*}
\]

where \( E\{ \cdot \} \rightarrow O(n^{-1}) + O(n^{-2}) \) 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 \( \partial \theta_0' / \partial \pi_0' \) and \( \partial \theta_0' / (\partial \pi_0')^{<j>} \) \((j = 2, 3)\) are given by (2.13) and Lemma 2, respectively. The result of \( b = -2q \) is given as follows. From (2.13), it follows that

\[
\begin{align*}
b &= \frac{2}{\phi_2''} \left[ K \sum_{k=1}^{K} \frac{\phi_2''}{\pi_{0k}} \frac{\partial \pi_{0k}}{\partial \theta_0'} \frac{\partial \theta_0'}{\partial \pi_0'} E\{ (p - \pi_0)(p_k - \pi_{0k}) \} \right] \\
&= -2K \sum_{k=1}^{K} \frac{\partial \pi_{0k}}{\partial \theta_0'} I_0^{-1} \frac{\partial \pi_0'}{\partial \theta_0'} \text{diag}^{-1}(\pi_0) \{ \text{diag}(\pi_0) - \pi_0 \pi_0' \} \}_{kj} \\
&= -2K \sum_{k=1}^{K} \frac{\partial \pi_{0k}}{\partial \theta_0'} I_0^{-1} \frac{\partial \pi_0'}{\partial \theta_0'} (I_{(K)})_{kj} \\
&= -2 \text{tr} \left\{ I_0^{-1} \frac{\partial \pi_0'}{\partial \theta_0'} \text{diag}^{-1}(\pi_0) \frac{\partial \pi_0}{\partial \theta_0'} \right\} \\
&= -2q,
\end{align*}
\]

where \( I_0 = \frac{\partial \pi_0'}{\partial \theta_0'} \text{diag}^{-1}(\pi_0) \frac{\partial \pi_0}{\partial \theta_0'} \), with \( \text{diag}^{-1}(\pi_0) = \{ \text{diag}(\pi_0) \}^{-1} \); \( I_{(K)} \) is the \( K \times K \) identity matrix; and \( \cdot \}_{kj} \) is the \( k \)-th column of a matrix.

For \( b_\Delta \) in (2.12), we require the following
\[ \kappa_3(a,b,c) = n^2 \{(p_a - \pi_{0a})(p_b - \pi_{0b})(p_c - \pi_{0c})\} \]
\[ = \delta_{abc} \pi_{0a} - \sum_{(a,b,c)} \delta_{ab} \pi_{0a} \pi_{0c} + 2 \pi_{0a} \pi_{0b} \pi_{0c} \]  

\[ (\text{Stuart & Ort, 1994, Equation (7.18)) and} \]
\[ m_4(a,b,c,d) = n^2 \{(p_a - \pi_{0a})(p_b - \pi_{0b})(p_c - \pi_{0c})(p_d - \pi_{0d})\} \rightarrow O(n^{-2}) \]
\[ = (\delta_{ab} \pi_{0a} - \pi_{0a} \pi_{0b})(\delta_{cd} \pi_{0c} - \pi_{0c} \pi_{0d}) \]
\[ + (\delta_{ac} \pi_{0a} - \pi_{0a} \pi_{0c})(\delta_{bd} \pi_{0b} - \pi_{0b} \pi_{0d}) \]
\[ + (\delta_{ad} \pi_{0a} - \pi_{0a} \pi_{0d})(\delta_{cd} \pi_{0b} - \pi_{0b} \pi_{0c}) \]  

\[ (a,b,c,d = 1,...,K). \]

A.2 Proof of Theorem 2

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

Note that under the parametrization with \( \hat{\theta} = p_{(K-1)} \) for the saturated model, the \( K \times K \) matrix \( \frac{\partial \pi_0}{\partial \theta_0} \frac{\partial \theta_0}{\partial \pi_0} \), becomes singular, whose rank is \( K - 1 \) and

\[ \partial^j \theta_0 \ (\partial \pi_0)^{<j>} \ (j = 2, 3) \] do not vanish. For the results under \( \hat{\theta} = p_{(K-1)} \), see Ogasawara (2018).

Let
\[ \kappa_3(k) = n^2 \{(p_k - \pi_{0k})^3\} = \pi_{0k}(1 - \pi_{0k})(1 - 2\pi_{0k}) \] and
\[ m_4(k) = n^2 \{(p_k - \pi_{0k})^4\} \rightarrow O(n^{-2}) = 3\pi_{0k}^2(1 - \pi_{0k})^2 \] \( (k = 1,...,K) \).

Then, \( b_{\Delta} \) in Theorem 1 becomes
\[
b_\lambda = \frac{2}{\phi_2''} \sum_{k=1}^{K} \left\{ -\frac{1}{2\varpi_{0k}^2} (\phi_2^{(3)} + \phi_2^{''})\kappa_3(k) + \frac{1}{2\varpi_{0k}^2} (\phi_2^{(3)} + 2\phi_2^{''})\kappa_3(k) \right. \\
- \frac{1}{6\varpi_{0k}^3} (\phi_2^{(4)} + 2\phi_2^{(3)})m_4(k) + \frac{1}{4\varpi_{0k}^3} (\phi_2^{(4)} + 4\phi_2^{(3)} + 2\phi_2^{''})m_4(k) \\
- \frac{1}{6\varpi_{0k}^3} (\phi_2^{(4)} + 6\phi_2^{(3)} + 6\phi_2^{''})m_4(k) \left\} \right.
\]

\[
= \frac{1}{\phi_2''} \sum_{k=1}^{K} \left\{ -(\phi_2^{(3)} + \phi_2^{''}) + \phi_2^{(3)} + 2\phi_2^{''} \right\} \left( \frac{1}{\varpi_{0k}} - 3 + 2\varpi_{0k} \right) \\
+ \left\{ -(\phi_2^{(4)} + 2\phi_2^{(3)}) + (3/2)(\phi_2^{(4)} + 4\phi_2^{(3)} + 2\phi_2^{''}) - (\phi_2^{(4)} + 6\phi_2^{(3)} + 6\phi_2^{''}) \right\} \\
\times \left( \frac{1}{\varpi_{0k}} - 2 + \varpi_{0k} \right) \\
= \frac{1}{\phi_2''} \sum_{k=1}^{K} \phi_2'' \left( \frac{1}{\varpi_{0k}} - 3 + 2\varpi_{0k} \right) - \left( \phi_2^{(4)} + 2\phi_2^{(3)} + 3\phi_2^{''} \right) \left( \frac{1}{\varpi_{0k}} - 2 + \varpi_{0k} \right),
\]

which gives (3.1).

References


practical information-theoretic approach (2nd ed.). New York: Springer.


Mathematical Statistics, 22, 79-86.
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.


Table 1. 10,000 times the proportions of models selected by the $\phi$IC's (the number of replications = 10,000)

<table>
<thead>
<tr>
<th>Model:</th>
<th>$\lambda = 0$ ($G^2$)</th>
<th>$\lambda = -1$ ($GM^2$)</th>
<th>$\lambda = -2$ (Neyman)</th>
<th>$\lambda = -0.5$ ($T^2$)</th>
<th>$\lambda = 0.5$</th>
<th>$\lambda = 2/3$ (C-R)</th>
<th>$\lambda = 1$ ($X^2$)</th>
<th>$\lambda = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 50$</td>
<td>$n = 200$</td>
<td>$n = 800$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>8176</td>
<td>1061</td>
<td>763</td>
<td>8022</td>
<td>1180</td>
<td>798</td>
<td>7853</td>
<td>1342</td>
</tr>
<tr>
<td></td>
<td>8139</td>
<td>1066</td>
<td>795</td>
<td>8009</td>
<td>1178</td>
<td>813</td>
<td>7853</td>
<td>1344</td>
</tr>
<tr>
<td></td>
<td>8078</td>
<td>1147</td>
<td>775</td>
<td>7968</td>
<td>1212</td>
<td>820</td>
<td>7837</td>
<td>1358</td>
</tr>
<tr>
<td></td>
<td>8166</td>
<td>1062</td>
<td>772</td>
<td>8022</td>
<td>1168</td>
<td>810</td>
<td>7860</td>
<td>1340</td>
</tr>
<tr>
<td></td>
<td>8194</td>
<td>1069</td>
<td>737</td>
<td>8014</td>
<td>1189</td>
<td>797</td>
<td>7860</td>
<td>1343</td>
</tr>
<tr>
<td></td>
<td>8203</td>
<td>1112</td>
<td>685</td>
<td>8023</td>
<td>1193</td>
<td>784</td>
<td>7859</td>
<td>1348</td>
</tr>
<tr>
<td></td>
<td>8205</td>
<td>1122</td>
<td>673</td>
<td>8046</td>
<td>1189</td>
<td>765</td>
<td>7863</td>
<td>1344</td>
</tr>
<tr>
<td></td>
<td>8241</td>
<td>1150</td>
<td>609</td>
<td>8069</td>
<td>1183</td>
<td>748</td>
<td>7872</td>
<td>1335</td>
</tr>
<tr>
<td></td>
<td>8399</td>
<td>1601</td>
<td>8537</td>
<td>1463</td>
<td>8394</td>
<td>1606</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8389</td>
<td>1611</td>
<td>8484</td>
<td>1516</td>
<td>8374</td>
<td>1626</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8389</td>
<td>1611</td>
<td>8452</td>
<td>1548</td>
<td>8367</td>
<td>1633</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8398</td>
<td>1602</td>
<td>8501</td>
<td>1499</td>
<td>8379</td>
<td>1621</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8402</td>
<td>1598</td>
<td>8572</td>
<td>1428</td>
<td>8411</td>
<td>1589</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8402</td>
<td>1598</td>
<td>8573</td>
<td>1427</td>
<td>8417</td>
<td>1583</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8356</td>
<td>1644</td>
<td>8573</td>
<td>1427</td>
<td>8434</td>
<td>1566</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8306</td>
<td>1694</td>
<td>8549</td>
<td>1451</td>
<td>8452</td>
<td>1548</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. $n$ = the number of observations, $Z$ = the number of deleted cases with zero frequency, $NC$ = 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)

<table>
<thead>
<tr>
<th>Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda = 0 ) ((G^2))</td>
<td>7900</td>
<td>1287</td>
<td>813</td>
<td>7877</td>
<td>1318</td>
<td>805</td>
<td>7851</td>
<td>1367</td>
<td>782</td>
</tr>
<tr>
<td>( \lambda = -1 ) ((GM^2))</td>
<td>7786</td>
<td>1310</td>
<td>904</td>
<td>7843</td>
<td>1338</td>
<td>819</td>
<td>7836</td>
<td>1383</td>
<td>781</td>
</tr>
<tr>
<td>( \lambda = -2 ) (\text{(Neyman)})</td>
<td>7696</td>
<td>1365</td>
<td>939</td>
<td>7817</td>
<td>1353</td>
<td>830</td>
<td>7840</td>
<td>1377</td>
<td>783</td>
</tr>
<tr>
<td>( \lambda = -0.5 ) ((T^2))</td>
<td>7813</td>
<td>1324</td>
<td>863</td>
<td>7859</td>
<td>1329</td>
<td>812</td>
<td>7846</td>
<td>1373</td>
<td>781</td>
</tr>
<tr>
<td>( \lambda = 0.5 )</td>
<td>7912</td>
<td>1293</td>
<td>795</td>
<td>7880</td>
<td>1319</td>
<td>801</td>
<td>7854</td>
<td>1364</td>
<td>782</td>
</tr>
<tr>
<td>( \lambda = 2/3 ) (\text{(C-R)})</td>
<td>7934</td>
<td>1276</td>
<td>790</td>
<td>7884</td>
<td>1316</td>
<td>800</td>
<td>7855</td>
<td>1363</td>
<td>782</td>
</tr>
<tr>
<td>( \lambda = 1 ) ((X^2))</td>
<td>7955</td>
<td>1265</td>
<td>780</td>
<td>7891</td>
<td>1311</td>
<td>798</td>
<td>\textbf{7858}</td>
<td>1361</td>
<td>781</td>
</tr>
<tr>
<td>( \lambda = 2 )</td>
<td>\textbf{7962}</td>
<td>1278</td>
<td>760</td>
<td>\textbf{7905}</td>
<td>1309</td>
<td>786</td>
<td>7853</td>
<td>1371</td>
<td>776</td>
</tr>
</tbody>
</table>

The \( \phi_2 \) in a \( \phi \)IC matches the \( \phi_1 \) for estimation of parameters.
Table 2. 10,000 times the proportions of the correct (simplest) models selected by the φICs (the number of replications = 10,000)

<table>
<thead>
<tr>
<th>n</th>
<th>50</th>
<th>200</th>
<th>800</th>
<th>50</th>
<th>200</th>
<th>800</th>
<th>50</th>
<th>200</th>
<th>800</th>
</tr>
</thead>
<tbody>
<tr>
<td>The genetics of plants (Fisher, 1970; 4 categories)</td>
<td>The genetics of plants (Fisher, 1970; 4 categories)</td>
<td>The genetics of plants (Fisher, 1970; 4 categories)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-category truncated Poisson (Bishop et al., 1975)</td>
<td>3-category truncated Poisson (Bishop et al., 1975)</td>
<td>3-category truncated Poisson (Bishop et al., 1975)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n:</td>
<td>50</td>
<td>200</td>
<td>800</td>
<td>50</td>
<td>200</td>
<td>800</td>
<td>50</td>
<td>200</td>
<td>800</td>
</tr>
<tr>
<td>(\lambda = 0) ((G^2))</td>
<td>7753</td>
<td>7806</td>
<td>7867</td>
<td>8481</td>
<td>8488</td>
<td>8416</td>
<td>7875</td>
<td>7909</td>
<td>7886</td>
</tr>
<tr>
<td>(\lambda = -1) ((GM^2))</td>
<td>7491</td>
<td>7575</td>
<td>7654</td>
<td>8475</td>
<td>8449</td>
<td>8399</td>
<td>7772</td>
<td>7882</td>
<td>7884</td>
</tr>
<tr>
<td>(\lambda = -2) (Neyman)</td>
<td>7333</td>
<td>7698</td>
<td>7846</td>
<td>8475</td>
<td>8413</td>
<td>8383</td>
<td>7675</td>
<td>7834</td>
<td>7875</td>
</tr>
<tr>
<td>(\lambda = -0.5) ((T^2))</td>
<td>7589</td>
<td>7787</td>
<td>7864</td>
<td>8475</td>
<td>8451</td>
<td>8406</td>
<td>7810</td>
<td>7888</td>
<td>7882</td>
</tr>
<tr>
<td>(\lambda = 0.5)</td>
<td>7798</td>
<td>7808</td>
<td><strong>7874</strong></td>
<td>8487</td>
<td>8548</td>
<td>8429</td>
<td>7899</td>
<td>7914</td>
<td>7894</td>
</tr>
<tr>
<td>(\lambda = 2/3) (C-R)</td>
<td>7794</td>
<td>7814</td>
<td>7872</td>
<td><strong>8487</strong></td>
<td>8548</td>
<td>8437</td>
<td>7907</td>
<td>7914</td>
<td>7895</td>
</tr>
<tr>
<td>(\lambda = 1) ((X^2))</td>
<td>7833</td>
<td><strong>7829</strong></td>
<td>7867</td>
<td>8454</td>
<td><strong>8552</strong></td>
<td>8448</td>
<td>7932</td>
<td>7916</td>
<td><strong>7899</strong></td>
</tr>
<tr>
<td>(\lambda = 2)</td>
<td><strong>7874</strong></td>
<td>7816</td>
<td>7860</td>
<td>8409</td>
<td>8531</td>
<td><strong>8464</strong></td>
<td><strong>7948</strong></td>
<td><strong>7919</strong></td>
<td>7892</td>
</tr>
</tbody>
</table>

The MLEs \((\lambda = 0)\) are used for all power divergences.

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 MI $\phi$ICs over those by the IC $\phi$s (the number of replications = 10,000)

<table>
<thead>
<tr>
<th>The genetics of plants (Fisher, 1970; 4 categories)</th>
<th>3-category truncated Poisson (Bishop et al., 1975)</th>
<th>4-category truncated Poisson</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$: 50 200 800</td>
<td>50 200 800</td>
<td>50 200 800</td>
</tr>
</tbody>
</table>

The $\phi_1$ in a IC matches the $\phi_1$ for estimation of parameters.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$G^2$</th>
<th>$GM^2$</th>
<th>Neyman</th>
<th>$T^2$</th>
<th>C-R</th>
<th>$X^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.96</td>
<td>1.90</td>
<td>4.18</td>
<td>1.81</td>
<td>2.96</td>
<td>4.09</td>
</tr>
<tr>
<td>-1</td>
<td>0.57</td>
<td>0.47</td>
<td>0.99</td>
<td>0.45</td>
<td>1.05</td>
<td>1.10</td>
</tr>
<tr>
<td>-2 (Neyman)</td>
<td>.21</td>
<td>.16</td>
<td>.26</td>
<td>.08</td>
<td>.10</td>
<td>.14</td>
</tr>
<tr>
<td>-0.5</td>
<td>.35</td>
<td>.42</td>
<td>.65</td>
<td>.33</td>
<td>.33</td>
<td>.41</td>
</tr>
<tr>
<td>0.5</td>
<td>.59</td>
<td>.12</td>
<td>.23</td>
<td>.36</td>
<td>.22</td>
<td>.22</td>
</tr>
<tr>
<td>2/3 (C-R)</td>
<td>.10</td>
<td>.02</td>
<td>.02</td>
<td>.08</td>
<td>.06</td>
<td>.06</td>
</tr>
<tr>
<td>1</td>
<td>1.23</td>
<td>1.63</td>
<td>2.26</td>
<td>1.70</td>
<td>2.05</td>
<td>2.09</td>
</tr>
<tr>
<td>2</td>
<td>.27</td>
<td>.41</td>
<td>.50</td>
<td>.39</td>
<td>.44</td>
<td>.47</td>
</tr>
</tbody>
</table>

The MLEs ($\lambda = 0$) are used for all power divergences.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$G^2$</th>
<th>$GM^2$</th>
<th>Neyman</th>
<th>$T^2$</th>
<th>C-R</th>
<th>$X^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.30</td>
<td>3.02</td>
<td>4.23</td>
<td>3.07</td>
<td>3.07</td>
<td>5.30</td>
</tr>
<tr>
<td>-1</td>
<td>.38</td>
<td>.43</td>
<td>1.23</td>
<td>.57</td>
<td>.81</td>
<td>.88</td>
</tr>
<tr>
<td>-2 (Neyman)</td>
<td>.08</td>
<td>.10</td>
<td>.24</td>
<td>.06</td>
<td>.11</td>
<td>.34</td>
</tr>
<tr>
<td>-0.5</td>
<td>.41</td>
<td>.41</td>
<td>.24</td>
<td>.42</td>
<td>.35</td>
<td>.88</td>
</tr>
<tr>
<td>0.5</td>
<td>.77</td>
<td>.13</td>
<td>.27</td>
<td>.42</td>
<td>.20</td>
<td>.80</td>
</tr>
<tr>
<td>2/3 (C-R)</td>
<td>.08</td>
<td>.00</td>
<td>.06</td>
<td>.06</td>
<td>.20</td>
<td>.16</td>
</tr>
<tr>
<td>1</td>
<td>1.36</td>
<td>1.55</td>
<td>1.96</td>
<td>1.54</td>
<td>1.99</td>
<td>2.39</td>
</tr>
<tr>
<td>2</td>
<td>.28</td>
<td>.21</td>
<td>.63</td>
<td>.33</td>
<td>.47</td>
<td>.60</td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>λ</th>
<th>50</th>
<th>200</th>
<th>800</th>
<th>50</th>
<th>200</th>
<th>800</th>
<th>50</th>
<th>200</th>
<th>800</th>
</tr>
</thead>
<tbody>
<tr>
<td>λ = 0 (G²)</td>
<td>2.84</td>
<td>.56</td>
<td>.09</td>
<td>.53</td>
<td>.58</td>
<td>.07</td>
<td>1.27</td>
<td>.35</td>
<td>.12</td>
</tr>
<tr>
<td>λ = -1 (GM²)</td>
<td>2.36</td>
<td>.64</td>
<td>.16</td>
<td>.48</td>
<td>.08</td>
<td>.01</td>
<td>1.48</td>
<td>.27</td>
<td>.10</td>
</tr>
<tr>
<td>λ = -2 (Neyman)</td>
<td>4.15</td>
<td>1.01</td>
<td>.20</td>
<td>.73</td>
<td>.33</td>
<td>.10</td>
<td>2.30</td>
<td>.50</td>
<td>.20</td>
</tr>
<tr>
<td>λ = -0.5 (T²)</td>
<td>2.17</td>
<td>.38</td>
<td>.11</td>
<td>.48</td>
<td>.43</td>
<td>.06</td>
<td>1.63</td>
<td>.40</td>
<td>.10</td>
</tr>
<tr>
<td>λ = 0.5</td>
<td>3.19</td>
<td>.74</td>
<td>.19</td>
<td>.41</td>
<td>.17</td>
<td>.04</td>
<td>1.92</td>
<td>.31</td>
<td>.05</td>
</tr>
<tr>
<td>λ = 2/3 (C-R)</td>
<td>3.33</td>
<td>.80</td>
<td>.16</td>
<td>.43</td>
<td>.15</td>
<td>.14</td>
<td>2.00</td>
<td>.38</td>
<td>.10</td>
</tr>
<tr>
<td>λ = 1 (X²)</td>
<td>4.43</td>
<td>1.00</td>
<td>.30</td>
<td>1.11</td>
<td>.15</td>
<td>.18</td>
<td>2.28</td>
<td>.45</td>
<td>.09</td>
</tr>
<tr>
<td>λ = 2</td>
<td>6.32</td>
<td>1.69</td>
<td>.54</td>
<td>2.25</td>
<td>.26</td>
<td>.01</td>
<td>4.53</td>
<td>1.01</td>
<td>.19</td>
</tr>
</tbody>
</table>

The parameter estimators by $\lambda = 1 (X^2)$ are used for all power divergences.
The γ in a φIC matches the φ for estimation of parameters.

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

<table>
<thead>
<tr>
<th>Model 1</th>
<th>S.B.</th>
<th>A.B.</th>
<th>H.A.B.</th>
<th>S. b_Δ</th>
<th>b_Δ</th>
<th>S.B.</th>
<th>A.B.</th>
<th>H.A.B.</th>
<th>S. b_Δ</th>
<th>b_Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>λ = 0 (G^2)</td>
<td>-2.10</td>
<td>-2</td>
<td>-2.13</td>
<td>-4.8</td>
<td>-6.6</td>
<td>-2.05</td>
<td>-2</td>
<td>-2.03</td>
<td>-10.1</td>
<td>-6.6</td>
</tr>
<tr>
<td>λ = 2/3 (C-R)</td>
<td>-2.14</td>
<td>-2</td>
<td>-2.14</td>
<td>-6.8</td>
<td>-7.1</td>
<td>-2.06</td>
<td>-2</td>
<td>-2.04</td>
<td>-12.7</td>
<td>-7.1</td>
</tr>
<tr>
<td>λ = 1 (X^2)</td>
<td>-2.22</td>
<td>-2</td>
<td>-2.19</td>
<td>-10.8</td>
<td>-9.3</td>
<td>-2.09</td>
<td>-2</td>
<td>-2.05</td>
<td>-17.2</td>
<td>-9.3</td>
</tr>
<tr>
<td>λ = 2</td>
<td>-2.72</td>
<td>-2</td>
<td>-2.48</td>
<td>-36.2</td>
<td>-24.0</td>
<td>-2.21</td>
<td>-2</td>
<td>-2.12</td>
<td>-42.9</td>
<td>-24.0</td>
</tr>
<tr>
<td>n = 200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. b_Δ</td>
<td>-2</td>
<td>-2</td>
<td>-2.01</td>
<td>-5.4</td>
<td>-6.6</td>
<td>-2.01</td>
<td>-2</td>
<td>-2.01</td>
<td>-10.8</td>
<td>-7.1</td>
</tr>
<tr>
<td>b_Δ</td>
<td>-2.06</td>
<td>-2</td>
<td>-2.06</td>
<td>-10.8</td>
<td>-7.1</td>
<td>-2.02</td>
<td>-2</td>
<td>-2.02</td>
<td>-16.7</td>
<td>-9.3</td>
</tr>
<tr>
<td>λ = 0 (G^2)</td>
<td>-2.01</td>
<td>-2</td>
<td>-2.01</td>
<td>-5.4</td>
<td>-6.6</td>
<td>-2.01</td>
<td>-2</td>
<td>-2.01</td>
<td>-10.8</td>
<td>-7.1</td>
</tr>
<tr>
<td>λ = 2/3 (C-R)</td>
<td>-2.01</td>
<td>-2</td>
<td>-2.01</td>
<td>-10.8</td>
<td>-7.1</td>
<td>-2.02</td>
<td>-2</td>
<td>-2.02</td>
<td>-16.7</td>
<td>-9.3</td>
</tr>
<tr>
<td>λ = 1 (X^2)</td>
<td>-2.02</td>
<td>-2</td>
<td>-2.02</td>
<td>-16.7</td>
<td>-9.3</td>
<td>-2.02</td>
<td>-2</td>
<td>-2.02</td>
<td>-16.7</td>
<td>-9.3</td>
</tr>
<tr>
<td>λ = 2</td>
<td>-2.06</td>
<td>-2</td>
<td>-2.06</td>
<td>-47.1</td>
<td>-24.0</td>
<td>-2.06</td>
<td>-2</td>
<td>-2.06</td>
<td>-47.1</td>
<td>-24.0</td>
</tr>
<tr>
<td>n = 800</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. b_Δ</td>
<td>-2</td>
<td>-2</td>
<td>-2.01</td>
<td>-5.4</td>
<td>-6.6</td>
<td>-2.01</td>
<td>-2</td>
<td>-2.01</td>
<td>-10.8</td>
<td>-7.1</td>
</tr>
<tr>
<td>b_Δ</td>
<td>-2.06</td>
<td>-2</td>
<td>-2.06</td>
<td>-10.8</td>
<td>-7.1</td>
<td>-2.02</td>
<td>-2</td>
<td>-2.02</td>
<td>-16.7</td>
<td>-9.3</td>
</tr>
<tr>
<td>λ = 0 (G^2)</td>
<td>-2.01</td>
<td>-2</td>
<td>-2.01</td>
<td>-5.4</td>
<td>-6.6</td>
<td>-2.01</td>
<td>-2</td>
<td>-2.01</td>
<td>-10.8</td>
<td>-7.1</td>
</tr>
<tr>
<td>λ = 2/3 (C-R)</td>
<td>-2.01</td>
<td>-2</td>
<td>-2.01</td>
<td>-10.8</td>
<td>-7.1</td>
<td>-2.02</td>
<td>-2</td>
<td>-2.02</td>
<td>-16.7</td>
<td>-9.3</td>
</tr>
<tr>
<td>λ = 1 (X^2)</td>
<td>-2.02</td>
<td>-2</td>
<td>-2.02</td>
<td>-16.7</td>
<td>-9.3</td>
<td>-2.02</td>
<td>-2</td>
<td>-2.02</td>
<td>-16.7</td>
<td>-9.3</td>
</tr>
<tr>
<td>λ = 2</td>
<td>-2.06</td>
<td>-2</td>
<td>-2.06</td>
<td>-47.1</td>
<td>-24.0</td>
<td>-2.06</td>
<td>-2</td>
<td>-2.06</td>
<td>-47.1</td>
<td>-24.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 2</th>
<th>n = 50</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>λ = 0 (G^2)</td>
<td>-3.79</td>
<td>-4</td>
<td>-4.24</td>
<td>10.5</td>
<td>-12.0</td>
<td>-3.98</td>
<td>-4</td>
<td>-4.06</td>
<td>4.8</td>
<td>-12.0</td>
</tr>
<tr>
<td>λ = 2/3 (C-R)</td>
<td>-3.82</td>
<td>-4</td>
<td>-4.34</td>
<td>8.8</td>
<td>-17.2</td>
<td>-4.00</td>
<td>-4</td>
<td>-4.09</td>
<td>-.3</td>
<td>-17.2</td>
</tr>
<tr>
<td>λ = 1 (X^2)</td>
<td>-3.92</td>
<td>-4</td>
<td>-4.46</td>
<td>4.2</td>
<td>-23.0</td>
<td>-4.04</td>
<td>-4</td>
<td>-4.12</td>
<td>-7.4</td>
<td>-23.0</td>
</tr>
<tr>
<td>λ = 2</td>
<td>-4.50</td>
<td>-4</td>
<td>-5.06</td>
<td>-25.1</td>
<td>-53.0</td>
<td>-4.24</td>
<td>-4</td>
<td>-4.27</td>
<td>-47.7</td>
<td>-53.0</td>
</tr>
<tr>
<td>n = 800</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. b_Δ</td>
<td>-4</td>
<td>-4</td>
<td>-5.06</td>
<td>-25.1</td>
<td>-53.0</td>
<td>-4.24</td>
<td>-4</td>
<td>-4.27</td>
<td>-47.7</td>
<td>-53.0</td>
</tr>
<tr>
<td>b_Δ</td>
<td>-4.09</td>
<td>-4</td>
<td>-4.07</td>
<td>-70.2</td>
<td>-53.0</td>
<td>-50.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. n = the number of observations, S.B. = simulated bias, A.B. = asymptotic bias = -2q, H.A.B. = b + n^{-1}b_Δ = -2q + n^{-1}b_Δ, S. b_Δ = simulated b_Δ = n(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.
The $\phi_2$ in a $\phi IC$ matches the $\phi_1$ for estimation of parameters.

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

<table>
<thead>
<tr>
<th>Model 3</th>
<th>S.B.</th>
<th>A.B.</th>
<th>H.A.B.</th>
<th>S. $b_\lambda$</th>
<th>$b_\lambda$</th>
<th>S.B.</th>
<th>A.B.</th>
<th>H.A.B.</th>
<th>S. $b_\lambda$</th>
<th>$b_\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$n = 50$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$n = 200$</td>
<td></td>
</tr>
<tr>
<td>$\lambda = 0$ ($G^2$)</td>
<td>-6.00</td>
<td>-6</td>
<td>-6.42</td>
<td>.1</td>
<td>-21.0</td>
<td>-6.03</td>
<td>-6</td>
<td>-6.11</td>
<td>-6.1</td>
<td>-21.0</td>
</tr>
<tr>
<td>$\lambda = 2/3$ (C-R)</td>
<td>-6.35</td>
<td>-6</td>
<td>-6.62</td>
<td>-17.5</td>
<td>-31.0</td>
<td>-6.11</td>
<td>-6</td>
<td>-6.16</td>
<td>-21.1</td>
<td>-31.0</td>
</tr>
<tr>
<td>$\lambda = 1$ ($X^2$)</td>
<td>-6.70</td>
<td>-6</td>
<td>-6.78</td>
<td>-35.2</td>
<td>-39.0</td>
<td>-6.18</td>
<td>-6</td>
<td>-6.20</td>
<td>-35.1</td>
<td>-39.0</td>
</tr>
<tr>
<td>$\lambda = 2$</td>
<td>-9.01</td>
<td>-6</td>
<td>-7.50</td>
<td>-150.5</td>
<td>-75.0</td>
<td>-6.53</td>
<td>-6</td>
<td>-6.38</td>
<td>-105.7</td>
<td>-75.0</td>
</tr>
<tr>
<td>$n = 800$</td>
<td></td>
<td></td>
<td></td>
<td>( $b_\lambda^* = b_\lambda$ )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( $b_\lambda^* = b_\lambda$ )</td>
<td></td>
</tr>
<tr>
<td>$\lambda = 0$ ($G^2$)</td>
<td>-5.99</td>
<td>-6</td>
<td>-6.03</td>
<td>6.1</td>
<td>-21.0</td>
<td>-21.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 2/3$ (C-R)</td>
<td>-6.02</td>
<td>-6</td>
<td>-6.04</td>
<td>-13.6</td>
<td>-31.0</td>
<td>-31.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 1$ ($X^2$)</td>
<td>-6.04</td>
<td>-6</td>
<td>-6.05</td>
<td>-29.6</td>
<td>-39.0</td>
<td>-39.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda = 2$</td>
<td>-6.13</td>
<td>-6</td>
<td>-6.09</td>
<td>-103.7</td>
<td>-75.0</td>
<td>-75.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Title</td>
<td>Author/s</td>
<td>Date</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>ホーキンズ=サイモンの条件に関する諸説の統合について</td>
<td>ダスグプタ,ディバンカー</td>
<td>Jul.1992</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Motivation and Causal Inferences in the Budgetary Control</td>
<td>Yoshihiro Naka</td>
<td>Aug.1992</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Побе лы управлен ия рабочей си ль на предприяти ях Дальнего Востока (социологические аспекты)</td>
<td>Михайловичи Шкурки</td>
<td>Nov.1992</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Business Cycles with Asset Price Bubbles and the Role of Monetary Policy</td>
<td>Hiroshi Shibuya</td>
<td>Jun.1993</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Continuous Double-Sided Auctions in Foreign Exchange Markets</td>
<td>Ryosuke Wada</td>
<td>Aug.1993</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>The Existence of Ramsey Equilibrium with Consumption Externality</td>
<td>Sadao Kanaya &amp; Tomoichi Shinotsuka</td>
<td>Sep.1993</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Money, Neutrality of Consumption Taxes, and Growth in Intertemporal Optimizing Models</td>
<td>Jun-ichi Itaya</td>
<td>Nov.1993</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Intra-industry Investment and Imperfect Markets</td>
<td>Laixun Zhao</td>
<td>Oct.1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Sit-Down to Split: Flint GM Workers in 1937-1939</td>
<td>Satoshi Takata</td>
<td>Dec.1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>The Complementarity between Endogenous Protection and Direct foreign Investment</td>
<td>Laixun Zhao</td>
<td>Feb.1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Regulatory System and Supervision of the Financial Institutions in Japan</td>
<td>Osamu Ito</td>
<td>May 1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>The Legacy of the Bubble Economy in Japan: Declining cross Shareholding and Capital Formation</td>
<td>Hiroo Hojo</td>
<td>May 1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Stockownership in the U.S.: Capital Formation and Regulation</td>
<td>Marshall E. Blume</td>
<td>May 1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>International Joint Ventures and Endogenous Protection</td>
<td>Laixun Zhao</td>
<td>Nov.1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>GM社をめぐるアパラジ労働史研究：ファインとエッ渚フォースの現場像の吟味</td>
<td>高田聡</td>
<td>Feb.1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>御売業の経営と戦略 - 御売流通研究会プロジェクト調査録(1):日用品雑貨御売業企業</td>
<td>御売流通研究会 (代表 高宮城朝則)</td>
<td>Apr.1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>御売業の経営と戦略 - 御売流通研究会プロジェクト調査録(2):食品・酒類御売業企業</td>
<td>御売流通研究会 (代表 高宮城朝則)</td>
<td>Apr.1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>A Note on the Impacts of Price Shocks on Wage in Unionized Economies</td>
<td>Laixun Zhao</td>
<td>May 1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Transfer Pricing and the Nature of the subsidiary firm</td>
<td>Laixun Zhao</td>
<td>Jun.1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Characteristics and Reforms of Public Health Insurance System in Japan</td>
<td>Takashi Nakahama</td>
<td>Sep.1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>The Role of Local Governments in Urban Development Policy</td>
<td>Yoshinori Akiyama</td>
<td>Sep.1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Comparison of Agricultural Policy in the U.S. and the Japan</td>
<td>Toshikazu Tateiwa</td>
<td>Oct.1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>US Health Insurance: Types, Patterns of Coverage and Constraints to Reform</td>
<td>Dwayne A. Banks</td>
<td>Oct.1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>A History of PBGC and Its Roles</td>
<td>C. David Gustafson</td>
<td>Nov.1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Authors</td>
<td>Publication Date</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------------------</td>
<td>------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>the Presence of Sector-Specific Unemployment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>An Econometric Study of Trade Creation and Trade Diversion in the</td>
<td>Wasahiro Endoh</td>
<td>Apr. 1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EEC, LAFTA and CMEA: A Simple Application of the Gravity Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A Dynamic Model of Fiscal Reconstruction</td>
<td>Toshihiro Itori</td>
<td>Apr. 1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Devolution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>An Econometric Study of Trade Creation and Trade Diversion in the</td>
<td>Masahiro Endoh</td>
<td>Apr. 1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EEC, LAFTA and CMEA: A Simple Application of the Gravity Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Transition of Postwar Asia-Pacific Trade Relations</td>
<td>Masahiro Endoh</td>
<td>Jul. 1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groups</td>
<td>&amp; Jun-ichi Itaya</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quellen zum Markvosen des Osnabrucker Landes in Niederschazen</td>
<td>Susumu Hirai</td>
<td>Sep. 1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quellen zum Markvosen des Osnabrucker Landes in Niederschazinem</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staatsarchiv Osnabruck (mit Schwerpunkt: Verfassung, Hoelting,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siedlung und Konflikten im 17. und 18. Jahrhundert)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equity and Continuity with a Continuum of Generations</td>
<td>Tomoichi Shinotsuka</td>
<td>Dec. 1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public Resources Allocation and Election System</td>
<td>Akihiko Kawaura</td>
<td>Mar. 1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discussion Paper Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Center for Business Creation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Otaru University of Commerce</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aftermath of the Flint Sit-Down Strike: Grass-Roots Unionism and</td>
<td>Satoshi Takata</td>
<td>Mar. 2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-American Workers, 1937-1939</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tariff induced dumping in the intermediate-good market</td>
<td>Chisato Shibayama</td>
<td>Apr. 2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deregulation, Monitoring and Ownership structure: A Case Study of</td>
<td>Akihiko Kawaura</td>
<td>Apr. 2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese Banks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>石油・ガス開発プロジェクトと北海道経済の活性化 第号</td>
<td>北東アジア-オホーツク研究会</td>
<td>May 1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>石油・ガス開発プロジェクトと北海道経済の活性化 第2号</td>
<td>高宮城雅則</td>
<td>Dec. 1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>長期的取引関係における資源蓄積と展開</td>
<td>近藤公彦 &amp; 坂川裕司</td>
<td>Dec. 1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aftermath of the Flint Sit-Down Strike: Grass-Roots Unionism and</td>
<td>Satoshi Takata</td>
<td>Mar. 2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-American Workers, 1937-1939</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tariff induced dumping in the intermediate-good market</td>
<td>Chisato Shibayama</td>
<td>Apr. 2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deregulation, Monitoring and Ownership structure: A Case Study of</td>
<td>Akihiko Kawaura</td>
<td>Apr. 2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese Banks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>石油・ガス開発プロジェクトと北海道経済の活性化 第号</td>
<td>北東アジア-オホーツク研究会</td>
<td>May 1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>石油・ガス開発プロジェクトと北海道経済の活性化 第2号</td>
<td>高宮城雅則</td>
<td>Dec. 1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>長期的取引関係における資源蓄積と展開</td>
<td>近藤公彦 &amp; 坂川裕司</td>
<td>Dec. 1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aftermath of the Flint Sit-Down Strike: Grass-Roots Unionism and</td>
<td>Satoshi Takata</td>
<td>Mar. 2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-American Workers, 1937-1939</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tariff induced dumping in the intermediate-good market</td>
<td>Chisato Shibayama</td>
<td>Apr. 2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deregulation, Monitoring and Ownership structure: A Case Study of</td>
<td>Akihiko Kawaura</td>
<td>Apr. 2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese Banks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>石油・ガス開発プロジェクトと北海道経済の活性化 第号</td>
<td>北東アジア-オホーツク研究会</td>
<td>May 1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>石油・ガス開発プロジェクトと北海道経済の活性化 第2号</td>
<td>高宮城雅則</td>
<td>Dec. 1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>長期的取引関係における資源蓄積と展開</td>
<td>近藤公彦 &amp; 坂川裕司</td>
<td>Dec. 1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
77. Monetary Policy in Bhutan: Implications of Indian Rupee Circulation
Akiko Kawaura Dec. 2001

78. Optimal Multiobject Auctions with Correlated Types
Tomoichi Shinotsuka & Simon Wilkie Feb. 2002

79. サハリン石油・ガス開発プロジェクトと北海道経済の活性化 第5号
北東アジアサハリン研究会 Mar. 2002

80. The Case Study of Retail Buying Organization
Hajime Itoh Mar. 2002

81. 宿泊業のサービス構成要素に関する重要度調査法に関する一考察 北海道への台湾人観光客の事情を中心に
稲葉由之 & 沈潔如 & 松尾睦 & 橋詰敦樹 & 近藤公彦 & 松尾睦 & 高宮城朝則 & 松尾睦 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考察-
伊藤一 & 横出敬樹 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. The Case Study of Retail Buying Organization
Hajime Itoh Mar. 2002

85. マーケティング行為からみた小売業による需要創造－明治期長崎市の経営行為を考察対象として－
坂川裕司 May 2002

86. The Comparison Study on Retail Buyer Behaviour between UK, Australia and Japan
Hajime Itoh Mar. 2002

87. 商品購入における「情報システムの無能」－リースと効果的バイヤーにみる合理性とその弊害－
坂川裕司 Sep. 2003

88. Interdependent Utility Functions in an Intergenerational Context
Tomoichi Shinotsuka Feb. 2002

89. Internal and External Views of the Corporate Reputation in the Japanese Hotel Industry
Hajime Itoh Feb. 2003

90. 社会科学研究の基礎－大学院生のための研究法－
松尾睦 May 2002

91. An Experiment of Round-Robin Tournament by Excel's Macro
Masaru Uzawa Apr. 2004

92. Product De-listing by Retail Buyers: Relational Antecedents and Consequences
Gary Davies & Hajime Itoh Dec. 2005

93. What is National Income in Jones' Model of Growth? An Expository Annotation
Mutsuhiro Kato Aug. 2006

94. A Note on Optimal Taxation in the Presence of Externalities
Tomoichi Shinotsuka & Ko Sumino Feb. 2005

95. A Note on Knapsack Problems with a Penalty
Iida Hiroshi Mar. 2007

96. A Critical Investigation of Long-run Properties of Endogenous Growth Models
Mutsuhiro Kato May 2006

97. Partitionのある風景
饭田浩志 Jun. 2008

98. Multiproduct Firms and Dumping
Chisato Shibayama Jul. 2008 & Yasunori Ishii

99. モスクワの低層住宅団地開発－2つのケース－
田中幹大 Apr. 2008

100. 小中学校における校長の経営観-探索的分析-
松尾睦 Dec. 2007

101. 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

102. 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 Mar. 2007

103. 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

104. 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

105. 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

106. 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

107. 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

108. 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

109. 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

110. 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

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

112. 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

113. 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

114. 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

115. 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

116. 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

117. 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

118. 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

119. 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

120. 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

121. 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

122. 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

123. 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

124. 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

125. 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

126. 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

127. 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

128. 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

129. 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

130. 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

131. 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

132. 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

133. 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

134. 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

135. 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

136. 目安ナップサックの周期性について
飯田浩志 Mar. 2009

137. 目安ナップサックの周期性について
飯田浩志 Mar. 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分担 浅野浩 Oct.2009
121. 部分線形モデルの差分推定量の漸近理論 刘慶豐 Oct.2009
123. 19世紀ドイツの農村ゲマインデ制と政治参加資格 平井進 Feb.2010
124. 19世紀北西ドイツの農村ゲマインデ制と政治参加資格 平井進 Feb.2010
125. 仮説の再考察 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
128. How to solve the collapsing subset-sum problem revisited Hiroshi Iida Jan.2011
129. 環境経営を考慮したSimkoモデルによる二重配当付けにおける前向き推計 角野浩 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 Ryoosuke 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. 企業家ネットワーキングによる地域経済の成長——サムライ日本プロジェクトの事例分析—— 加藤敬太 Oct.2011
134. The bargaining family with strategic interaction Atsue Mizushima & Koichi Futagami Mar.2012
135. Mechanism design to the budget constrained buyer: A canonical mechanism approach Naoki Kojima May 2012
137. 企業家ネットワーキングによる地域経済の成長——サムライ日本プロジェクトの事例分析—— 加藤敬太 Oct.2011