Economic Review (Otaru University of Commerce),
Vol.70, No.2 & 3, 65-86, December, 2019.

Supplement to the paper "Asymptotic cumulants of the minimum phi-divergence estimator for categorical data under possible model misspecification"

Haruhiko Ogasawara

This article supplements Ogasawara (2019a), and gives
[0] miscellaneous issues in Section S0 with errata in Subsection S0.1 and some limiting results in Subsection S0.2,
[1] the partial derivatives of $\hat{\boldsymbol{\theta}}$ with respect to $\mathbf{p}$ evaluated at $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$ and $\mathbf{p} = \boldsymbol{\tau}$ in Section S1,
[2] added numerical results under model misspecification in Section S2 with Tables S1.1 to S1.11, where the applications of the asymptotic cumulants in interval estimation (for interval estimation, see Section S3) are shown in Tables S1.9 to S1.11,
[3] the corresponding results under correct model specification in Tables S2.1 to S2.13.

In the tables, ".00" indicates a rounded value of zero up to the second place while "0" indicates an exactly zero value. Note that Tables S2.1 to S2.13 are not included in this supplement, but available in Ogasawara (2019b).

References

Hall, P. (1992). On the removal of skewness by transformation. *Journal of the Royal Statistical Society, B, 54*, 221-228.

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

Ogasawara, H. (2012). Cornish-Fisher expansions using sample cumulants and monotonic transformations. *Journal of Multivariate Analysis, 103*, 1-18.

Ogasawara, H. (2019a). Asymptotic cumulants of the minimum phi-divergence estimator for categorical data under possible model misspecification. *Communications in Statistics – Theory and Methods* (online published). https://doi.org/10.1080/03610926.2019.1576888.

Ogasawara, H. (2019b). *Supplemental tables and R codes for the paper "Asymptotic cumulants of the minimum phi-divergence estimator for* *categorical data under possible model misspecification"*. Unpublished document. http://www.res.otaru-uc.ac.jp/~emt-hogasa/, http://hdl.handle.net/10252/00005864.

S0. Miscellaneous issues

S0.1 Errata

The term $-3\mathrm{E}[\{\hat{\theta}-\mathrm{E}(\hat{\theta})\}^2]$ on the right-hand side of the first equation for $\kappa_4(\hat{\theta})$ in (2.2) should be $-3\{\mathrm{E}[\{\hat{\theta}-\mathrm{E}(\hat{\theta})\}^2]\}^2$.

The factor $\otimes n^3\boldsymbol{\kappa}_3(\mathbf{p})$ on the left-hand side of the last equation for $\kappa_4(\hat{\theta})$ in (2.2) should be $\otimes n^2\boldsymbol{\kappa}_3(\mathbf{p})$.

The second duplicate inequality "$i \neq j$" in the last line of (2.4) should be deleted.

The phrase "*n* = 25, 200 and 800" in the fifth line of Section 3 should be "*n* = 50, 200 and 800".

The URLs in Ogasawara (2019a) for the supplement in the reference list of the published paper should be "http://www.res.otaru-uc.ac.jp/~emt-hogasa/, https://barrel.repo.nii.ac.jp/".

The URLs in Ogasawara (2019b) for the supplemental tables and R codes in the reference list of the published paper should be "http://www.res.otaru-uc.ac.jp/~emt-hogasa/, http://hdl.handle.net/10252/00005864".

S0.2 Some limiting results

The limiting results after (1.4) are derived as follows:

$$\lim_{\lambda\to 0}\left(\frac{x^{\lambda+1}-x}{\lambda(\lambda+1)}-\frac{x-1}{\lambda+1}\right)=\lim_{\lambda\to 0}\frac{x^{\lambda+1}-x}{\lambda}-x+1$$

$$=\lim_{\lambda\to 0}\frac{x(x^{\lambda}-1)}{\lambda}-x+1=x\ln x-x+1$$

and

$$\lim_{\lambda\to -1}\left(\frac{x^{\lambda+1}-x}{\lambda(\lambda+1)}-\frac{x-1}{\lambda+1}\right)=\lim_{\lambda\to -1}\frac{x^{\lambda+1}-x-\lambda(x-1)}{\lambda(\lambda+1)}$$

$$=\lim_{\lambda\to -1}\frac{x^{\lambda+1}-1-(\lambda+1)(x-1)}{\lambda(\lambda+1)}=-\lim_{\lambda\to -1}\frac{x^{\lambda+1}-1}{\lambda+1}+x-1$$

$$=-\ln x+x-1.$$

S1. The partial derivatives of $\hat{\boldsymbol{\theta}}$ with respect to p evaluated at $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$ and $\mathbf{p} = \boldsymbol{\tau}$

We use the following lemma.

Lemma 1. *Define* $\rho_i = \tau_i / \pi_{0i}$ $(i = 1, ..., K)$ *and* $\phi^{(i)}(x^*)$ $(i = 3, 4)$ *as the third and fourth derivatives of* $\phi(x)$ *with respect to x at* x^**, respectively. Then, with the assumption of the existence of the derivatives of* $\phi(\cdot)$,

$$\frac{\partial \pi_k \phi(p_k / \pi_k)}{\partial \pi_k}\Big|_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}} = \left\{ -\frac{p_k}{\pi_k}\phi'(p_k / \pi_k) + \phi(p_k / \pi_k) \right\}_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}}$$

$$= -\rho_k \phi'(\rho_k) + \phi(\rho_k),$$

$$\frac{\partial^2 \pi_k \phi(p_k / \pi_k)}{\partial \pi_k^2}\Big|_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}}$$

$$= \left\{ \frac{p_k^2}{\pi_k^3}\phi''(p_k / \pi_k) + \frac{p_k}{\pi_k^2}\phi'(p_k / \pi_k) - \frac{p_k}{\pi_k^2}\phi'(p_k / \pi_k) \right\}_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}}$$

$$= \frac{\rho_k^2}{\pi_{0k}}\phi''(\rho_k),$$

$$\frac{\partial^3 \pi_k \phi(p_k / \pi_k)}{\partial \pi_k^3}\Big|_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}} = \left\{ -\frac{p_k^3}{\pi_k^5}\phi^{(3)}(p_k / \pi_k) - 3\frac{p_k^2}{\pi_k^4}\phi''(p_k / \pi_k) \right\}_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}}$$

$$= -\frac{1}{\pi_{0k}^2}\{\rho_k^3 \phi^{(3)}(\rho_k) + 3\rho_k^2 \phi''(\rho_k)\},$$

(A.1)

$$\frac{\partial^4 \pi_k \phi(p_k / \pi_k)}{\partial \pi_k^4}\Big|_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}} = \left\{ \frac{p_k^4}{\pi_k^7}\phi^{(4)}(p_k / \pi_k) + \left(5\frac{p_k^3}{\pi_k^6} + 3\frac{p_k^3}{\pi_k^6} \right)\phi^{(3)}(p_k / \pi_k) \right.$$

$$\left. + 12\frac{p_k^2}{\pi_k^5}\phi''(p_k / \pi_k) \right\}_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}}$$

$$= \frac{1}{\pi_{0k}^3}\{\rho_k^4 \phi^{(4)}(\rho_k) + 8\rho_k^3 \phi^{(3)}(\rho_k) + 12\rho_k^2 \phi''(\rho_k)\},$$

$$\frac{\partial^2 \pi_k \phi(p_k / \pi_k)}{\partial \pi_k \partial p_k}\Big|_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}} = \left\{ -\frac{p_k}{\pi_k^2}\phi''(p_k / \pi_k) + (-1+1)\frac{1}{\pi_k}\phi'(p_k / \pi_k) \right\}_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}}$$

$$= \left\{ -\frac{p_k}{\pi_k^2}\phi''(p_k / \pi_k) \right\}_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}} = -\frac{\rho_k}{\pi_{0k}}\phi''(\rho_k),$$

$$\frac{\partial^3 \pi_k \phi(p_k / \pi_k)}{\partial \pi_k \partial p_k^2}\Big|_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}} = \left\{ -\frac{p_k}{\pi_k^3}\phi^{(3)}(p_k / \pi_k) - \frac{1}{\pi_k^2}\phi''(p_k / \pi_k) \right\}_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}}$$

$$= -\frac{1}{\pi_{0k}^2}\{\rho_k \phi^{(3)}(\rho_k) + \phi''(\rho_k)\},$$

$$\frac{\partial^3 \pi_k \phi(p_k / \pi_k)}{\partial \pi_k^2 \partial p_k}\Big|_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}} = \left\{ \frac{p_k^2}{\pi_k^4}\phi^{(3)}(p_k / \pi_k) + 2\frac{p_k}{\pi_k^3}\phi''(p_k / \pi_k) \right\}_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}}$$

$$= \frac{1}{\pi_{0k}^2}\{\rho_k^2 \phi^{(3)}(\rho_k) + 2\rho_k \phi''(\rho_k)\},$$

$$\frac{\partial^4 \pi_k \phi(p_k / \pi_k)}{\partial \pi_k \partial p_k^3}\Big|_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}} = \left\{ -\frac{p_k}{\pi_k^4}\phi^{(4)}(p_k / \pi_k) - \frac{2}{\pi_k^3}\phi^{(3)}(p_k / \pi_k) \right\}_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}}$$

$$= -\frac{1}{\pi_{0k}^3}\{\rho_k \phi^{(4)}(\rho_k) + 2\phi^{(3)}(\rho_k)\},$$

$$\frac{\partial^4 \pi_k \phi(p_k / \pi_k)}{\partial \pi_k^2 \partial p_k^2}\Big|_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}}$$

$$= \left\{ \frac{p_k^2}{\pi_k^5}\phi^{(4)}(p_k / \pi_k) + 4\frac{p_k}{\pi_k^4}\phi^{(3)}(p_k / \pi_k) + \frac{2}{\pi_k^3}\phi''(p_k / \pi_k) \right\}_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}}$$

$$= \frac{1}{\pi_{0k}^3}\{\rho_k^2 \phi^{(4)}(\rho_k) + 4\rho_k \phi^{(3)}(\rho_k) + 2\phi''(\rho_k)\},$$

$$\frac{\partial^4 \pi_k \phi(p_k / \pi_k)}{\partial \pi_k^3 \partial p_k}\Big|_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}}$$

$$= \left\{ -\frac{p_k^3}{\pi_k^6}\phi^{(4)}(p_k / \pi_k) - 6\frac{p_k^2}{\pi_k^5}\phi^{(3)}(p_k / \pi_k) - 6\frac{p_k}{\pi_k^4}\phi''(p_k / \pi_k) \right\}_{\substack{p_k=\tau_k \\ \pi_k=\pi_{0k}}}$$

$$= -\frac{1}{\pi_{0k}^3}\{\rho_k^3\phi^{(4)}(\rho_k) + 6\rho_k^2\phi^{(3)}(\rho_k) + 6\rho_k\phi''(\rho_k)\}$$

$(k = 1,...,K)$.

Proof. The results are given by direct derivation. Q.E.D.

Lemma 2. *With the assumption of the existence of the derivatives of* $\phi(x)$ *up to the fourth order at* $x = \rho_k (k = 1,...,K)$,

$$\frac{\partial \hat{\boldsymbol{\theta}}}{\partial \mathbf{p}'}\Big|_{\substack{\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}_0 \\ \mathbf{p}=\boldsymbol{\tau}}} \equiv \frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\tau}'}$$

$$= \underset{(A)}{\Bigg[} \sum_{a=1}^{K} \left[\frac{\rho_a^2}{\pi_{0a}}\phi''(\rho_a)\frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0}\frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0'} + \{-\rho_a\phi'(\rho_a) + \phi(\rho_a)\}\frac{\partial^2 \pi_{0a}}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'} \right] \underset{(A)}{\Bigg]}^{-1} \quad \text{(A.2)}$$

$$\times \frac{\partial \boldsymbol{\pi}_0'}{\partial \boldsymbol{\theta}_0} \mathrm{diag}\left\{ \frac{\rho_1}{\pi_{01}}\phi''(\rho_1),...,\frac{\rho_K}{\pi_{0K}}\phi''(\rho_K) \right\},$$

$$\frac{\partial^2 \hat{\boldsymbol{\theta}}}{\partial p_i \partial p_j}\Big|_{\substack{\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}_0 \\ \mathbf{p}=\boldsymbol{\tau}}} \equiv \frac{\partial^2 \boldsymbol{\theta}_0}{\partial \tau_i \partial \tau_j}$$

$$= -\underset{(A)}{\Bigg[} \sum_{a=1}^{K} \left[\frac{\rho_a^2}{\pi_{0a}}\phi''(\rho_a)\frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0}\frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0'} + \{-\rho_a\phi'(\rho_a) + \phi(\rho_a)\}\frac{\partial^2 \pi_{0a}}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'} \right] \underset{(A)}{\Bigg]}^{-1}$$

$$\times \underset{(B)}{\Bigg[} \sum_{a=1}^{K} \underset{(C)}{\Bigg[} -\frac{1}{\pi_{0a}^2}\{\rho_a^3\phi^{(3)}(\rho_a) + 3\rho_a^2\phi''(\rho_a)\}\frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0}\left(\frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0'}\right)^{<2>}$$

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

$$+\{-\rho_a\phi'(\rho_a)+\phi(\rho_a)\}\frac{\partial^3\pi_{0a}}{\partial\boldsymbol{\theta}_0(\partial\boldsymbol{\theta}_0')^{<2>}}\Bigg]_{(C)}\frac{\partial\boldsymbol{\theta}_0}{\partial\tau_i}\otimes\frac{\partial\boldsymbol{\theta}_0}{\partial\tau_j}$$

$$+\sum_{(i,j)}^{2}\Bigg[_{(D)}\frac{1}{\pi_{0i}^2}\{\rho_i^2\phi^{(3)}(\rho_i)+2\rho_i\phi''(\rho_i)\}\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0}\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0'}-\frac{\rho_i}{\pi_{0i}}\phi''(\rho_i)\frac{\partial^2\pi_{0i}}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0'}\Bigg]_{(D)}\frac{\partial\boldsymbol{\theta}_0}{\partial\tau_j}$$

$$-\frac{\delta_{ij}}{\pi_{0i}^2}\{\rho_i\phi^{(3)}(\rho_i)+\phi''(\rho_i)\}\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0}\Bigg]_{(B)},$$

$$\frac{\partial^3\hat{\boldsymbol{\theta}}}{\partial p_i\partial p_j\partial p_k}\Big|_{\substack{\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}_0\\ \mathbf{p}=\boldsymbol{\tau}}}\equiv\frac{\partial^3\boldsymbol{\theta}_0}{\partial\tau_i\partial\tau_j\partial\tau_k}$$

$$=-\Bigg[_{(A)}\sum_{a=1}^{K}\Bigg[\frac{\rho_a^2}{\pi_{0a}}\phi''(\rho_a)\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0}\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0'}+\{-\rho_a\phi'(\rho_a)+\phi(\rho_a)\}\frac{\partial^2\pi_{0a}}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0'}\Bigg]\Bigg]_{(A)}^{-1}$$

$$\times\Bigg[_{(B)}\sum_{a=1}^{K}\Bigg[_{(C)}\Bigg[_{(D)}\frac{1}{\pi_{0a}^3}\{\rho_a^4\phi^{(4)}(\rho_a)+8\rho_a^3\phi^{(3)}(\rho_a)+12\rho_a^2\phi''(\rho_a)\}\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0}\left(\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0'}\right)^{<3>}$$

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

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

$$+\frac{\rho_a^2}{\pi_{0a}}\phi''(\rho_a)$$

$$\times\Bigg\{\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0}\otimes\frac{\partial^3\pi_{0a}}{(\partial\boldsymbol{\theta}_0')^{<3>}}+2\frac{\partial^2\pi_{0a}}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0'}\otimes\frac{\partial^2\pi_{0a}}{(\partial\boldsymbol{\theta}_0')^{<2>}}+\frac{\partial^3\pi_{0a}}{\partial\boldsymbol{\theta}_0(\partial\boldsymbol{\theta}_0')^{<2>}}\otimes\frac{\partial\pi_{0a}}{\partial\boldsymbol{\theta}_0'}\Bigg\}$$

$$+\{-\rho_a\phi'(\rho_a)+\phi(\rho_a)\}\frac{\partial^4\pi_{0a}}{\partial\boldsymbol{\theta}_0(\partial\boldsymbol{\theta}_0')^{<3>}}\Bigg]_{(D)}\frac{\partial\boldsymbol{\theta}_0}{\partial\tau_i}\otimes\frac{\partial\boldsymbol{\theta}_0}{\partial\tau_j}\otimes\frac{\partial\boldsymbol{\theta}_0}{\partial\tau_k}$$

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

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

$$\left.+\{-\rho_a\phi'(\rho_a)+\phi(\rho_a)\}\frac{\partial^3\pi_{0a}}{\partial\boldsymbol{\theta}_0(\partial\boldsymbol{\theta}_0')^{<2>}}\right]\frac{\partial\boldsymbol{\theta}_0}{\partial\tau_i}\otimes\frac{\partial^2\boldsymbol{\theta}_0}{\partial\tau_j\partial\tau_k}\Bigg]_{\text{(C)}}$$

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

$$+\frac{1}{\pi_{0i}^{2}}\{\rho_i^2\phi^{(3)}(\rho_i)+2\rho_i\phi''(\rho_i)\}\left\{\frac{\partial^2\pi_{0i}}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0'}\otimes\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0'}+\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0}\left(\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0'}\right)^{<2>}\right\}$$

$$\left.-\frac{\rho_i}{\pi_{0i}}\phi''(\rho_i)\frac{\partial^3\pi_{0i}}{\partial\boldsymbol{\theta}_0(\partial\boldsymbol{\theta}_0')^{<2>}}\right]\frac{\partial\boldsymbol{\theta}_0}{\partial\tau_j}\otimes\frac{\partial\boldsymbol{\theta}_0}{\partial\tau_k}$$

$$+\left[\frac{1}{\pi_{0i}^{2}}\{\rho_i^2\phi^{(3)}(\rho_i)+2\rho_i\phi''(\rho_i)\}\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0}\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0'}-\frac{\rho_i}{\pi_{0i}}\phi''(\rho_i)\frac{\partial^2\pi_{0i}}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0'}\right]\frac{\partial^2\boldsymbol{\theta}_0}{\partial\tau_j\partial\tau_k}$$

$$+\delta_{ij}\left[\frac{1}{\pi_{0i}^{3}}\{\rho_i^2\phi^{(4)}(\rho_i)+4\rho_i\phi^{(3)}(\rho_i)+2\phi''(\rho_i)\}\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0}\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0'}\right.$$

$$\left.-\frac{1}{\pi_{0i}^{2}}\{\rho_i\phi^{(3)}(\rho_i)+\phi''(\rho_i)\}\frac{\partial^2\pi_{0i}}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0'}\right]\frac{\partial\boldsymbol{\theta}_0}{\partial\tau_k}\Bigg]_{\text{(E)}}$$

$$-\delta_{ijk}\frac{1}{\pi_{0i}^{3}}\{\rho_i\phi^{(4)}(\rho_i)+2\phi^{(3)}(\rho_i)\}\frac{\partial\pi_{0i}}{\partial\boldsymbol{\theta}_0}\Bigg]_{\text{(B)}}\quad(i,j,k=1,...,K),$$

where e.g., $\underset{\text{(A)}}{[}\cdot\underset{\text{(A)}}{]}$ *is for ease of finding correspondence;* $\sum_{(i,j)}^{2}(\cdot)$ *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* $\delta_{ijk}\equiv\delta_{ij}\delta_{jk}$.

Proof. Recalling that $\partial\hat{D}_\phi/\partial\hat{\boldsymbol{\theta}}=\mathbf{0}$ with $\mathbf{0}$ being the $q\times1$ zero vector and using the formulas of the partial derivatives in implicit functions (see

Ogasawara, 2009, Equation (A.2)), we obtain

$$
\begin{aligned}
\frac{\partial \boldsymbol{\theta}_0}{\partial \boldsymbol{\tau}'} &= -\left(\frac{\partial^2 D_\phi}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'}\right)^{-1} \frac{\partial^2 D_\phi}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\tau}'} \\
&= -\underset{(\mathrm{A})}{\Bigg[} \sum_{a=1}^{K} \left[\frac{\rho_a^2}{\pi_{0a}} \phi''(\rho_a) \frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0} \frac{\partial \pi_{0a}}{\partial \boldsymbol{\theta}_0'} + \{-\rho_a \phi'(\rho_a) + \phi(\rho_a)\} \frac{\partial^2 \pi_{0a}}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'} \right] \underset{(\mathrm{A})}{\Bigg]}^{-1} \qquad \text{(A.3)} \\
&\quad \times \left\{ -\frac{\rho_1}{\pi_{01}} \phi''(\rho_1) \frac{\partial \pi_{01}}{\partial \boldsymbol{\theta}_0}, \ldots, -\frac{\rho_K}{\pi_{0K}} \phi''(\rho_K) \frac{\partial \pi_{0K}}{\partial \boldsymbol{\theta}_0} \right\},
\end{aligned}
$$

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

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

Using Lemma 1, (A.3) gives (A.2) in Lemma 2. Q.E.D.

Note that in the first equation of (A.3), $\underset{(\mathrm{A})}{[}\cdot\underset{(\mathrm{A})}{]}$ under correct model specification becomes $\mathbf{I}_0$ since $\rho_k = 1\ (k = 1, \ldots, K)$, $\phi''(1) = 1$, $\phi'(1) = 0$ and $\phi(1) = 0$.

S2. Additional numerical results under model misspecification

Table S1.1 gives the simulated and theoretical ratios i.e., Sim. = SD/ASE and Th. = HASE/ASE, where SD is the standard deviation from simulation, ASE $= n^{-1/2}\beta_2^{1/2}$ and HASE $= (n^{-1}\beta_2 + n^{-2}\beta_{\Delta 2})^{1/2}$. The large simulated values e.g., 5.446 for Case B when n = 200 and 5.501 for Case C when n = 50 by $\lambda = -2$ are due to unstable results in estimation. Except the results when n = 50, the values of the ratios are approximately 1.000.

Table S1.2 shows the simulated and theoretical (β_1) biases multiplied by *n*. In the table, "0" indicates an exactly zero value. Again, $\lambda = -2$ gives the largest biases while $\lambda = 0$ gives the smallest ones. The largest (absolute) β_1 is -4.12 for Case B when $\lambda = -2$, whose actual value when *n* = 50 is $-4.12/50 \doteq -.08$ while the corresponding ASE in Table 2 is $1.53/\sqrt{50} \doteq .22$. That is, the asymptotic bias is approximately 40 % of the ASE. When *n* becomes larger, the relative asymptotic bias becomes smaller since $n^{-1}\beta_1 / n^{-1/2}\beta_2^{1/2} = O(n^{-1/2})$.

Table S1.3 gives the simulated and theoretical ($\beta_3 / \beta_2^{3/2}$) skewness multiplied by $n^{1/2}$. The values in the table are mostly positive. It is of interest to see that the largest absolute values of $\beta_3 / \beta_2^{3/2}$ in Cases A to C are given by $\lambda = -1, -2$ and 1, respectively.

Table S1.4 shows the simulated and theoretical (β_4 / β_2^2) kurtoses multiplied by *n*. The methods by $\lambda = -1$ and -2 give unstable results when *n* = 50 and 200.

In Table S1.5, the simulated and theoretical (β_1^2, $\beta_{\Delta 2}$) values are shown. Note that the asymptotic mean square error up to order $O(n^{-1})$ is equal to the asymptotic variance $n^{-1}\beta_2$ and that the value up to order $O(n^{-2})$ is given by $n^{-1}\beta_2 + n^{-2}(\beta_1^2 + \beta_{\Delta 2})$. The simulated $\beta_{\Delta 2}$ is given by $n^2(\mathrm{SD}^2 - \mathrm{ASE}^2)$. The table shows the contribution of β_1^2 and $\beta_{\Delta 2}$ in the added value of order $O(n^{-2})$. It is found that on average, the relative contribution of β_1^2 is much smaller than that of $\beta_{\Delta 2}$. Recall that HASE/ASE when *n* = 200 and 800 are close to 1.000 in Table S1.1 indicating that the contribution of β_1^2 in $n^{-1}\beta_2 + n^{-2}(\beta_1^2 + \beta_{\Delta 2})$ is rather small.

The results of Table S1.6 to S1.8 are those for studentized MϕEs. Table S1.6 gives the simulated and theoretical ($\beta_2^{1/2}{}' = 1$) standard errors. The methods by $\lambda = -1$ and -2 give unstable results when *n* = 50 and 200. When *n* = 800, the simulated values are mostly close to 1.000 with some exceptions e.g., 1.034 for Case B by $\lambda = -2$.

Table S1.7 shows the simulated and theoretical ($\beta_1{}'$) biases of the studentized MϕEs multiplied by $n^{1/2}$. Most of the values are negative. The method by $\lambda = -1$ gives unstable results when n = 50. The largest absolute $\beta_1{}'$ is given by $\lambda = 2$ for Case A yielding $n^{-1/2}\beta_1{}' = -1.20/\sqrt{50} \doteq -.17$ when n = 50 which is approximately 20% of the unit ASE.

Table S1.8 gives the simulated and theoretical ($\beta_3{}'$) skewnesses multiplied by $n^{1/2}$ for the studentized MϕEs. The values are mostly negative. Recall that the corresponding results for the non-studentized MϕEs are mostly positive. This type of sign reversal typically happens after studentization. The methods by $\lambda = -1$ and -2 give unstable results when n = 50 or 200. The values by $\lambda = 0,\ 2/3,\ 1$ and 2 in Table S1.8 seem to be larger than those of the corresponding absolute values in Table S1.3.

The corresponding numerical results under correct model specification are given in Tables S2.1 to S2.10 of Ogasawara (2019b), where $\boldsymbol{\pi}_0$'s in Cases A, B, C and D are given by θ_0= .4, 1, 1.5 and $\boldsymbol{\theta}_0 = (1.5, .3549)'$, respectively. Note that $\boldsymbol{\pi}_0$ in Case C is the same as that in Case D. Though there are differences in the two sets of tables, the relative values among different λ's are similar in a crude sense.

S3. Applications of the asymptotic cumulants in interval estimation

The asymptotic cumulants derived earlier can typically be used for interval estimation. In this section, simulations for one-sided confidence intervals are shown under model misspecification. Interval estimation under model misspecification may seem odd. However, as mentioned earlier, the population value corresponding to a parameter estimator is reasonably defined as that when infinitely many observations are available. Consequently, it is also reasonable to estimate the population value under model misspecification. Note that in many or most of the cases in practice especially in the behavioral and social sciences, models are approximations to reality. In these cases, estimation of the population value under correct model specification becomes meaningless. Recall, however, that the population values corresponding to MϕEs with different $\phi(\cdot)$ s, when $\boldsymbol{\tau}(=\mathrm{E}(\mathbf{p}))$ is given with $\boldsymbol{\tau} \neq \boldsymbol{\pi}_0$ i.e., under model misspecification, are different from MϕE to MϕE. We use the following four lower endpoints in one-sided confidence intervals:

(i) the normal approximation by the Fisher information matrix (NF)

$$L^{(1)} = \hat{\theta} - n^{-1/2}\{(\hat{\mathbf{I}}^{-1})_{\theta\theta}\}^{1/2} z_\alpha\,, \tag{A.4}$$

(ii) the normal approximation by the robust ASE estimate (NR)

$$L^{(2)} = \hat{\theta} - n^{-1/2}\hat{\beta}_2^{1/2} z_\alpha\,, \tag{A.5}$$

(iii) the Cornish-Fisher expansion denoted by C-F (see Ogasawara, 2012, Equation (2.5))

$$L^{(3)} = \hat{\theta} - n^{-1/2}\hat{\beta}_2^{1/2} z_\alpha - n^{-1}\hat{\beta}_2^{1/2}\{\hat{\beta}_1' + (\hat{\beta}_3'/6)(z_\alpha^2 - 1)\} \quad \text{(A.6)}$$

and (iv) Hall's (1992) monotonic transformation with bias correction before cubic transformation denoted by Hall (see Ogasawara, 2012, Theorem 4)

$$L^{(4)} = \hat{\theta} - n^{-1}\hat{\beta}_2^{1/2}\hat{\beta}_1' - \frac{6\hat{\beta}_2^{1/2}}{\hat{\beta}_3'}\left[\left\{\frac{\hat{\beta}_3'}{2}\left(n^{-1/2}z_\alpha - n^{-1}\frac{\hat{\beta}_3'}{6}\right) - 1\right\}^{1/3} + 1\right], \quad \text{(A.7)}$$

where $\hat{\theta}$ is an MϕE; $\hat{\mathbf{I}}$ is a sample version of the Fisher information matrix per observation; $(\cdot)_{\theta\theta}$ indicates the diagonal element of a matrix corresponding to $\hat{\theta}$; $\alpha = \int_{-\infty}^{z_\alpha} (1/\sqrt{2\pi})\exp(-z^2/2)\,\mathrm{d}z$; $n^{-1/2}\hat{\beta}_2^{1/2}$ is a sample version of the robust ASE under possible model misspecification (see (2.2)); and $\hat{\beta}_1'$ and $\hat{\beta}_3'$ are sample versions of β_1' and β_3' (see (2.7)), respectively.

It is known that $L^{(1)}$ is invalid under model misspecification but is included for comparison while

$$\Pr(\theta_0 > L^{(2)}) = \alpha + O(n^{-1/2}), \quad \text{(A.8)}$$

and when we neglect the discreteness of a categorical variable, we have

$$\Pr(\theta_0 > L^{(i)}) = \alpha + O(n^{-1}) \ (i = 3, 4). \quad \text{(A.9)}$$

Simulations for interval estimations are performed. Tables S1.9 to S1.11 show selected results when n = 50 with $\lambda = -2$ (Neyman's statistic), 2/3 (the Cressie-Read statistic) and 2 for the proportions of a population value below the one-sided confidence intervals given by (A.4) to (A.7) with the number of replications being 10,000. the values of Z (the number of deleted cases with zero frequencies or empty cells) and NC (the number of deleted cases due to no-convergence) defined as before have been reduced due to the reduced number of replications.

In the tables, the results by NF look similar to those by NR in many points in a crude sense. However, when we look at the tables carefully, we find that NR improves NF as is expected. For instance, in Table S1.9 for Case B when .1000 is a nominal value, the corresponding proportions by NF and NR are .2077 and .1146, respectively. While among the results by NF, NR, C-F and Hall, no method gives best results under all conditions, overall C-F and Hall seem to give improvements over NF and NR.

The results for the confidence intervals under correct model specification are also available in Tables S2.11 to S2.13 of Ogasawara (2019b). The confidence intervals are constructed in the same manners as those under model misspecification except that $\hat{\beta}_1'$ and $\hat{\beta}_3'$ are given from the formulas of β_1' and β_3' under correct model specification, respectively. Note that $\hat{\beta}_2$ is given by the robust ASE estimate even under correct model specification.

Table S1.1. Simulated and theoretical ratios of the higher- and lower-order asymptotic standard errors for the MϕEs when models are misspecified

Case	n	50		200		800		50		200		800	
Parameter		Sim.	Th.	Sim.	Th.	Sim.	Th.	Sim.	Th.	Sim.	Th.	Sim.	Th.
		$\lambda = 0$ (G^2, ML)						$\lambda = -1$ (GM^2)					
A	θ	.933	1.013	.997	1.003	1.007	1.001	.969	1.037	1.003	1.009	1.007	1.002
B	θ	.986	.939	1.001	.985	1.003	.996	.700	1.003	1.015	1.001	1.005	1.000
C	θ	1.010	.964	1.000	.991	1.001	.998	1.144	1.008	1.010	1.002	1.003	1.001
D	θ_1	.992	.984	1.002	.996	.999	.999	1.017	1.006	1.008	1.001	1.000	1.000
	θ_2	.973	.991	1.002	.998	1.001	.999	.985	1.004	1.004	1.001	1.001	1.000
		$\lambda = -2$ (Neyman)						$\lambda = 2/3$ (C-R)					
A	θ	.992	1.082	1.015	1.021	1.007	1.005	.922	1.001	.995	1.000	1.007	1.000
B	θ	.707	1.244	5.446	1.066	.987	1.017	.990	.945	1.001	.987	1.003	.997
C	θ	5.501	1.101	1.032	1.026	1.008	1.007	.994	.946	.996	.987	1.000	.997
D	θ_1	1.417	1.029	1.014	1.007	1.002	1.002	.979	.971	.999	.993	.998	.998
	θ_2	1.001	1.017	1.007	1.004	1.002	1.001	.966	.983	1.000	.996	1.000	.999
		$\lambda = 1$ (X^2, Pearson)						$\lambda = 2$					
A	θ	.918	.995	.993	.999	1.006	1.000	.908	.980	990	.995	1.006	.999
B	θ	.991	.948	1.001	.987	1.003	.997	.995	.957	1.002	.990	1.003	.997
C	θ	.987	.939	.994	.985	1.000	.996	.973	.920	.990	.981	.999	.995
D	θ_1	.974	.964	.998	.991	.998	.998	.963	.948	.994	.987	.997	.997
	θ_2	.963	.979	.999	.995	1.000	.999	.956	.967	.997	.992	.999	.998

Note. n = the number of observations, Sim. = simulated value = SD/ASE, SD = the standard deviation from simulation, ASE = $n^{-1/2}\bar{\beta}_2^{1/2}$, Th. = theoretical value = HASE/ASE, HASE=$(n^{-1}\beta_2 + n^{-2}\beta_{\Delta 2})^{1/2}$, G^2 = the log-likelihood ratio statistic, GM^2 = the modified log-likelihood ratio statistic, Neyman = Neyman's statistic, C-R = the Cressie-Read statistic, X^2 = Pearson's statistic.

Table S1.2. Simulated and theoretical biases multiplied by n for the MϕEs when models are misspecified: β_1

Case	Parameter	Sim.(n) (50)	Sim.(n) (200)	Sim.(n) (800)	Th.	Sim.(n) (50)	Sim.(n) (200)	Sim.(n) (800)	Th.
		$\lambda = 0$ (G^2, ML)				$\lambda = -1$ (GM^2)			
A	θ	.54	.01	.05	0	.83	.22	.28	.20
B	θ	.41	.20	.06	.11	2.26	-.45	-.66	-.76
C	θ	.18	.25	.07	.04	.04	.12	-.08	-.15
D	θ_1	.85	.51	.26	.00	1.26	.88	.62	.36
	θ_2	.07	.06	-.03	-.04	.14	.12	.03	.02
		$\lambda = -2$ (Neyman)				$\lambda = 2/3$ (C-R)			
A	θ	1.06	.20	.30	.17	.40	-.12	-.09	-.14
B	θ	4.84	1.39	-2.26	-4.12	.38	.20	.08	.15
C	θ	1.61	-.39	-.60	-.70	.15	.21	.04	.02
D	θ_1	1.80	1.25	.97	.70	.59	.27	.02	-.24
	θ_2	.21	.19	.09	.08	.02	.02	-.07	-.08
		$\lambda = 1$ (X^2, Pearson)				$\lambda = 2$			
A	θ	.34	-.18	-.15	-.19	.22	-.30	-.29	-.32
B	θ	.35	.17	.06	.13	.26	.08	-.02	.05
C	θ	.11	.16	-.00	-.02	-.04	-.05	-.21	-.22
D	θ_1	.47	.15	-.10	-.36	.12	-.21	-.47	-.72
	θ_2	.00	-.00	-.09	-.10	-.05	-.06	-.16	-.17

Note. n = the number of observations, Sim. = simulated value, Th. = theoretical value = β_1, G^2 = the log-likelihood ratio statistic, GM^2 = the modified log-likelihood ratio statistic, Neyman = Neyman's statistic, C-R = the Cressie-Read statistic, X^2 = Pearson's statistic. The "0" indicates an exactly zero value.

Table S1.3. Simulated and theoretical skewnesses multiplied by $n^{1/2}$ for the MϕEs when models are misspecified: $\beta_3 / \beta_2^{3/2}$

Case	Parameter	Sim.(n) (50)	Sim.(n) (200)	Sim.(n) (800)	Th.	Sim.(n) (50)	Sim.(n) (200)	Sim.(n) (800)	Th.
		$\lambda = 0$ (G^2, ML)				$\lambda = -1$ (GM^2)			
A	θ	3.28	1.89	1.72	1.66	4.92	3.54	3.35	3.42
B	θ	2.46	2.16	2.80	1.63	159.3	1.09	1.87	-.31
C	θ	1.41	1.53	1.57	.67	131.2	1.41	1.52	.41
D	θ_1	2.55	2.06	2.11	.02	2.75	2.09	2.11	.01
	θ_2	2.33	1.39	1.79	1.31	2.41	1.40	1.81	1.31
		$\lambda = -2$ (Neyman)				$\lambda = 2/3$ (C-R)			
A	θ	5.43	2.45	2.17	2.27	2.96	1.52	1.35	1.24
B	θ	107.0	555.4	-1.70	-10.9	2.00	1.82	2.42	1.46
C	θ	163.6	.53	.96	-.39	1.36	1.50	1.51	.71
D	θ_1	274.1	2.16	2.13	.01	2.56	2.07	2.11	.03
	θ_2	2.58	1.42	1.82	1.31	2.31	1.39	1.79	1.31
		$\lambda = 1$ (X^2, Pearson)				$\lambda = 2$			
A	θ	2.88	1.43	1.26	1.15	2.76	1.33	1.17	1.05
B	θ	1.82	1.66	2.24	1.34	1.46	1.29	1.83	1.01
C	θ	1.35	1.47	1.47	.72	1.31	1.39	1.35	.68
D	θ_1	2.59	2.07	2.11	.03	2.70	2.10	2.12	.04
	θ_2	2.30	1.39	1.78	1.31	2.31	1.40	1.77	1.31

Note. n = the number of observations, Sim. = simulated value, Th. = theoretical value = $\beta_3 / \beta_2^{3/2}$, G^2 = the log-likelihood ratio statistic, GM^2 = the modified log-likelihood ratio statistic, Neyman = Neyman's statistic, C-R = the Cressie-Read statistic, X^2 = Pearson's statistic.

Table S1.4. Simulated and theoretical kurtoses multiplied by n for the MϕEs when models are misspecified: β_4 / β_2^2

Case	Parameter	Sim.(n) (50)	(200)	(800)	Th.	Sim.(n) (50)	(200)	(800)	Th.
		$\lambda = 0$ (G^2, ML)				$\lambda = -1$ (GM^2)			
A	θ	-.7	2.7	21.0	2.1	25.3	14.8	22.0	12.6
B	θ	5.1	8.0	5.2	-21.8	2.6e4	17.1	21.3	-13.7
C	θ	4.2	4.3	7.9	-14.5	1.1e5	5.5	9.5	-12.7
D	θ_1	12.4	13.3	15.0	-7.8	17.6	13.7	15.1	-9.5
	θ_2	-2.6	3.6	14.0	.4	-1.4	3.6	13.9	-1.0
		$\lambda = -2$ (Neyman)				$\lambda = 2/3$ (C-R)			
A	θ	48.2	21.3	29.7	25.7	-5.0	-1.4	19.7	-1.4
B	θ	1.2e4	3.2e5	-35.5	251.0	3.4	7.3	.1	-18.8
C	θ	2.8e4	19.5	18.4	-.5	4.0	4.2	7.0	-14.7
D	θ_1	1.7e5	15.0	15.4	-11.2	12.7	13.3	15.0	-6.7
	θ_2	2.7	3.8	13.8	-2.4	-2.8	3.7	14.0	1.3
		$\lambda = 1$ (X^2, Pearson)				$\lambda = 2$			
A	θ	-6.0	-2.6	18.9	-2.5	-7.6	-4.7	16.6	-4.5
B	θ	2.9	7.2	-1.5	-17.9	2.1	7.1	-4.0	-15.7
C	θ	3.9	4.2	6.6	-14.7	3.5	4.1	5.1	-15.0
D	θ_1	13.0	13.3	15.0	-6.2	13.9	13.4	15.0	-4.5
	θ_2	-2.7	3.7	14.1	1.8	-2.5	3.9	14.2	3.2

Note. n = the number of observations, Sim. = simulated value, Th. = theoretical value = β_4 / β_2^2, G^2 = the log-likelihood ratio statistic, GM^2 = the modified log-likelihood ratio statistic, Neyman = Neyman's statistic, C-R = the Cressie-Read statistic, X^2 = Pearson's statistic, $x\,\mathrm{e}\,y = x10^y$.

Table S1.5. Simulated and squared biases and added higher-order asymptotic biases multiplied by n^2 for the MϕEs when models are misspecified: β_1^2 and $\beta_{\Delta 2}$

Case	Parameter	β_1^2 Sim.(n) (50)	(200)	(800)	Th.	$\beta_{\Delta 2}$ Sim.(n) (50)	(200)	(800)	Th.= $\beta_{\Delta 2}$
		λ = -2 (Neyman)							
A	θ	1.11	.04	.09	.03	-.4	3.0	5.5	4.2
B	θ	23.47	1.94	5.12	16.98	5731	1.3e4	-47.1	64.0
C	θ	2.60	.15	.36	.49	2660	23.7	22.4	19.3
D	θ_1	3.24	1.56	.94	.49	86.5	10.0	4.8	5.1
	θ_2	.05	.04	.01	.01	.0	.4	.5	.3
		λ = 2/3 (C-R)							
A	θ	.16	.02	.01	.02	-2.7	-.8	3.8	.02
B	θ	.14	.04	.01	.02	-.9	.3	4.2	-4.7
C	θ	.02	.04	.00	.00	-1.1	-2.9	.4	-8.9
D	θ_1	.35	.07	.00	.06	-3.6	-.6	-5.6	-5.0
	θ_2	.00	.00	.01	.01	-.5	.0	.0	-.3
		λ = 2							
A	θ	.05	.09	.08	.10	-3.5	-1.6	3.5	-.8
B	θ	.07	.01	.00	.00	-.4	.6	4.5	-3.4
C	θ	.00	.00	.04	.05	-4.4	-6.7	-2.6	-12.8
D	θ_1	.02	.05	.22	.52	-6.3	-4.0	-9.3	-8.8
	θ_2	.00	.00	.02	.03	-.7	-.2	-.2	-.5

Note. n = the number of observations, Sim. = simulated value, Th. = theoretical value = β_1^2 or $\beta_{\Delta 2}$, Neyman = Neyman's statistic, C-R = the Cressie-Read statistic, x e $y = x10^y$.

Table S1.6. Simulated and theoretical standard errors of the studentized MϕEs when models are misspecified: $\beta_2^{1/2}$'

Case	Parameter	Sim.(n) (50)	(200)	(800)	Th.	Sim.(n) (50)	(200)	(800)	Th.
		$\lambda = 0$ (G^2, ML)				$\lambda = -1$ (GM^2)			
A	θ	.940	1.025	1.014	1	.930	1.032	1.014	1
B	θ	1.011	1.010	1.004	1	1.4e8	1.008	1.002	1
C	θ	1.036	1.005	1.002	1	4.1	1.006	1.002	1
D	θ_1	.982	1.003	.999	1	.978	1.002	.999	1
	θ_2	1.009	1.017	1.004	1	1.007	1.017	1.004	1
		$\lambda = -2$ (Neyman)				$\lambda = 2/3$ (C-R)			
A	θ	.874	.997	1.002	1	.945	1.025	1.014	1
B	θ	15.9	11.6	1.034	1	1.012	1.008	1.004	1
C	θ	10.9	.999	1.000	1	1.029	1.004	1.002	1
D	θ_1	1.676	1.001	.999	1	.981	1.003	.999	1
	θ_2	1.005	1.016	1.004	1	1.008	1.017	1.004	1
		$\lambda = 1$ (X^2, Pearson)				$\lambda = 2$			
A	θ	.948	1.026	1.014	1	.955	1.027	1.015	1
B	θ	1.012	1.007	1.004	1	1.010	1.005	1.004	1
C	θ	1.026	1.003	1.002	1	1.017	1.001	1.002	1
D	θ_1	.981	1.003	.999	1	.986	1.002	.999	1
	θ_2	1.008	1.017	1.004	1	1.007	1.017	1.004	1

Note. n = the number of observations, Sim. = simulated value, Th. = theoretical value = $\beta_2^{1/2}$' =1, G^2 = the log-likelihood ratio statistic, GM^2 = the modified log-likelihood ratio statistic, Neyman = Neyman's statistic, C-R = the Cressie-Read statistic, X^2 = Pearson's statistic, $x\text{e}\,y = x10^y$.

Table S1.7. Simulated and theoretical biases multiplied by $n^{1/2}$ for the studentized MϕEs when models are misspecified: $\beta_1{}'$

Case	Parameter	Sim.(n) (50)	(200)	(800)	Th.	Sim.(n) (50)	(200)	(800)	Th.
		$\lambda = 0$ (G^2, ML)				$\lambda = -1$ (GM^2)			
A	θ	.28	-.85	-.76	-.83	.43	-.93	-.79	-.90
B	θ	-.40	-.62	-.76	-.59	2.2e7	-.88	-1.08	-.67
C	θ	-.44	-.36	-.50	-.28	-.42	-.42	-.58	-.31
D	θ_1	.00	-.28	-.46	-.01	.28	-.00	-.19	.27
	θ_2	-.53	-.58	-.80	-.76	-.37	-.42	-.64	-.60
		$\lambda = -2$ (Neyman)				$\lambda = 2/3$ (C-R)			
A	θ	.74	-.48	-.31	-.47	.06	-.99	-.91	-.96
B	θ	6.39	2.90	-.94	.81	-.33	-.52	-.64	-.52
C	θ	2.42	-.55	-.75	-.42	-.46	-.39	-.52	-.32
D	θ_1	.57	.25	.07	.53	-.21	-.47	-.65	-.19
	θ_2	-.22	-.27	-.49	-.45	-.64	-.68	-.91	-.87
		$\lambda = 1$ (X^2, Pearson)				$\lambda = 2$			
A	θ	-.04	-1.06	-.99	-1.04	-.25	-1.24	-1.18	-1.20
B	θ	-.32	-.50	-.61	-.50	-.31	-.48	-.58	-.49
C	θ	-.49	-.42	-.55	-.36	-.59	-.57	-.68	-.52
D	θ_1	-.32	-.57	-.75	-.28	-.66	-.87	-1.03	-.57
	θ_2	-.69	-.74	-.96	-.92	-.86	-.90	-1.12	-1.08

Note. n = the number of observations, Sim. = simulated value, Th. = theoretical value = $\beta_1{}'$, G^2 = the log-likelihood ratio statistic, GM^2 = the modified log-likelihood ratio statistic, Neyman = Neyman's statistic, C-R = the Cressie-Read statistic, X^2 = Pearson's statistic, $x\text{e}\,y = x10^y$.

Table S1.8. Simulated and theoretical skewnesses multiplied by $n^{1/2}$ for the studentized MϕEs when models are misspecified: β_3'

Case	Parameter	Sim.(n) (50)	Sim.(n) (200)	Sim.(n) (800)	Th.	Sim.(n) (50)	Sim.(n) (200)	Sim.(n) (800)	Th.
		$\lambda = 0$ (G^2, ML)				$\lambda = -1$ (GM^2)			
A	θ	-.01	-3.60	-3.46	-3.32	-.45	-4.55	-4.42	-4.15
B	θ	-2.45	-2.89	-2.11	-2.57	323	-1.82	-.95	-.17
C	θ	-2.19	-1.85	-1.77	-1.17	2024	-1.69	-1.57	-.76
D	θ_1	-1.49	-2.01	-1.88	-.28	-1.48	-2.00	-1.87	-.01
	θ_2	-1.44	-3.16	-2.57	-2.63	-1.45	-3.17	-2.56	-2.62
		$\lambda = -2$ (Neyman)				$\lambda = 2/3$ (C-R)			
A	θ	1.17	-1.88	-2.17	-2.04	-.02	-3.47	-3.29	-3.17
B	θ	110	578	.13	10.11	-2.48	-2.66	-1.92	-2.57
C	θ	178	-1.02	-.89	.20	-2.20	-1.87	-1.80	-1.30
D	θ_1	506	-2.06	-1.88	.00	-1.66	-2.05	-1.90	-.04
	θ_2	-2.04	-3.20	-2.56	-2.62	-1.47	-3.16	-2.58	-2.63
		$\lambda = 1$ (X^2, Pearson)				$\lambda = 2$			
A	θ	-.06	-3.46	-3.26	-3.15	-.20	-3.49	-3.24	-3.13
B	θ	-2.44	-2.54	-1.81	-2.49	-2.28	-2.22	-1.53	-2.24
C	θ	-2.22	-1.87	-1.80	-1.34	-2.35	-1.88	-1.79	-1.39
D	θ_1	-1.78	-2.08	-1.91	-.04	-2.37	-2.20	-1.95	-.06
	θ_2	-1.49	-3.17	-2.58	-2.63	-1.57	-3.19	-2.60	-2.63

Note. n = the number of observations, Sim. = simulated value, Th. = theoretical value = β_3', G^2 = the log-likelihood ratio statistic, GM^2 = the modified log-likelihood ratio statistic, Neyman = Neyman's statistic, C-R = the Cressie-Read statistic, X^2 = Pearson's statistic.

Table S1.9. Proportions of a population value below the one-sided confidence intervals under model misspecification: n = 50 and $\lambda = -2$ (Neyman's statistic)

Case	Parameter	Method	.0050	.0250	.1000	.5000	.9000	.9750	.9950
			Nominal values						
A	θ		Z = 50, NC = 939						
		NF	.0108	.0344	.1118	.5366	.8841	.9476	.9659
		NR	.0028	.0173	.0902	.5366	.9569	.9971	1.0000
		C-F	.0016	.0222	.1026	.5433	.9513	.9975	1.0000
		Hall	.0016	.0218	.1010	.5452	.9515	.9975	1.0000
B	θ		Z = 0, NC = 334						
		NF	.0383	.0933	.2077	.4655	.7081	.7927	.8573
		NR	.0081	.0303	.1146	.4655	.8614	.9346	.9761
		C-F	.0118	.0352	.1089	.5136	.8805	.9503	.9781
		Hall	.0104	.0352	.1089	.5103	.8805	.9880	.9967
C	θ		Z = 0, NC = 22						
		NF	.0086	.0340	.1228	.4774	.8193	.9128	.9608
		NR	.0032	.0202	.0877	.4774	.8763	.9591	.9878
		C-F	.0021	.0181	.0834	.5034	.8911	.9664	.9892
		Hall	.0021	.0180	.0831	.5013	.8924	.9691	.9926
D	θ_1		Z = 0, NC = 349						
		NF	.0045	.0264	.1087	.5526	.9071	.9757	.9947
		NR	.0019	.0181	.0973	.5526	.9145	.9769	.9960
		C-F	.0004	.0102	.0835	.5013	.9143	.9767	.9969
		Hall	.0003	.0102	.0835	.5013	.9143	.9767	.9970
	θ_2								
		NF	.0029	.0205	.0907	.5100	.8776	.9629	.9908
		NR	.0020	.0167	.0859	.5100	.8829	.9641	.9924
		C-F	.0026	.0214	.0959	.5120	.9150	.9884	.9998
		Hall	.0023	.0211	.0953	.5120	.9194	.9976	1.0000

Note. NF = the normal approximation by the Fisher information matrix, NR = the normal approximation by the robust ASE estimate, C-F = the Cornish-Fisher expansion, Hall = Hall's (1992) monotonic cubic transformation, Z = the number of deleted cases with zero frequenc(ies), NC = the number of deleted case(s) due to non-convergence.

Table S1.10. Proportions of a population value below the one-sided confidence intervals under model misspecification: n = 50 and $\lambda = 2/3$ (the Cressie-Read statistic)

Case	Parameter	Method	Nominal values						
			.0050	.0250	.1000	.5000	.9000	.9750	.9950
A	θ		Z = 50, NC = 939						
		NF	.0008	.0077	.0526	.5181	.9442	.9982	1.0000
		NR	.0034	.0214	.0874	.5181	.9012	.9806	1.0000
		C-F	.0034	.0253	.1049	.5352	.9500	1.0000	1.0000
		Hall	.0033	.0252	.1049	.5352	.9557	1.0000	1.0000
B	θ		Z = 0, NC = 334						
		NF	.0027	.0252	.1144	.5105	.8538	.9447	.9792
		NR	.0016	.0130	.0887	.5105	.8786	.9588	.9828
		C-F	.0017	.0171	.1007	.5221	.9012	.9762	.9971
		Hall	.0014	.0171	.1007	.5221	.9012	.9839	.9992
C	θ		Z = 0, NC = 22						
		NF	.0026	.0209	.0968	.4955	.8760	.9575	.9861
		NR	.0025	.0183	.0904	.4955	.8812	.9585	.9864
		C-F	.0040	.0232	.1007	.4974	.8897	.9675	.9911
		Hall	.0039	.0229	.1006	.4974	.8898	.9686	.9923
D	θ_1		Z = 0, NC = 349						
		NF	.0019	.0160	.0831	.4956	.9030	.9727	.9947
		NR	.0019	.0155	.0880	.4956	.9048	.9683	.9930
		C-F	.0020	.0153	.0919	.5273	.9056	.9748	.9964
		Hall	.0013	.0151	.0919	.5273	.9056	.9748	.9964
	θ_2								
		NF	.0016	.0147	.0778	.4875	.8733	.9576	.9894
		NR	.0017	.0145	.0787	.4875	.8728	.9587	.9892
		C-F	.0040	.0246	.1012	.5099	.9058	.9878	1.0000
		Hall	.0038	.0241	.0991	.5099	.9074	.9953	1.0000

Note. NF = the normal approximation by the Fisher information matrix, NR = the normal approximation by the robust ASE estimate, C-F = the Cornish-Fisher expansion, Hall = Hall's (1992) monotonic cubic transformation, Z = the number of deleted cases with zero frequenc(ies), NC = the number of deleted cases due to non-convergence.

Table S1.11. Proportions of a population value below the one-sided confidence intervals under model misspecification: n = 50 and $\lambda = 2$

Case	Parameter	Method	.0050	.0250	.1000	.5000	.9000	.9750	.9950
			Nominal values						
A	θ		Z = 50, NC = 939						
		NF	.0010	.0080	.0529	.4978	.9313	.9960	.9999
		NR	.0033	.0196	.0821	.4978	.8886	.9748	.9996
		C-F	.0040	.0264	.1036	.5321	.9451	1.0000	1.0000
		Hall	.0040	.0264	.1036	.5321	.9470	1.0000	1.0000
B	θ		Z = 0, NC = 334						
		NF	.0015	.0163	.0925	.5083	8655	.9417	.9770
		NR	.0013	.0151	.0906	.5083	8756	.9592	.9864
		C-F	.0020	.0190	.0958	.5083	.8997	.9784	.9963
		Hall	.0020	.0190	.0958	.5083	.8997	.9813	.9980
C	θ		Z = 0, NC = 22						
		NF	.0019	.0183	.0879	.4896	.8800	.9592	.9877
		NR	.0022	.0156	.0828	.4896	.8806	.9586	.9847
		C-F	.0039	.0217	.1006	.4969	.8900	.9671	.9914
		Hall	.0038	.0216	.1006	.4969	.8909	.9713	.9931
D	θ_1		Z = 0, NC = 349						
		NF	.0016	.0126	.0793	.4801	.8942	.9717	.9948
		NR	.0013	.0122	.0754	.4801	.8838	.9638	.9890
		C-F	.0014	.0174	.0928	.5158	.9035	.9733	.9950
		Hall	.0014	.0173	.0928	.5158	.9035	.9744	.9953
	θ_2								
		NF	.0012	.0127	.0733	.4791	.8705	.9559	.9886
		NR	.0013	.0124	.0746	.4791	.8691	.9549	.9886
		C-F	.0043	.0251	.1019	.5046	.9004	.9860	.9999
		Hall	.0039	.0242	.1014	.5046	.9021	.9931	1.0000

Note. NF = the normal approximation by the Fisher information matrix, NR = the normal approximation by the robust ASE estimate, C-F = the Cornish-Fisher expansion, Hall = Hall's (1992) monotonic cubic transformation, Z = the number of deleted cases with zero frequenc(ies), NC = the number of deleted cases due to non-convergence.