

Third-order average local powers of Bartlett-type adjusted tests: Ordinary *versus* adjusted profile likelihood

Yoshihide Kakizawa

Faculty of Economics, Hokkaido University, Nishi 7, Kita 9, Kita-ku, Sapporo 060-0809, Japan

ARTICLE INFO

Article history:

Received 27 May 2015

Available online 23 September 2016

AMS subject classifications:

62E20

62F03

62F05

Keywords:

Adjusted profile likelihood

Average local power

Bartlett-type adjustment

Local alternative

Nuisance parameter

ABSTRACT

Statistical inference in the presence of a nuisance parameter is often based on profile likelihood. Because it is not a genuine likelihood function, several adjustments to the profile likelihood function for eliminating score/information bias were proposed in the 1980s and 1990s, under the so-called global parameter orthogonality. On the basis of Stern's (1997) adjusted profile likelihood, which is applicable even without the global parameter orthogonality, we discuss higher-order average local power properties after several Bartlett-type adjustments. It turns out that Rao's statistic arising from Stern's adjusted profile likelihood continues to enjoy desirable average local power properties, as in the ordinary likelihood inference. We also investigate, using a simulation, the performance of Rao's test, compared with the likelihood ratio test and Wald's test.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Statistical inference in the presence of a nuisance parameter is a widely encountered and fundamental issue. Suppose that $\mathbf{X}_1, \dots, \mathbf{X}_N$ are independent and identically distributed (i.i.d.) random vectors (taking values of \mathbb{R}^{d_X}) according to a parametric density $f(\mathbf{x}, \boldsymbol{\theta})$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_{(1)}^\top, \boldsymbol{\theta}_{(2)}^\top)^\top \in \boldsymbol{\Theta} = \boldsymbol{\Theta}_{(1)} \times \boldsymbol{\Theta}_{(2)}$ (say), where $\boldsymbol{\theta}_{(1)} = (\theta_1, \dots, \theta_{p_1})^\top$ is a parameter of interest, and the remaining parameter $\boldsymbol{\theta}_{(2)} = (\theta_{p_1+1}, \dots, \theta_{p_1+p_2})^\top$ is the nuisance parameter ($p = p_1 + p_2$). Let $\mathcal{L}^{(N)}(\boldsymbol{\theta}) = \sum_{i=1}^N \log f(\mathbf{X}_i, \boldsymbol{\theta})$ be the log-likelihood. A commonly used test statistic for testing $\boldsymbol{\theta}_{(1)} = \boldsymbol{\theta}_{(1)0}$ against $\boldsymbol{\theta}_{(1)} \neq \boldsymbol{\theta}_{(1)0}$, where $\boldsymbol{\theta}_{(1)0} \in \boldsymbol{\Theta}_{(1)}$ is specified, while $\boldsymbol{\theta}_{(2)} \in \boldsymbol{\Theta}_{(2)}$ remains unspecified, is the likelihood ratio (LR), for which the (unrestricted) maximum likelihood estimator (MLE) $\hat{\boldsymbol{\theta}}_{\text{ML}}^{(N)}$ of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and the restricted MLE $\tilde{\boldsymbol{\theta}}_{(2)\text{ML}}^{(N)}$ of $\boldsymbol{\theta}_{(2)} \in \boldsymbol{\Theta}_{(2)}$ under the constraint $\boldsymbol{\theta}_{(1)} = \boldsymbol{\theta}_{(1)0}$ are computed. Historically, this has often been interpreted through the so-called profile likelihood function

$$\mathcal{L}^{P(N)}(\boldsymbol{\theta}_{(1)}) = \sup_{\boldsymbol{\theta}_{(2)} \in \boldsymbol{\Theta}_{(2)}} \mathcal{L}^{(N)}(\boldsymbol{\theta}) = \mathcal{L}^{(N)}\{\check{\boldsymbol{\theta}}^{(N)}(\boldsymbol{\theta}_{(1)})\}, \quad \check{\boldsymbol{\theta}}^{(N)}(\boldsymbol{\theta}_{(1)}) = \begin{pmatrix} \boldsymbol{\theta}_{(1)} \\ \check{\boldsymbol{\theta}}_{(2)}^{(N)}(\boldsymbol{\theta}_{(1)}) \end{pmatrix}, \quad (1)$$

where $\check{\boldsymbol{\theta}}_{(2)}^{(N)}(\boldsymbol{\theta}_{(1)})$ is the MLE of $\boldsymbol{\theta}_{(2)} \in \boldsymbol{\Theta}_{(2)}$ for a given $\boldsymbol{\theta}_{(1)} \in \boldsymbol{\Theta}_{(1)}$. Then, noting that

$$\hat{\boldsymbol{\theta}}_{\text{ML}}^{(N)} = \begin{pmatrix} \hat{\boldsymbol{\theta}}_{(1)\text{ML}}^{(N)} \\ \hat{\boldsymbol{\theta}}_{(2)\text{ML}}^{(N)} \end{pmatrix} = \check{\boldsymbol{\theta}}^{(N)}(\hat{\boldsymbol{\theta}}_{(1)\text{ML}}^{(N)}) \quad \text{and} \quad \tilde{\boldsymbol{\theta}}_{\text{ML}}^{(N)} = \begin{pmatrix} \boldsymbol{\theta}_{(1)0} \\ \tilde{\boldsymbol{\theta}}_{(2)\text{ML}}^{(N)} \end{pmatrix} = \check{\boldsymbol{\theta}}^{(N)}(\boldsymbol{\theta}_{(1)0}),$$

E-mail address: kakizawa@econ.hokudai.ac.jp.

<http://dx.doi.org/10.1016/j.jmva.2016.09.006>

0047-259X/© 2016 Elsevier Inc. All rights reserved.

the LR test statistic $T_{\text{LR}}^{(N)}$ is given by

$$T_{\text{LR}}^{(N)} = 2\{\mathcal{L}^{(N)}(\hat{\theta}_{\text{ML}}^{(N)}) - \mathcal{L}^{(N)}(\hat{\theta}_{\text{ML}}^{(N)})\} = 2\{\mathcal{L}^{(N)}(\hat{\theta}_{(1)\text{ML}}^{(N)}) - \mathcal{L}^{(N)}(\theta_{(1)0})\}.$$

The latter form indicates that the statistical inference in the presence of the nuisance parameter $\theta_{(2)}$ can be performed as if the likelihood were given by $\mathcal{L}^{(N)}(\theta_{(1)})$. However, since the nuisance parameter $\theta_{(2)} \in \Theta_2$ is replaced by the MLE for a fixed value $\theta_{(1)} \in \Theta_{(1)}$ of the parameter of interest, the function $\mathcal{L}^{(N)}(\theta_{(1)})$ is not a genuine likelihood function, and has not only a score bias

$$E_{\theta}^{(N)}\left[\frac{\partial}{\partial \theta_a} \mathcal{L}^{(N)}(\theta_{(1)})\right] = O(1), \quad a = 1, \dots, p_1,$$

but also an information bias

$$E_{\theta}^{(N)}\left[\frac{\partial}{\partial \theta_{a_1}} \mathcal{L}^{(N)}(\theta_{(1)}) \frac{\partial}{\partial \theta_{a_2}} \mathcal{L}^{(N)}(\theta_{(1)}) + \frac{\partial^2}{\partial \theta_{a_1} \partial \theta_{a_2}} \mathcal{L}^{(N)}(\theta_{(1)})\right] = O(1), \quad a_1, a_2 \in \{1, \dots, p_1\},$$

where $E_{\theta}^{(N)}$ denotes the expectation under the θ -distribution $P_{\theta}^{(N)}$ of $\mathbf{X}_1, \dots, \mathbf{X}_N$.

Various adjustments to the profile likelihood function $\mathcal{L}^{(N)}(\theta_{(1)})$ were proposed in the 1980s and 1990s, including a conditional or marginal likelihood. See, e.g., [7,32] for such pseudo-likelihoods. To justify these adjustments, higher-order asymptotic theory has been considered, where the $N^{-1/2}$ -term is referred to as the $(i+1)$ th-order. Mukerjee and Chandra [29] and DiCiccio and Stern [9] studied the Bartlett correction to the adjusted profile LR test statistic, in order to improve the large sample χ^2 -approximation of the null distribution. Mukerjee [24] considered a Bartlett-type correction for a conditional version of Rao's statistic. Under a sequence of local alternatives, Mukerjee [23,25] and Ghosh and Mukerjee [11] discussed the third-order power comparison between the ordinary and conditional version in a class of test statistics. However, the above-mentioned results in this area, except for DiCiccio and Stern [9], were often developed in the scalar parameter case ($p_1 = 1$), because conditional or marginal likelihood approaches in the 1980s required restrictions:

- global parameter orthogonality, in the sense of Cox and Reid [8],
- explicit knowledge of an ancillary statistic, and so on.

Although there are specific statistical models in which global parameter orthogonality under a suitable reparameterization holds for the case $p_1 > 1$, such a reparameterization can only be guaranteed to exist, in general, when $p_1 = 1$.

We focus on the adjusted profile likelihood function suggested in Stern [33], which is applicable to a general framework in the sense that both the parameter of interest and the nuisance parameter are allowed to be multi-dimensional, and no assumption is made about global parameter orthogonality or curved exponentiality. To the best of knowledge, few studies have examined Stern's adjusted profile likelihood inference, exceptions being related second-order local power analyses [18,26,34], and the Bartlett correction to Stern's adjusted profile LR test statistic [10]. Our goal is to extend the previous third-order results of the ordinary likelihood inference [16,17,19,20] to those of Stern's adjusted profile likelihood inference.

The novel results are as follows. Firstly, in Section 2.3, we introduce a class $\mathcal{T}_{N,3,A} (\supset \mathcal{T}_{N,3})$ of test statistics for testing $\theta_{(1)} = \theta_{(1)0}$ against $\theta_{(1)} \neq \theta_{(1)0}$ (the subclass $\mathcal{T}_{N,3}$ is found in [16,17,19,20]), under which we can develop the third-order asymptotic testing theory, regardless of the ordinary likelihood inference (with $\mathbf{A}(\cdot) \equiv \mathbf{0}_p$) or Stern's adjusted profile likelihood inference (with $\mathbf{A}(\cdot) \equiv \mathbf{M}(\cdot)$). Here, the enlarged class $\bigcup_A \mathcal{T}_{N,3,A}$ under consideration covers the LR test statistic, several variations of Rao's and Wald's statistics using the expected/observed information, and Terrell's gradient test statistic [36]. In principle, it is worth dealing with a larger class than $\mathcal{T}_{N,3,A}$, because the expected Fisher information matrix can also be estimated via the so-called "outer-product type" and "sandwich type". Such a generalization, being more complicated than that of Kakizawa [18] for the second-order analysis, is not pursued further, for simplicity.

Secondly, beyond the Bartlett correction (see [1]), "the Bartlett-type correction" is indispensable since, under the null hypothesis, the statistic other than the LR test statistic is, in general, not Bartlett correctable. Here, Bartlett correctability (e.g., [3,13,21]) refers to the remarkable fact that the scalar multiplication to $T^{(N)} (\xrightarrow{d} \chi_f^2)$ in the form $T^{\text{Bart}(N)} = (1 - \rho/N)T^{(N)}$, for some constant ρ , makes the large sample χ^2 -approximation of the null distribution accurate. Recall that the Bartlett-type adjustment dates back to Chandra and Mukerjee [5], Cordeiro and Ferrari [6], and Taniguchi [35]. The present two forms (the GCF and GB; see Definitions 1 and 2) were discussed by Kakizawa [16,17,19,20] in the class $\mathcal{T}_{N,3}$ for the ordinary likelihood inference. In Proposition 1, under both the null hypothesis $\theta_{(1)} = \theta_{(1)0}$ and a sequence of local alternatives $\theta_{(1)} = \theta_{(1)0} + N^{-1/2}\mathbf{h}_{(1)}$, where $\mathbf{h}_{(1)} = (h_1, \dots, h_{p_1})^\top$, we present the third-order asymptotic expansions of the distributions of the Bartlett-type adjusted test statistics $T_A^{\text{GCF}(N)}$ and $T_A^{\text{GB}(N)}$, applied to the test statistic $T_A^{(N)} \in \mathcal{T}_{N,3,A}$. As a corollary, the necessary and sufficient condition for the Bartlett correctability members of the enlarged class $\bigcup_A \mathcal{T}_{N,3,A}$ is shown to be the same as that for the class $\mathcal{T}_{N,3}$ (see Theorem 1 of [16]), which is of independent interest.

Thirdly, within the class $\mathcal{T}_{N,3,A}$ of the test statistics, which links the ordinary likelihood inference (with $\mathbf{A}(\cdot) \equiv \mathbf{0}_p$) and Stern's adjusted profile likelihood inference (with $\mathbf{A}(\cdot) \equiv \mathbf{M}(\cdot)$), we are primarily concerned with the higher-order average local power properties after the above-mentioned Bartlett-type adjustments. Although the third-order point-by-point/average local power comparison among several Bartlett-type adjustments is not new (e.g., [15,27,28,30,31]) in the absence of nuisance parameter, our results are substantial extensions of a series of papers [16,17,19,20] to the enlarged class

$\mathcal{T}_{N,3,A}$ in the presence of nuisance parameter. Remarkably, Theorem 5 indicates that Rao's test continues to enjoy desirable average local power properties, as in Kakizawa [19,20].

The rest of this paper is organized as follows. Section 2 illustrates Stern's adjusted profile likelihood, and then considers a general class $\mathcal{T}_{N,3,A}$ of the test statistics for testing $\theta_{(1)} = \theta_{(1)0}$ against $\theta_{(1)} \neq \theta_{(1)0}$. In Section 3, we present the third-order asymptotic expansions of the distributions of the Bartlett-type adjusted test statistics $T_A^{GCF(N)}$ and $T_A^{GB(N)}$, applied to the test statistic $T_A^{(N)} \in \mathcal{T}_{N,3,A}$, under both the null hypothesis and a sequence of local alternatives. The third-order average local power properties after the Bartlett-type adjustments are given in Section 4. Section 5 contains a simulation study to assess the effectiveness of the Bartlett-type adjustments in the small sample case. Concluding remarks are given in Section 6.

2. Framework

2.1. Notation

We denote by $P_\theta^{(N)}$ the θ -distribution of $\mathbf{X}_1, \dots, \mathbf{X}_N$, which are i.i.d. random vectors (taking values of \mathbb{R}^{d_X}) according to a density $f(\mathbf{x}, \theta)$, $\theta \in \Theta \subset \mathbb{R}^p$. For any sequence $\{Y^{(N)}\}_{N \geq 1}$ of random variables $Y^{(N)} = g_N(\mathbf{X}_1, \dots, \mathbf{X}_N)$, we write $Y^{(N)} = o_\theta^{(N)}(q, \beta)$, if $P_\theta^{(N)}[|Y^{(N)}| > d(\log N)^\beta] = o(N^{-q})$ as $N \rightarrow \infty$, for some $d > 0$, $q \geq 0$, and $\beta \geq 0$.

Under the same regularity conditions as in Kakizawa [19], the partial derivative of the log-density $\log f(\mathbf{x}, \theta)$ with respect to θ is denoted by $\ell_{j_1 \dots j_R}(\mathbf{x}, \theta) = (\partial/\partial\theta_{j_1}) \dots (\partial/\partial\theta_{j_R}) \log f(\mathbf{x}, \theta)$, for $R \in \mathbb{N}$, $j_1, \dots, j_R \in \{1, \dots, p\}$. We introduce $I_R = j_1 \dots j_R$ for notational simplicity and denote the cumulants of the $\ell_{I_R}(\mathbf{X}, \theta)$'s with respect to $\mathbf{X} \sim f(\cdot, \theta)$ by

$$v_{I_{R_1} \dots I_{R_v}}(\theta) = \text{cum}_\theta[\ell_{I_{R_1}}(\mathbf{X}, \theta), \dots, \ell_{I_{R_v}}(\mathbf{X}, \theta)]. \quad (2)$$

We assume that

$$\begin{aligned} v_{j_1}(\theta) &= 0, & v_{j_1 j_2}(\theta) + v_{j_1 j_2}(\theta) &= 0, & v_{j_1 j_2 j_3}(\theta) + \langle 3 \rangle v_{j_1 j_2 j_3}(\theta) + v_{j_1 j_2 j_3}(\theta) &= 0, \\ v_{j_1 j_2 j_3 j_4}(\theta) + \langle 4 \rangle v_{j_1 j_2 j_3 j_4}(\theta) + \langle 3 \rangle v_{j_1 j_2 j_3 j_4}(\theta) + \langle 6 \rangle v_{j_1 j_2 j_3 j_4}(\theta) + v_{j_1 j_2 j_3 j_4}(\theta) &= 0 \end{aligned}$$

for all $\theta \in \Theta$, where $\langle n \rangle$ before a term with indices is a sum of n similar terms obtained by an index permutation. These Bartlett identities for the cumulants enable us to eliminate $v_{j_1 j_2}(\theta)$, $v_{j_1 j_2 j_3}(\theta)$, and $v_{j_1 j_2 j_3 j_4}(\theta)$ in subsequent calculations. We write

$$Z_{j_1 \dots j_R}^{(N)}(\theta) = \begin{cases} \frac{1}{N^{1/2}} \sum_{i=1}^N \ell_{j_1}(\mathbf{X}_i, \theta), & R = 1, \\ \frac{1}{N^{1/2}} \sum_{i=1}^N \{\ell_{j_1 \dots j_R}(\mathbf{X}_i, \theta) - v_{j_1 \dots j_R}(\theta)\}, & R = 2, 3, 4. \end{cases}$$

We want to test $\theta_{(1)} = \theta_{(1)0}$ against $\theta_{(1)} \neq \theta_{(1)0}$, where $\theta_{(1)0} \in \Theta_{(1)}$ is specified, while $\theta_{(2)} \in \Theta_{(2)}$ remains unspecified. According to the partition $\theta = (\theta_{(1)}^\top, \theta_{(2)}^\top)^\top$, we stack $Z_j^{(N)}(\theta)$ and $v_{j,k}(\theta) = -v_{jk}(\theta)$ as

$$[Z_j^{(N)}(\theta)]_{j=1, \dots, p} = \mathbf{Z}^{(N)}(\theta) = \begin{pmatrix} \mathbf{Z}_{(1)}^{(N)}(\theta) \\ \mathbf{Z}_{(2)}^{(N)}(\theta) \end{pmatrix} \quad (\text{the score vector})$$

and

$$[v_{j,k}(\theta)]_{j,k \in \{1, \dots, p\}} = \mathbf{v}(\theta) = \begin{pmatrix} \mathbf{v}_{(11)}(\theta) & \mathbf{v}_{(12)}(\theta) \\ \mathbf{v}_{(21)}(\theta) & \mathbf{v}_{(22)}(\theta) \end{pmatrix} \quad (\text{the (expected) Fisher information matrix}),$$

respectively. Similarly, we stack $J_{jk}^{(N)}(\theta) = -N^{-1} \sum_{i=1}^N \ell_{jk}(\mathbf{X}_i, \theta)$ as

$$[J_{jk}^{(N)}(\theta)]_{j,k \in \{1, \dots, p\}} = \mathbf{J}^{(N)}(\theta) = \begin{pmatrix} \mathbf{J}_{(11)}^{(N)}(\theta) & \mathbf{J}_{(12)}^{(N)}(\theta) \\ \mathbf{J}_{(21)}^{(N)}(\theta) & \mathbf{J}_{(22)}^{(N)}(\theta) \end{pmatrix} \quad (\text{the observed Fisher information matrix}).$$

No assumption is made about the global parameter orthogonality $\mathbf{v}_{(12)}(\cdot) \equiv \mathbf{O}_{p_1, p_2}$ (see [8]), where \mathbf{O}_{p_1, p_2} is $p_1 \times p_2$ zero matrix.

We assume that $\mathbf{v}(\theta)$ is positive definite for all $\theta \in \Theta$; in this case, $\mathbf{v}_{(22)}(\theta)$ is positive definite because it is the principal submatrix of $\mathbf{v}(\theta)$, and the Schur complement of $\mathbf{v}(\theta)$, defined by

$$\mathbf{v}_{(11 \cdot 2)}(\theta) = [\mathbf{v}_{(11 \cdot 2) a, a'}(\theta)]_{a, a' \in \{1, \dots, p_1\}} = \mathbf{v}_{(11)}(\theta) - \mathbf{v}_{(12)}(\theta) \mathbf{v}_{(22)}^{-1}(\theta) \mathbf{v}_{(21)}(\theta),$$

is nonsingular. We denote by $v^{j,k}(\theta)$ the (j, k) th element of $\mathbf{v}^{-1}(\theta)$, and by $v_{(11 \cdot 2)}^{a,b}(\theta)$ the (a, b) th element of $\mathbf{v}_{(11 \cdot 2)}^{-1}(\theta)$. Further, let $[v_{(22)}^{r,s}(\cdot)]_{r,s \in \{p_1+1, \dots, p\}}$ be the inverse of the matrix $\mathbf{v}_{(22)}(\cdot) = [v_{r,s}(\cdot)]_{r,s \in \{p_1+1, \dots, p\}}$. Note that

$$\mathbf{v}_{(11 \cdot 2)}(\theta) = \mathbf{g}^\top(\theta) \mathbf{v}(\theta) \mathbf{g}(\theta) \quad \text{and} \quad \mathbf{g}(\theta) \mathbf{v}_{(11 \cdot 2)}^{-1}(\theta) \mathbf{g}^\top(\theta) = \mathbf{v}^{-1}(\theta) - \begin{pmatrix} \mathbf{O}_{p_1, p_1} & \mathbf{O}_{p_1, p_2} \\ \mathbf{O}_{p_2, p_1} & \mathbf{v}_{(22)}^{-1}(\theta) \end{pmatrix},$$

where

$$\mathfrak{g}(\theta) = [\mathfrak{g}_{j,a}(\theta)]_{j=1,\dots,p; a=1,\dots,p_1} = \begin{pmatrix} \mathbf{I}_{p_1} \\ -\mathbf{v}_{(22)}^{-1}(\theta) \mathbf{v}_{(21)}(\theta) \end{pmatrix}$$

(hereafter, \mathbf{I}_{p_1} is $p_1 \times p_1$ identity matrix). We write $\mathbf{J}_{(11,2)}^{(N)}(\theta) = \mathbf{J}_{(11)}^{(N)}(\theta) - \mathbf{J}_{(12)}^{(N)}(\theta) \{\mathbf{J}_{(22)}^{(N)}(\theta)\}^{-1} \mathbf{J}_{(21)}^{(N)}(\theta)$.

Remark 1. (i) We employ the standard summation convention that if an index occurs twice in a product of two or more terms, it denotes the summation over all values that this index may assume. Throughout this paper, unless otherwise stated, we use the letters $\{j, k\}$ as indices of θ , running from 1 to p , the letters $\{a, b\}$ as indices of $\theta_{(1)}$, running from 1 to p_1 , and the letters $\{r, s\}$ as indices of $\theta_{(2)}$, running from $p_1 + 1$ to p . Thus, $Q_{\dots j \dots}(\cdot) \mathfrak{g}_{j,a}(\cdot) = Q_{\dots a \dots}(\cdot)$ (say) is understood to be $\sum_{j=1}^p Q_{\dots j \dots}(\cdot) \mathfrak{g}_{j,a}(\cdot)$. In what follows, we use the notation $C_{a_1 a_2 a_3}^{\mathfrak{g}}(\cdot) = C_{j_1 j_2 j_3}(\cdot) \prod_{i=1}^3 \mathfrak{g}_{j_i, a_i}(\cdot)$ and $C_{a_1 a_2, k_1 k_2}^{\mathfrak{g}}(\cdot) = C_{j_1 j_2, k_1 k_2}(\cdot) \prod_{i=1}^2 \mathfrak{g}_{j_i, a_i}(\cdot)$ (and similar notation $D_{a_1 a_2 a_3, k_1 \dots k_v}^{\mathfrak{g}}(\cdot)$, $D_{a_1 a_2 a_3, k_1 \dots k_v}^{\mathfrak{g}}(\cdot)$, $D_{a_1 a_2, k_1 k_2, k_3 k_4}^{\mathfrak{g}}(\cdot)$, etc.) for C -functions $C_{j_1 j_2 j_3}(\cdot)$ and $C_{j_1 j_2, k_1 k_2}(\cdot)$, and D -functions $D_{j_1 j_2 j_3 j_4}(\cdot)$, $D_{j_1 j_2 j_3, k_1 \dots k_v}(\cdot)$, and $D_{j_1 j_2, k_1 k_2, k_3 k_4}(\cdot)$; see [16,17,19,20].

(ii) We often use the pattern $Q_{\dots a \dots}(\theta) [\mathbf{v}_{(11,2)}^{-1}(\theta) \mathbf{z}_{(1)}(\theta)]_a$ for some $\mathbf{z}_{(1)}(\theta) \in \mathbb{R}^{p_1}$, where $[\mathbf{v}]_i$ denotes the i th element v_i of any vector \mathbf{v} . Such a pattern arises naturally as the matrix manipulation

$$\mathbf{v}^{-1}(\theta) \begin{pmatrix} \mathbf{z}_{(1)}(\theta) \\ \mathbf{0}_{p_2} \end{pmatrix} = \mathfrak{g}(\theta) \mathbf{v}_{(11,2)}^{-1}(\theta) \mathbf{z}_{(1)}(\theta).$$

(iii) For notational simplicity, we write $\widehat{Q} = Q(\widehat{\theta}_{\text{ML}}^{(N)})$, $\widetilde{Q} = Q(\widetilde{\theta}_{\text{ML}}^{(N)})$, and $\check{Q}(\theta_{(1)}) = Q\{\check{\theta}^{(N)}(\theta_{(1)})\}$ for any (nonrandom/random) scalar or vector or matrix function $Q(\cdot)$.

2.2. Stern's adjusted profile likelihood inference

In order to eliminate the score bias of the profile likelihood (1) (see the introduction), Stern [33] considered

$$\begin{aligned} \mathcal{L}^{\text{AP}(N)}(\theta_{(1)}) &= \mathcal{L}^{\text{P}(N)}(\theta_{(1)}) - \frac{1}{N^{1/2}} \check{\mathbf{M}}^{\top}(\theta_{(1)}) \check{\mathbf{v}}^{-1}(\theta_{(1)}) \begin{pmatrix} \check{\mathbf{z}}_{(1)}^{(N)}(\theta_{(1)}) \\ \mathbf{0}_{p_2} \end{pmatrix} \\ &= \mathcal{L}^{\text{P}(N)}(\theta_{(1)}) - \frac{1}{N^{1/2}} \check{M}_{b_1}^{\mathfrak{g}}(\theta_{(1)}) [\check{\mathbf{v}}_{(11,2)}^{-1}(\theta_{(1)}) \check{\mathbf{z}}_{(1)}^{(N)}(\theta_{(1)})]_{b_1}, \end{aligned} \quad (3)$$

where

$$M_j(\cdot) = \frac{1}{2} v_{(22)}^{r,r'}(\cdot) \{v_{rr',j}(\cdot) + v_{r,r',j}(\cdot)\}, \quad j = 1, \dots, p \text{ (we stack } M_j(\cdot) \text{ as } \mathbf{M}(\cdot) = [M_j(\cdot)]_{j=1,\dots,p}).$$

Such an additive adjustment is applicable to a general framework, because both the parameter of interest and the nuisance parameter are allowed to be multi-dimensional, and no assumption is made about the global parameter orthogonality.

Recall that the ordinary likelihood-based test statistics for testing $\theta_{(1)} = \theta_{(1)0}$ against $\theta_{(1)} \neq \theta_{(1)0}$ are the LR test statistic, variants of Rao's and Wald's statistics, and the gradient test statistic, defined by

$$\begin{aligned} T_{\text{LR}}^{(N)} &= 2(\widehat{\mathcal{L}}^{(N)} - \widetilde{\mathcal{L}}^{(N)}), \\ T_{\text{R}}^{(N)} &= (\widetilde{\mathbf{Z}}_{(1)}^{(N)})^{\top} \widetilde{\mathbf{v}}_{(11,2)}^{-1} \widetilde{\mathbf{Z}}_{(1)}^{(N)}, & T_{\text{W}}^{(N)} &= N(\widehat{\theta}_{(1)\text{ML}}^{(N)} - \theta_{(1)0})^{\top} \widehat{\mathbf{v}}_{(11,2)}(\widehat{\theta}_{(1)\text{ML}}^{(N)} - \theta_{(1)0}), \\ T_{\text{MR}}^{(N)} &= (\widetilde{\mathbf{Z}}_{(1)}^{(N)})^{\top} \widetilde{\mathbf{v}}_{(11,2)}^{-1} \widetilde{\mathbf{Z}}_{(1)}^{(N)}, & T_{\text{MW}}^{(N)} &= N(\widehat{\theta}_{(1)\text{ML}}^{(N)} - \theta_{(1)0})^{\top} \widehat{\mathbf{v}}_{(11,2)}(\widehat{\theta}_{(1)\text{ML}}^{(N)} - \theta_{(1)0}), \\ T_{\text{R}^{\text{ob}}}^{(N)} &= (\widetilde{\mathbf{Z}}_{(1)}^{(N)})^{\top} (\widetilde{\mathbf{J}}_{(11,2)}^{(N)})^{-1} \widetilde{\mathbf{Z}}_{(1)}^{(N)}, & T_{\text{W}^{\text{ob}}}^{(N)} &= N(\widehat{\theta}_{(1)\text{ML}}^{(N)} - \theta_{(1)0})^{\top} \widetilde{\mathbf{J}}_{(11,2)}^{(N)}(\widehat{\theta}_{(1)\text{ML}}^{(N)} - \theta_{(1)0}), \\ T_{\text{MR}^{\text{ob}}}^{(N)} &= (\widetilde{\mathbf{Z}}_{(1)}^{(N)})^{\top} (\widetilde{\mathbf{J}}_{(11,2)}^{(N)})^{-1} \widetilde{\mathbf{Z}}_{(1)}^{(N)}, & T_{\text{MW}^{\text{ob}}}^{(N)} &= N(\widehat{\theta}_{(1)\text{ML}}^{(N)} - \theta_{(1)0})^{\top} \widetilde{\mathbf{J}}_{(11,2)}^{(N)}(\widehat{\theta}_{(1)\text{ML}}^{(N)} - \theta_{(1)0}), \\ T_{\text{grad}}^{(N)} &= (\widetilde{\mathbf{Z}}_{(1)}^{(N)})^{\top} N^{1/2}(\widehat{\theta}_{(1)\text{ML}}^{(N)} - \theta_{(1)0}). \end{aligned} \quad (4)$$

There are many ways to estimate the Fisher information matrix. We need to distinguish Rao's and Wald's statistics carefully, as follows. $\text{R}(\text{R}^{\text{ob}})$, depending only on the restricted MLE, is Rao's statistic using the expected (observed) information, whereas $\text{W}(\text{W}^{\text{ob}})$, depending only on the unrestricted MLE, is Wald's statistic using the expected (observed) information. Here, the variants using outer-product/sandwich type estimators are not pursued for simplicity, because, as mentioned in the introduction, such a generalization would necessarily require additional formulas. On the other hand, MR/MW ($\text{MR}^{\text{ob}}/\text{MW}^{\text{ob}}$) are the abbreviations of modified Rao/Wald statistics using the expected (observed) information, for which both the restricted MLE and the unrestricted MLE are computed, as in the case of the LR and gradient test statistics.

In keeping with the analogy to these test statistics (4) from the ordinary likelihood, and noting that $\check{\theta}^{(N)}(\theta_{(1)0}) = \widetilde{\theta}_{\text{ML}}^{(N)}$ and $\check{\theta}^{(N)}(\widehat{\theta}_{(1)\text{ML}}^{(N)}) = \widehat{\theta}_{\text{ML}}^{(N)}$, an adjusted profile version of the LR, Rao, Wald, and gradient test statistics can be constructed on

the basis of (3), as follows.

$$\begin{aligned}
 T_{LR}^{AP(N)} &= 2\{\mathcal{L}^{AP(N)}(\bar{\boldsymbol{\theta}}_{(1)}^{(N)}) - \mathcal{L}^{AP(N)}(\boldsymbol{\theta}_{(1)0})\}, \\
 T_R^{AP(N)} &= \frac{1}{N} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}_{(1)}} \mathcal{L}^{AP(N)}(\boldsymbol{\theta}_{(1)0}) \right\}^\top \check{\mathbf{v}}_{(11-2)}^{-1}(\boldsymbol{\theta}_{(1)0}) \frac{\partial}{\partial \boldsymbol{\theta}_{(1)}} \mathcal{L}^{AP(N)}(\boldsymbol{\theta}_{(1)0}), \\
 T_W^{AP(N)} &= N(\bar{\boldsymbol{\theta}}_{(1)}^{(N)} - \boldsymbol{\theta}_{(1)0})^\top \check{\mathbf{v}}_{(11-2)}(\bar{\boldsymbol{\theta}}_{(1)}^{(N)})(\bar{\boldsymbol{\theta}}_{(1)}^{(N)} - \boldsymbol{\theta}_{(1)0}), \\
 T_{MR}^{AP(N)} &= \frac{1}{N} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}_{(1)}} \mathcal{L}^{AP(N)}(\boldsymbol{\theta}_{(1)0}) \right\}^\top \check{\mathbf{v}}_{(11-2)}^{-1}(\bar{\boldsymbol{\theta}}_{(1)}^{(N)}) \frac{\partial}{\partial \boldsymbol{\theta}_{(1)}} \mathcal{L}^{AP(N)}(\boldsymbol{\theta}_{(1)0}), \\
 T_{MW}^{AP(N)} &= N(\bar{\boldsymbol{\theta}}_{(1)}^{(N)} - \boldsymbol{\theta}_{(1)0})^\top \check{\mathbf{v}}_{(11-2)}(\boldsymbol{\theta}_{(1)0})(\bar{\boldsymbol{\theta}}_{(1)}^{(N)} - \boldsymbol{\theta}_{(1)0}), \\
 T_{\text{grad}}^{AP(N)} &= \left\{ \frac{\partial}{\partial \boldsymbol{\theta}_{(1)}} \mathcal{L}^{AP(N)}(\boldsymbol{\theta}_{(1)0}) \right\}^\top (\bar{\boldsymbol{\theta}}_{(1)}^{(N)} - \boldsymbol{\theta}_{(1)0}),
 \end{aligned} \tag{5}$$

where $\bar{\boldsymbol{\theta}}_{(1)}^{(N)}$ is the adjusted profile MLE

$$\mathcal{L}^{AP(N)}(\bar{\boldsymbol{\theta}}_{(1)}^{(N)}) = \sup_{\boldsymbol{\theta}_{(1)} \in \Theta_{(1)}} \mathcal{L}^{AP(N)}(\boldsymbol{\theta}_{(1)}).$$

The observed variants $T_{Rob}^{AP(N)}$, $T_{Wob}^{AP(N)}$, $T_{MRob}^{AP(N)}$, and $T_{MWob}^{AP(N)}$ are similarly defined after replacing $\check{\mathbf{v}}_{(11-2)}(\cdot)$ with $\check{\mathbf{J}}_{(11-2)}^{(N)}(\cdot)$, and, in principle, $\check{\mathbf{J}}_{(11-2)}^{(N)}(\cdot)$ can be replaced by $[-\frac{\partial^2}{\partial \theta_{a_1} \partial \theta_{a_2}} \mathcal{L}^{AP(N)}(\cdot)]_{a_1, a_2 \in \{1, \dots, p_1\}}$.

2.3. A class of test statistics admitting the stochastic expansion

Let $\mathbf{A}(\cdot) = [A_j(\cdot)]_{j=1, \dots, p}$, where functions $A_j(\cdot)$ are of class $\mathcal{C}^2(\boldsymbol{\Theta})$. In principle, one may consider

$$\begin{aligned}
 \mathcal{L}_A^{(N)}(\boldsymbol{\theta}_{(1)}) &= \mathcal{L}^{P(N)}(\boldsymbol{\theta}_{(1)}) - \frac{1}{N^{1/2}} \check{\mathbf{A}}^\top(\boldsymbol{\theta}_{(1)}) \check{\mathbf{v}}^{-1}(\boldsymbol{\theta}_{(1)}) \begin{pmatrix} \check{\mathbf{Z}}_{(1)}^{(N)}(\boldsymbol{\theta}_{(1)}) \\ \mathbf{0}_{p_2} \end{pmatrix} \\
 &= \mathcal{L}^{P(N)}(\boldsymbol{\theta}_{(1)}) - \frac{1}{N^{1/2}} \check{A}_{b_1}^g(\boldsymbol{\theta}_{(1)}) [\check{\mathbf{v}}_{(11-2)}^{-1}(\boldsymbol{\theta}_{(1)}) \check{\mathbf{Z}}_{(1)}^{(N)}(\boldsymbol{\theta}_{(1)})]_{b_1}.
 \end{aligned}$$

Then, the A -variant of the test statistics $T_{LR,A}^{(N)}$, $T_{R,A}^{(N)}$, $T_{W,A}^{(N)}$, $T_{MR,A}^{(N)}$, $T_{MW,A}^{(N)}$, $T_{Rob,A}^{(N)}$, $T_{Wob,A}^{(N)}$, $T_{MRob,A}^{(N)}$, $T_{MWob,A}^{(N)}$, and $T_{\text{grad},A}^{(N)}$ are defined in (5), with $\mathcal{L}^{AP(N)}(\boldsymbol{\theta}_{(1)}) = \mathcal{L}_M^{(N)}(\boldsymbol{\theta}_{(1)})$ replaced by $\mathcal{L}_A^{(N)}(\boldsymbol{\theta}_{(1)})$. As mentioned in Section 2.2, $\mathbf{A}(\cdot) = \mathbf{M}(\cdot)$ has been suggested by Stern [33] for eliminating score bias. We emphasize that, at this stage, it is not clear which choice of $\mathbf{A}(\cdot) \neq \mathbf{0}_p$ is appropriate from the point of view of higher-order asymptotic testing theory. In Remark 5 of Section 3, we give another interpretation for choosing $\mathbf{A}(\cdot) = \mathbf{M}(\cdot)$. Section 4.3, however, reveals no definitive guide about $\mathbf{A}(\cdot)$, in general.

After tedious algebra (see also [9,16]), we can show that each of the (A -variant) test statistics belongs to a class $\mathcal{T}_{N,3,A}$, as follows. Every test statistic $T_A^{(N)} = T_N(\mathbf{X}_1, \dots, \mathbf{X}_N; \boldsymbol{\theta}_{(1)0}) \in \mathcal{T}_{N,3,A}$ for testing $\boldsymbol{\theta}_{(1)} = \boldsymbol{\theta}_{(1)0}$ against $\boldsymbol{\theta}_{(1)} \neq \boldsymbol{\theta}_{(1)0}$ admits a stochastic expansion of the form

$$\begin{aligned}
 T_A^{(N)} &= (\tilde{\mathbf{Z}}_{(1)}^{(N)})^\top \tilde{\mathbf{v}}_{(11-2)}^{-1} \tilde{\mathbf{Z}}_{(1)}^{(N)} \\
 &+ \frac{2}{N^{1/2}} \left(\tilde{C}_{b_1 b_2 b_3}^g \prod_{i=1}^3 [\tilde{\mathbf{v}}_{(11-2)}^{-1} \tilde{\mathbf{Z}}_{(1)}^{(N)}]_{b_i} + \tilde{C}_{b_1 b_2, k_1 k_2}^g \tilde{Z}_{k_1 k_2}^{(N)} \prod_{i=1}^2 [\tilde{\mathbf{v}}_{(11-2)}^{-1} \tilde{\mathbf{Z}}_{(1)}^{(N)}]_{b_i} + \tilde{A}_{b_1}^g [\tilde{\mathbf{v}}_{(11-2)}^{-1} \tilde{\mathbf{Z}}_{(1)}^{(N)}]_{b_1} \right) \\
 &+ \frac{2}{N} \left\{ \tilde{D}_{b_1 b_2 b_3 b_4}^g \prod_{i=1}^4 [\tilde{\mathbf{v}}_{(11-2)}^{-1} \tilde{\mathbf{Z}}_{(1)}^{(N)}]_{b_i} + (\tilde{D}_{b_1 b_2 b_3, k_1 k_2}^g \tilde{Z}_{k_1 k_2}^{(N)} + \tilde{D}_{b_1 b_2 b_3, k_1 k_2 k_3}^g \tilde{Z}_{k_1 k_2 k_3}^{(N)}) \prod_{i=1}^3 [\tilde{\mathbf{v}}_{(11-2)}^{-1} \tilde{\mathbf{Z}}_{(1)}^{(N)}]_{b_i} \right. \\
 &+ (\tilde{D}_{b_1 b_2, k_1 k_2, k_3 k_4}^g \tilde{Z}_{k_1 k_2}^{(N)} \tilde{Z}_{k_3 k_4}^{(N)} + {}_A \tilde{D}_{b_1 b_2}^g) \prod_{i=1}^2 [\tilde{\mathbf{v}}_{(11-2)}^{-1} \tilde{\mathbf{Z}}_{(1)}^{(N)}]_{b_i} \\
 &+ {}_A \tilde{D}_{b_1, k_1 k_2}^g \tilde{Z}_{k_1 k_2}^{(N)} [\tilde{\mathbf{v}}_{(11-2)}^{-1} \tilde{\mathbf{Z}}_{(1)}^{(N)}]_{b_1} + \left. \frac{1}{2} \tilde{A}_{b_1}^g \tilde{\mathbf{v}}_{(11-2)}^{b_1, b_2} \tilde{A}_{b_2}^g \right\} \\
 &+ \frac{1}{N^{3/2}} o_{\theta^\dagger}^{(N)}(1 + \xi, \beta) \quad \text{for some fixed } \beta, \xi > 0, \text{ with } \theta^\dagger = \begin{pmatrix} \boldsymbol{\theta}_{(1)0} \\ \boldsymbol{\theta}_{(2)} \end{pmatrix}
 \end{aligned} \tag{6}$$

(hereafter, $\theta_{(2)}^\dagger \in \boldsymbol{\Theta}_{(2)}$ is the irrelevant true value of the nuisance parameter $\boldsymbol{\theta}_{(2)}$), where the A D-functions ${}_A D_{j_1 j_2}(\cdot) = {}_A D_{j_2 j_1}(\cdot)$ and ${}_A D_{j_1, k_1 k_2}(\cdot)$, which may vary from one test statistic to another, are of class $\mathcal{C}^1(\boldsymbol{\Theta})$. Throughout this paper, the notation with pre-subscript A (e.g., ${}_A D_{j_1 j_2}(\cdot)$) indicates that the A D-functions vanish if $\mathbf{A}(\cdot) \equiv \mathbf{0}_p$ (in that case, (6) reduces to the test

statistic belonging to the class $\mathcal{T}_{N,3}$ in [16,17,19,20] for the ordinary likelihood inference). Therefore, we can develop the third-order asymptotic testing theory, regardless of the ordinary likelihood inference (with $\mathbf{A}(\cdot) \equiv \mathbf{0}_p$) or Stern's adjusted profile likelihood inference (with $\mathbf{A}(\cdot) = \mathbf{M}(\cdot)$). Remark 1(ii) will help to understand terms such as $\tilde{C}_{b_1 b_2 b_3}^{\mathcal{G} \mathcal{G} \mathcal{G}} \prod_{i=1}^3 [\tilde{\mathbf{v}}_{(11,2)}^{-1} \tilde{\mathbf{Z}}_{(1)}^{(N)}]_{b_i}$.

Remark 2. For the class $\mathcal{T}_{N,3,A}$ above, we assume that the C (or D)-functions, which may vary from one test statistic to another, are of class $\mathcal{C}^2(\Theta)$ (or $\mathcal{C}^1(\Theta)$). Further, without loss of generality, we assume that $C_{j_1 j_2 j_3}(\cdot)$, $C_{j_1 j_2, k_1 k_2}(\cdot)$, $D_{j_1 j_2 j_3 j_4}(\cdot)$, $D_{j_1 j_2 j_3, k_1 \dots k_4}(\cdot)$, and $D_{j_1 j_2, k_1 k_2, k_3 k_4}(\cdot)$ are symmetric under permutation of $\{j_1, j_2, j_3, j_4\}$, and that $D_{j_1 j_2, k_1 k_2, k_3 k_4}(\cdot) = D_{j_1 j_2, k_3 k_4, k_1 k_2}(\cdot)$. We define

$$C_{a_1 a_2 a_3}^{+\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) = C_{a_1 a_2 a_3}^{\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) + \frac{\langle 3 \rangle}{3} C_{a_1 a_2, k_1 k_2}^{\mathcal{G} \mathcal{G}}(\cdot) v_{k_1 k_2, a_3}^{\mathcal{G}}(\cdot), \quad a_1, a_2, a_3 \in \{1, \dots, p_1\}. \quad (7)$$

It turns out (see [16,18]) that the C^+ -functions for the A -variant of the LR, Rao, Wald, and gradient test statistics are given by

$$\begin{aligned} \text{LR} C_{a_1 a_2 a_3}^{+\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) &= -\frac{1}{6} v_{a_1, a_2, a_3}^{\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot), \quad \text{R} C_{a_1 a_2 a_3}^{+\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) \equiv 0, \\ \text{W} C_{a_1 a_2 a_3}^{+\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) &= \text{Rob} C_{a_1 a_2 a_3}^{+\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) = \text{Wob} C_{a_1 a_2 a_3}^{+\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) = \frac{1}{6} \langle 3 \rangle v_{a_1 a_2, a_3}^{\mathcal{G} \mathcal{G}}(\cdot), \\ \text{MR} C_{a_1 a_2 a_3}^{+\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) &= \text{MRob} C_{a_1 a_2 a_3}^{+\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) = \text{MWob} C_{a_1 a_2 a_3}^{+\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) = -\frac{1}{6} \{2 \langle 3 \rangle v_{a_1 a_2, a_3}^{\mathcal{G} \mathcal{G}}(\cdot) + 3 v_{a_1, a_2, a_3}^{\mathcal{G} \mathcal{G}}(\cdot)\}, \\ \text{MW} C_{a_1 a_2 a_3}^{+\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) &= -\frac{1}{6} \{\langle 3 \rangle v_{a_1 a_2, a_3}^{\mathcal{G} \mathcal{G}}(\cdot) + 3 v_{a_1, a_2, a_3}^{\mathcal{G} \mathcal{G}}(\cdot)\}, \\ \text{grad} C_{a_1 a_2 a_3}^{+\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) &= -\frac{1}{12} \{\langle 3 \rangle v_{a_1 a_2, a_3}^{\mathcal{G} \mathcal{G}}(\cdot) + 3 v_{a_1, a_2, a_3}^{\mathcal{G} \mathcal{G}}(\cdot)\}. \end{aligned}$$

Remark 3. As in Kakizawa [19,20], a class $\mathcal{T}_{N,3,A}^c$ can be defined as a subclass of $\mathcal{T}_{N,3,A}$ consisting of those members in $\mathcal{T}_{N,3,A}$ for which $C_{a_1 a_2, k_1 k_2}^{\mathcal{G} \mathcal{G}}(\cdot)$, where $a_1, a_2 \in \{1, \dots, p_1\}$, $k_1, k_2 \in \{1, \dots, p\}$ have the form

$$C_{a_1 a_2, k_1 k_2}^{\mathcal{G} \mathcal{G}}(\cdot) = \frac{c}{2} \{\mathcal{G}_{k_1, a_1}(\cdot) \mathcal{G}_{k_2, a_2}(\cdot) + \mathcal{G}_{k_1, a_2}(\cdot) \mathcal{G}_{k_2, a_1}(\cdot)\},$$

with c being a constant. In particular, we know that

- $T_{R,A}^{(N)}, T_{MR,A}^{(N)} \in \mathcal{T}_{N,3,A}^0$,
- $T_{LR,A}^{(N)}, T_{Rob,A}^{(N)}, T_{Wob,A}^{(N)}, T_{MRob,A}^{(N)}, T_{MWob,A}^{(N)}, T_{grad,A}^{(N)} \in \mathcal{T}_{N,3,A}^{1/2}$, and
- $T_{W,A}^{(N)}, T_{MW,A}^{(N)} \in \mathcal{T}_{N,3,A}^1$.

However, we employ the enlarged class $\mathcal{T}_{N,3,A}$, for greater generality.

2.4. Bartlett-type adjustments: GCF and GB

We now formulate the Bartlett-type adjustments (the GCF and GB) applied to the class $\mathcal{T}_{N,3,A}$, regardless of the ordinary likelihood inference (with $\mathbf{A}(\cdot) \equiv \mathbf{0}_p$) or Stern's adjusted profile likelihood inference (with $\mathbf{A}(\cdot) = \mathbf{M}(\cdot)$). Both GCF and GB ensure that the resulting test has size $\alpha + o(N^{-1})$, where $0 < \alpha < 1$ is the significance level. In what follows, unless otherwise stated, the term with superscript C (or CD) indicates that it depends on the C (or C, D)-functions, e.g.,

$$\Gamma_{a_1 a_2 a_3}^C(\cdot) = -\frac{1}{6} \{v_{a_1, a_2, a_3}^{\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) + 6 C_{a_1 a_2 a_3}^{+\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot)\}, \quad a_1, a_2, a_3 \in \{1, \dots, p_1\}$$

according to the test statistic $T_A^{(N)} \in \mathcal{T}_{N,3,A}$ under consideration.

Definition 1 (GCF Adjustment [17]). Let $T_A^{(N)} \in \mathcal{T}_{N,3,A}$. We say that

$$T_A^{\text{GCF}(N)} = T_A^{(N)} + \frac{2}{N} \sum_{R=2,4,6} \tilde{I}_{b_1 \dots b_R} \prod_{i=1}^R [\tilde{\mathbf{v}}_{(11,2)}^{-1} \tilde{\mathbf{Z}}_{(1)}^{(N)}]_{b_i} \quad (8)$$

is a generalized Cordeiro–Ferrari Bartlett-type (GCF, in short) adjustment of the test statistic $T_A^{(N)}$, if

$$P_{\theta^\dagger}^{(N)} [T_A^{\text{GCF}(N)} \leq x] = \Pr[\chi_{p_1}^2 \leq x] + o(N^{-1}),$$

where functions $\Gamma_{a_1 \dots a_R}(\cdot)$ are of class $\mathcal{C}^1(\Theta)$.

Definition 2 (GB (double) Adjustment [16]). Let $T_A^{(N)} \in \mathcal{T}_{N,3,A}$. With

$$T_A^{*(N)} = T_A^{(N)} + \frac{2}{N^{1/2}} \tilde{r}_{b_1 b_2 b_3}^C \prod_{i=1}^3 [\tilde{v}_{(11.2)}^{-1} \tilde{\mathbf{z}}_{(1)}^{(N)}]_{b_i},$$

we say that

$$T_A^{GB(N)} = T_A^{*(N)} + \frac{2}{N} \sum_{R=2,4} \tilde{r}_{b_1 \dots b_R}^{GB} \prod_{i=1}^R [\tilde{v}_{(11.2)}^{-1} \tilde{\mathbf{z}}_{(1)}^{(N)}]_{b_i} \quad (9)$$

is a generalized Bartlett-type (GB, in short) adjustment of the test statistic $T_A^{(N)}$, if

$$P_{\theta^\dagger}^{(N)} [T_A^{GB(N)} \leq x] = \Pr[\chi_{p_1}^2 \leq x] + o(N^{-1}),$$

where functions $\Gamma_{a_1 a_2}^{GB}(\cdot)$ and $\Gamma_{a_1 a_2 a_3 a_4}^{GB}(\cdot)$ are of class $\mathcal{C}^1(\Theta)$.

It is worth noting that, for the case $\mathbf{A}(\cdot) \equiv \mathbf{0}_p$, Kakizawa [16] gave a general definition of

$$\begin{aligned} T_A^{GB\Delta(N)} = & T_A^{(N)} + \frac{2}{N^{1/2}} \tilde{r}_{b_1 b_2 b_3}^C \prod_{i=1}^3 [\tilde{v}_{(11.2)}^{-1} \tilde{\mathbf{z}}_{(1)}^{(N)}]_{b_i} \\ & + \frac{2}{N} \left\{ (\tilde{r}_{b_1 b_2 b_3 b_4}^{GB} + \tilde{\Delta}_{b_1 b_2 b_3 b_4}^{GB}) \prod_{i=1}^4 [\tilde{v}_{(11.2)}^{-1} \tilde{\mathbf{z}}_{(1)}^{(N)}]_{b_i} \right. \\ & \left. + \tilde{\Delta}_{b_1 b_2 b_3, k_1 k_2}^{GB} \tilde{\mathbf{z}}_{k_1 k_2}^{(N)} \prod_{i=1}^3 [\tilde{v}_{(11.2)}^{-1} \tilde{\mathbf{z}}_{(1)}^{(N)}]_{b_i} + \tilde{r}_{b_1 b_2}^{GB} \prod_{i=1}^2 [\tilde{v}_{(11.2)}^{-1} \tilde{\mathbf{z}}_{(1)}^{(N)}]_{b_i} \right\} \end{aligned} \quad (10)$$

and interpreted it as “double adjustment” (see (A.1)). Technically, such an interpretation for the GB adjustment in Definition 2 may be more convenient since, for the derivations of Propositions 1 and 2 (see also (23) and (24)), the GB adjustment of $T_A^{(N)} (\in \mathcal{T}_{N,3,A})$ can be treated as a special case of the GCF adjustment of $T_A^{*(N)} (\in \mathcal{T}_{N,3,A})$. That is,

$$T_A^{GB(N)} = (T_A^*)^{GCF(N)} \quad (\text{set } \Gamma_{a_1 \dots a_6}(\cdot) \equiv 0 \text{ in Definition 1}),$$

where $T_A^{*(N)} \in \mathcal{T}_{N,3,A}$ has the stochastic expansion (6), with $C_{a_1 a_2 a_3}^{\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot)$ replaced by $C_{a_1 a_2 a_3}^{\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) + \Gamma_{a_1 a_2 a_3}^C(\cdot)$. The argument that the C^+ -function (7) should read $-(1/6) v_{a_1, a_2, a_3}^{\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot)$ for $T_A^{*(N)} \in \mathcal{T}_{N,3,A}$ will be made repeatedly below. One reason for the simplest choice of $\Delta_{b_1 b_2 b_3 b_4}^{GB}(\cdot) = \Delta_{b_1 b_2 b_3, k_1 k_2}^{GB}(\cdot) \equiv 0$ (see (9)) is that whichever choice is made with respect to $\mathbf{A}(\cdot)$, the terms $\tilde{\Delta}_{b_1 b_2 b_3 b_4}^{GB}$ and $\tilde{\Delta}_{b_1 b_2 b_3, k_1 k_2}^{GB}$ involved in (10) have no contribution to the resulting third-order average local power (see Proposition 2). On the other hand, as will be argued in Remarks 5(i), 6(ii), and 7, some local power properties of the GB-adjusted test are different from those of the GCF-adjusted test. For the sake of a better understanding of such distinctive results for the GCF/GB adjustments, we define the two types separately.

The name “GCF” and the idea behind the form of (8) stem from a generalization of the original proposal in Cordeiro and Ferrari [6]. As in the ordinary case (see [17]), if $p_1 > 1$, infinitely many functions

$$[\Gamma_{a_1 a_2}(\cdot), \Gamma_{a_1 a_2 a_3 a_4}(\cdot), \Gamma_{a_1 a_2 a_3 a_4 a_5 a_6}(\cdot)]_{a_1, a_2, a_3, a_4, a_5, a_6 \in \{1, \dots, p_1\}}$$

that satisfy (17)–(19) give rise to an improved test statistic (8). All resulting GCF-adjusted tests are, nonetheless, equivalent in the sense of the third-order average local power (see Proposition 2). Among infinitely many possibilities for Definition 1, it suffices to consider the adjustment

$$T_A^{CF_3(N)} = \left[1 - \frac{2}{N} \left\{ \frac{\tilde{\beta}_3^C}{p_1(p_1+2)(p_1+4)} (T_A^{(N)})^2 + \frac{\tilde{\beta}_{2,A}^{CD}}{p_1(p_1+2)} T_A^{(N)} + \frac{\tilde{\beta}_{1,A}^{CD}}{p_1} \right\} \right] T_A^{(N)} \quad \text{for } T_A^{(N)} \in \mathcal{T}_{N,3,A}, \quad (11)$$

where

$$\begin{aligned} \beta_{1,A}^{CD}(\cdot) = & \beta_1^{CD}(\cdot) + v_{(11.2)}^{b_1, b_2}(\cdot) \left[{}_A D_{b_1 b_2}^{\mathcal{G} \mathcal{G}}(\cdot) + \frac{\langle 2 \rangle}{2} {}_A D_{b_1, k_1 k_2}^{\mathcal{G}}(\cdot) v_{k_1 k_2, b_2}^{\mathcal{G}}(\cdot) + \frac{1}{2} A_{b_1}^{\mathcal{G}}(\cdot) A_{b_2}^{\mathcal{G}}(\cdot) - M_{b_1}^{\mathcal{G}}(\cdot) A_{b_2}^{\mathcal{G}}(\cdot) \right. \\ & \left. - \frac{1}{2} \{ v_{b_1, b_2, b}^{\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) + 6 C_{b_1 b_2 b}^{+\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) \} v_{(11.2)}^{b, b'}(\cdot) A_{b'}^{\mathcal{G}}(\cdot) \right], \end{aligned} \quad (12)$$

$$\beta_{2,A}^{CD}(\cdot) = \beta_2^{CD}(\cdot) + \frac{1}{2} v_{(11.2)}^{b_1, b_2}(\cdot) \{ v_{b_1, b_2, b}^{\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) + 6 C_{b_1 b_2 b}^{+\mathcal{G} \mathcal{G} \mathcal{G}}(\cdot) \} v_{(11.2)}^{b, b'}(\cdot) A_{b'}^{\mathcal{G}}(\cdot), \quad (13)$$

$$\beta_3^C(\cdot) = \frac{1}{24} \left[3v_{(11,2)}^{b_1, b_2}(\cdot) \{v_{b_1, b_2, b}^{\mathcal{G}}(\cdot) + 6C_{b_1 b_2 b}^{+\mathcal{G}}(\cdot)\} v_{(11,2)}^{b, b'}(\cdot) \{v_{b'_1, b'_2, b'}^{\mathcal{G}}(\cdot) + 6C_{b'_1 b'_2 b'}^{+\mathcal{G}}(\cdot)\} v_{(11,2)}^{b'_1, b'_2}(\cdot) \right. \\ \left. + 2\{v_{b_1, b_3, b_5}^{\mathcal{G}}(\cdot) + 6C_{b_1 b_3 b_5}^{+\mathcal{G}}(\cdot)\} v_{(11,2)}^{b_1, b_2}(\cdot) v_{(11,2)}^{b_3, b_4}(\cdot) v_{(11,2)}^{b_5, b_6}(\cdot) \{v_{b_2, b_4, b_6}^{\mathcal{G}}(\cdot) + 6C_{b_2 b_4 b_6}^{+\mathcal{G}}(\cdot)\} \right]. \quad (14)$$

The closed-form expressions for $\beta_1^{CD}(\cdot)$ and $\beta_2^{CD}(\cdot)$ are found in Kakizawa [17] for the ordinary likelihood inference.

Similarly, if $p_1 > 1$, infinitely many functions

$$[\Gamma_{a_1 a_2}^{GB}(\cdot), \Gamma_{a_1 a_2 a_3 a_4}^{GB}(\cdot)]_{a_1, a_2, a_3, a_4 \in \{1, \dots, p_1\}}$$

that satisfy (20) and (21) give rise to an improved test statistic (9); all resulting GB-adjusted tests are equivalent in the sense of the third-order average local power, so that, among infinitely many possibilities for Definition 2, it suffices to consider the adjustment

$$(T_A^*)^{CF_2(N)} = \left[1 - \frac{2}{N} \left\{ \frac{\tilde{\beta}_2^{CD}}{p_1(p_1 + 2)} T_A^{*(N)} + \frac{\tilde{\beta}_{1,A}^{+CD}}{p_1} \right\} \right] T_A^{*(N)} \quad \text{for } T_A^{(N)} \in \mathcal{T}_{N,3,A}, \quad (15)$$

where

$$\beta_{1,A}^{+CD}(\cdot) = \beta_1^{CD}(\cdot) + v_{(11,2)}^{b_1, b_2}(\cdot) \left[A D_{b_1, k_1 k_2}^{\mathcal{G}}(\cdot) + \frac{\langle 2 \rangle}{2} A D_{b_1, k_1 k_2}^{\mathcal{G}}(\cdot) v_{k_1 k_2, b_2}^{\mathcal{G}}(\cdot) + \frac{1}{2} A_{b_1}^{\mathcal{G}}(\cdot) A_{b_2}^{\mathcal{G}}(\cdot) - M_{b_1}^{\mathcal{G}}(\cdot) A_{b_2}^{\mathcal{G}}(\cdot) \right].$$

Remark 4. (i) In view of (11)–(14), the test statistic $T_A^{(N)} \in \mathcal{T}_{N,3,A}$ is Bartlett correctable (i.e., the scalar multiplication to $T_A^{(N)}$, given by

$$T_A^{\text{Bart}(N)} = \left(1 - \frac{2\tilde{\beta}_{1,A}^{CD}}{N p_1} \right) T_A^{(N)}, \quad (16)$$

makes the large sample χ^2 -approximation of the null distribution accurate up to $o(N^{-1})$), iff

$$v_{a_1, a_2, a_3}^{\mathcal{G}}(\cdot) + 6C_{a_1 a_2 a_3}^{+\mathcal{G}}(\cdot) \equiv 0, \quad a_1, a_2, a_3 \in \{1, \dots, p_1\} \text{ and } \beta_2^{CD}(\cdot) \equiv 0.$$

It turns out that the necessary and sufficient condition for the Bartlett correctability members of the enlarged class $\bigcup_A \mathcal{T}_{N,3,A}$ is the same as that for the class $\mathcal{T}_{N,3}$ (see Theorem 1 of [16]). This supports the Bartlett correctability of Stern's adjusted profile LR test statistic $T_{\text{LR}}^{\text{AP}(N)}$ (see also [9] for a conditional approach).

(ii) The CF adjustment (11) is not nonnegative, because it is polynomial of degree 3 in $T_A^{(N)}$, with the negative coefficient $-(2/N)\tilde{\beta}_3^C/\{p_1(p_1 + 2)(p_1 + 4)\}$ of $(T_A^{(N)})^3$ (hence, it is not monotone increasing). To preserve the monotonicity, the idea of Kakizawa [14] can be applied to suggest its monotone variant $T_A^{K_3(N)} = T_A^{CF_3(N)} + N^{-2} \Delta_A^{CD(N)}$ for $T_A^{(N)} \in \mathcal{T}_{N,3,A}$, where

$$\Delta_A^{CD(N)} = \int_0^{T_A^{(N)}} \left\{ \frac{\tilde{\beta}_3^C}{p_1(p_1 + 2)(p_1 + 4)} (3t^2) + \frac{\tilde{\beta}_{2,A}^{CD}}{p_1(p_1 + 2)} (2t) + \frac{\tilde{\beta}_{1,A}^{CD}}{p_1} \right\}^2 dt \\ = 9 \frac{(\tilde{\beta}_3^C)^2}{p_1^2(p_1 + 2)^2(p_1 + 4)^2} \frac{(T_A^{(N)})^5}{5} + 3 \frac{\tilde{\beta}_{2,A}^{CD} \tilde{\beta}_3^C}{p_1^2(p_1 + 2)^2(p_1 + 4)} \frac{(T_A^{(N)})^4}{2} \\ + \left\{ 4 \frac{(\tilde{\beta}_{2,A}^{CD})^2}{p_1^2(p_1 + 2)^2} + 6 \frac{\tilde{\beta}_{1,A}^{CD} \tilde{\beta}_3^C}{p_1^2(p_1 + 2)(p_1 + 4)} \right\} \frac{(T_A^{(N)})^3}{3} + 2 \frac{\tilde{\beta}_{1,A}^{CD} \tilde{\beta}_{2,A}^{CD}}{p_1^2(p_1 + 2)} (T_A^{(N)})^2 + \frac{(\tilde{\beta}_{1,A}^{CD})^2}{p_1^2} T_A^{(N)}.$$

Similarly, instead of (15), we can use $(T_A^*)^{K_2(N)} = (T_A^*)^{CF_2(N)} + N^{-2} \Delta_A^{+CD(N)}$ for $T_A^{(N)} \in \mathcal{T}_{N,3,A}$, where

$$\Delta_A^{+CD(N)} = \int_0^{T_A^{*(N)}} \left\{ \frac{\tilde{\beta}_2^{CD}}{p_1(p_1 + 2)} (2t) + \frac{\tilde{\beta}_{1,A}^{+CD}}{p_1} \right\}^2 dt \\ = 4 \frac{(\tilde{\beta}_2^{CD})^2}{p_1^2(p_1 + 2)^2} \frac{(T_A^{*(N)})^3}{3} + 2 \frac{\tilde{\beta}_{1,A}^{+CD} \tilde{\beta}_2^{CD}}{p_1^2(p_1 + 2)} (T_A^{*(N)})^2 + \frac{(\tilde{\beta}_{1,A}^{+CD})^2}{p_1^2} T_A^{*(N)}.$$

3. Asymptotic expansions of $T_A^{\text{GCF}(N)}$ and $T_A^{\text{GB}(N)}$

We are primarily concerned with the third-order local powers of several Bartlett-type adjusted tests from Stern's adjusted profile likelihood. Thus, in this section, we present the third-order asymptotic expansions of the distributions of the Bartlett-type adjusted test statistics $T_A^{\text{GCF}(N)}$ and $T_A^{\text{GB}(N)}$ under both the null hypothesis $\theta_{(1)} = \theta_{(1)0}$ and a sequence of local alternatives $\theta_{(1)} = \theta_{(1)0} + N^{-1/2} \mathbf{h}_{(1)}$, where $\mathbf{h}_{(1)} = (h_1, \dots, h_{p_1})^\top$ (the outlined proof of Proposition 1 is provided in the Appendix).

Unless otherwise stated, for any (nonrandom/random) scalar or vector or matrix function $Q(\cdot)$, we write Q instead of $Q(\theta^\dagger)$, to simplify the notation. Let $g_\nu(\cdot; \omega^2)$ be the density of the noncentral chi-squared distribution with ν degrees of freedom and noncentrality parameter ω^2 . Write $G_\nu(x; \omega^2) = \int_0^x g_\nu(t; \omega^2) dt$. Given a level $\alpha \in (0, 1)$, let $\chi_{p_1, \alpha}^2$ be the upper α -point of the chi-squared distribution with p_1 degrees of freedom.

Proposition 1. Let $T_A^{(N)} \in \mathcal{T}_{N,3,A}$.

(i) $P_{\theta^\dagger}^{(N)}[T_A^{\text{GCF}(N)} \leq x] = \Pr[\chi_{p_1}^2 \leq x] + o(N^{-1})$ iff

$$0 \equiv \beta_{1,A}^{\text{CD}}(\cdot) + \Gamma_{b_1 b_2}(\cdot) v_{(11,2)}^{b_1, b_2}(\cdot), \quad (17)$$

$$0 \equiv \beta_{2,A}^{\text{CD}}(\cdot) + 3\Gamma_{b_1 b_2 b_3 b_4}(\cdot) v_{(11,2)}^{b_1, b_2}(\cdot) v_{(11,2)}^{b_3, b_4}(\cdot), \quad (18)$$

$$0 \equiv \beta_3^{\text{C}}(\cdot) + 15\Gamma_{b_1 b_2 b_3 b_4 b_5 b_6}(\cdot) v_{(11,2)}^{b_1, b_2}(\cdot) v_{(11,2)}^{b_3, b_4}(\cdot) v_{(11,2)}^{b_5, b_6}(\cdot). \quad (19)$$

In addition, $P_{\theta^\dagger}^{(N)}[T_A^{\text{GB}(N)} \leq x] = \Pr[\chi_{p_1}^2 \leq x] + o(N^{-1})$ iff

$$0 \equiv \beta_{1,A}^{\dagger \text{CD}}(\cdot) + \Gamma_{b_1 b_2}^{\text{GB}}(\cdot) v_{(11,2)}^{b_1, b_2}(\cdot), \quad (20)$$

$$0 \equiv \beta_2^{\text{CD}}(\cdot) + 3\Gamma_{b_1 b_2 b_3 b_4}^{\text{GB}}(\cdot) v_{(11,2)}^{b_1, b_2}(\cdot) v_{(11,2)}^{b_3, b_4}(\cdot) \quad (21)$$

(the GB is a special case of the GCF in the sense that $T_A^{\text{GB}(N)} = (T_A^*)^{\text{GCF}(N)}$ with $\Gamma_{a_1 a_2 a_3 a_4 a_5 a_6}(\cdot) \equiv 0$; note that (19) holds for the GB, since substituting $-(1/6)v_{a_1 a_2 a_3}^{\text{G}}(\cdot)$ for $C_{a_1 a_2 a_3}^{+\text{G}}(\cdot)$ yields $\beta_3^{\text{C}}(\cdot) \equiv 0$).

(ii) The local power $P_{\theta^\dagger + N^{-1/2}(\mathbf{h}_{(1)}^\top, \mathbf{0}_{p_2}^\top)^\top}^{(N)}[T_A^{\#(N)} > \chi_{p_1, \alpha}^2] = \pi_\alpha^{(N)}(T_A^{\#(N)}; \mathbf{h}_{(1)})$ (say), $\# = \text{GCF, GB}$ of the resulting GCF-(GB-) adjusted test is expanded as

$$\pi_\alpha^{(N)}(T_A^{\#(N)}; \mathbf{h}_{(1)}) = 1 - G_{p_1}(\chi_{p_1, \alpha}^2; \mathbf{h}_{(1)}^\top \mathbf{v}_{(11,2)} \mathbf{h}_{(1)}) + \sum_{\ell=1}^2 \frac{2}{N^{\ell/2}} \pi_{\alpha, A}^{\#(\ell)}(\mathbf{h}_{(1)}) + o(N^{-1}),$$

where

$$\pi_{\alpha, A}^{\text{GCF}(\ell)}(\mathbf{h}_{(1)}) = \sum_{v=1}^{3\ell} \mathcal{P}_{v, A}^{\text{GCF}(\ell)}(\mathbf{h}_{(1)}) g_{p_1+2v}(\chi_{p_1, \alpha}^2; \mathbf{h}_{(1)}^\top \mathbf{v}_{(11,2)} \mathbf{h}_{(1)}),$$

$$\pi_{\alpha, A}^{\text{GB}(\ell)}(\mathbf{h}_{(1)}) = \sum_{v=1}^{2\ell} \mathcal{P}_{v, A}^{\text{GB}(\ell)}(\mathbf{h}_{(1)}) g_{p_1+2v}(\chi_{p_1, \alpha}^2; \mathbf{h}_{(1)}^\top \mathbf{v}_{(11,2)} \mathbf{h}_{(1)})$$

depend on the C,D-functions associated with $T_A^{(N)} \in \mathcal{T}_{N,3,A}$. In particular,

$$\mathcal{P}_{1,A}^{\text{GCF}(1)}(\mathbf{h}_{(1)}) = -M_{\diamond}^{\text{G}} + A_{\diamond}^{\text{G}} - \frac{1}{6}(3v_{\diamond, \diamond, \diamond}^{\text{G}} + 2v_{\diamond, \diamond, \diamond}^{\text{G}}) + \frac{1}{2}(2v_{\diamond, \diamond, \diamond}^{\text{G}} + v_{\diamond, \diamond, \diamond}^{\text{G}}),$$

$$\mathcal{P}_{2,A}^{\text{GCF}(1)}(\mathbf{h}_{(1)}) = \frac{1}{2}(v_{\diamond, \diamond, b, b'}^{\text{G}} + 6C_{\diamond b b'}^{+\text{G}}) v_{(11,2)}^{b, b'} + \frac{1}{6} v_{\diamond, \diamond, \diamond}^{\text{G}}, \quad \mathcal{P}_{3,A}^{\text{GCF}(1)}(\mathbf{h}_{(1)}) = \frac{1}{6}(v_{\diamond, \diamond, \diamond}^{\text{G}} + 6C_{\diamond \diamond \diamond}^{+\text{G}})$$

(we used the notation $Q_{\dots\dots} = Q_{\dots\dots} h_a$). The closed-form expressions of $\mathcal{P}_{v,A}^{\text{GCF}(2)}(\mathbf{h}_{(1)})$ are available from the author, and the $\mathcal{P}_{v,A}^{\text{GB}(\ell)}(\mathbf{h}_{(1)})$'s are given by the $\mathcal{P}_{v,A}^{\text{GCF}(\ell)}(\mathbf{h}_{(1)})$'s, with $C_{a_1 a_2 a_3}^{+\text{G}}(\cdot)$ replaced by $-(1/6)v_{a_1 a_2 a_3}^{\text{G}}(\cdot)$, i.e., $\mathcal{P}_{2,A}^{\text{GB}(1)}(\mathbf{h}_{(1)}) = (1/6)v_{\diamond, \diamond, \diamond}^{\text{G}}$, $\mathcal{P}_{3,A}^{\text{GB}(1)}(\mathbf{h}_{(1)}) = 0$, and so on.

Remark 5. (i) All GB-adjusted tests $T_A^{\text{GB}(N)} > \chi_{p_1, \alpha}^2$ for $T_A^{(N)} \in \mathcal{T}_{N,3,A}$ are equivalent up to the second-order in terms of the point-by-point local power (see [18]).

(ii) The second-order locally unbiasedness $(\partial/\partial \mathbf{h}_{(1)})\pi_{\alpha, A}^{\text{GB}(1)}(\mathbf{0}_{p_1}) = \mathbf{0}_{p_1}$ holds if $\mathbf{A}(\cdot) = \mathbf{M}(\cdot)$.

(iii) The ordinary LR test $T_{\text{LR}}^{(N)} > \chi_{p_1, \alpha}^2$ is, in general, second-order locally biased (see [12]). On the other hand, Stern's adjusted profile LR test $T_{\text{LR}}^{\text{AP}(N)} > \chi_{p_1, \alpha}^2$ is second-order locally unbiased. The score bias is probably the cause of this phenomenon.

Although it also reveals that the third-order point-by-point local power of the Cordeiro–Ferrari (CF, in short) adjusted test [6], given by $T_A^{\text{CF}_3(N)} > \chi_{p_1, \alpha}^2$ for $T_A^{(N)} \in \mathcal{T}_{N,3,A}$ (see (11)), is the same as that of the size-adjusted test based on Cornish–Fisher's type expansion for the percentile

$$q_{p_1, \alpha}^{\text{CD}(N)} = \left[1 + \frac{2}{N} \left\{ \frac{\tilde{\beta}_3^{\text{C}}}{p_1(p_1+2)(p_1+4)} (\chi_{p_1, \alpha}^2)^2 + \frac{\tilde{\beta}_{2,A}^{\text{CD}}}{p_1(p_1+2)} \chi_{p_1, \alpha}^2 + \frac{\tilde{\beta}_{1,A}^{\text{CD}}}{p_1} \right\} \right] \chi_{p_1, \alpha}^2,$$

i.e.,

$$P_{\theta^\dagger + N^{-1/2}(\mathbf{h}'_{(1)}, 0'_{p_2})'}[T_A^{(N)} > q_{p_1, \alpha}^{CD(N)}] = P_{\theta^\dagger + N^{-1/2}(\mathbf{h}'_{(1)}, 0'_{p_2})'}[T_A^{CF_3(N)} > \chi_{p_1, \alpha}^2] + o(N^{-1}), \quad (22)$$

no definitive conclusion is obtained there, because the formulas $\mathcal{P}_{v,A}^{GCF(2)}(\mathbf{h}_{(1)})$ depending on the choice of $[\Gamma_{a_1 a_2}(\cdot), \Gamma_{a_1 a_2 a_3 a_4}(\cdot), \Gamma_{a_1 a_2 a_3 a_4 a_5 a_6}(\cdot)]_{a_1, a_2, a_3, a_4, a_5, a_6 \in \{1, \dots, p_1\}}$ (see (17)–(19)) are very complicated, e.g., $\mathcal{P}_{6,A}^{GCF(2)}(\mathbf{h}_{(1)}) = \Gamma_{\diamond, \diamond, \diamond, \diamond, \diamond, \diamond} + (1/2)\{\mathcal{P}_{3,A}^{GCF(1)}(\mathbf{h}_{(1)})\}^2$.

We turn from the point-by-point local power comparison to the average local power comparison [27]. That is, let

$$\text{ave}_{S_\lambda}\{\pi(\mathbf{h}_{(1)})\} = \frac{\int_{S_\lambda} \pi(\mathbf{h}_{(1)}) d\mathbf{h}_{(1)}}{\int_{S_\lambda} d\mathbf{h}_{(1)}}$$

be the average of $\pi(\mathbf{h}_{(1)})$ along the sphere $S_\lambda = \{\mathbf{h}_{(1)} \in \mathbb{R}^{p_1} : \mathbf{h}_{(1)}^\top \mathbf{v}_{(11.2)} \mathbf{h}_{(1)} = \lambda\}$, where $\lambda > 0$. Noting that $\text{ave}_{S_\lambda}\{\pi_{\alpha,A}^{(1)}(\mathbf{h}_{(1)})\} = \text{ave}_{S_\lambda}\{\mathcal{P}_{6,A}^{GCF(2)}(\mathbf{h}_{(1)})\} = 0$, we have the third-order term in the average local power of the GCF-(or GB-) adjusted test, i.e.,

$$\begin{aligned} \text{ave}_{S_\lambda}\{\pi_{\alpha,A}^{GCF(2)}(\mathbf{h}_{(1)})\} &= \sum_{v=1}^5 \text{ave}_{S_\lambda}\{\mathcal{P}_{v,A}^{GCF(2)}(\mathbf{h}_{(1)})\} g_{p_1+2v}(\chi_{p_1, \alpha}^2; \lambda), \\ \text{ave}_{S_\lambda}\{\pi_{\alpha,A}^{GB(2)}(\mathbf{h}_{(1)})\} &= \sum_{v=1}^4 \text{ave}_{S_\lambda}\{\mathcal{P}_{v,A}^{GB(2)}(\mathbf{h}_{(1)})\} g_{p_1+2v}(\chi_{p_1, \alpha}^2; \lambda) \end{aligned}$$

(we used $\text{ave}_{S_\lambda}(\prod_{i=1}^{2m} h_{a_i}) = (\lambda^m / E[\chi_{p_1}^{2m}]) \frac{(2m)!}{2^m m!} v_{(11.2)}^{a_1, a_2} \cdots v_{(11.2)}^{a_{2m-1}, a_{2m}}$), as follows.

Proposition 2. We write $(A_{a_1}^g)_k = (\partial/\partial\theta_k)A_{a_1}^g(\boldsymbol{\theta}^\dagger)$ and define

$$\begin{aligned} AU_1 &= [-A_{b_1}^g A_{b_2}^g + 2M_{b_1}^g A_{b_2}^g + v_{(11.2)}^{b, b'} v_{b' b_1}^{g g g} A_{b_2}^g + 2\{-(A_{b_1}^g)_{b_2}^g + (A_{b_1}^g)_{b_2}\} v_{(11.2)}^{b_1, b_2}, \\ AU^C &= -6v_{(11.2)}^{b, b'} C_{b b' b_1}^{+g g g} v_{(11.2)}^{b_1, b_2} A_{b_2}^g. \end{aligned}$$

Then,

$$\begin{aligned} \text{ave}_{S_\lambda}\{\mathcal{P}_{1,A}^{GCF(2)}(\mathbf{h}_{(1)})\} &= \text{ave}_{S_\lambda}\{\mathcal{P}_1^{GCF(2)}(\mathbf{h}_{(1)})\} + \frac{\lambda}{2p_1} AU_1, \\ \text{ave}_{S_\lambda}\{\mathcal{P}_{2,A}^{GCF(2)}(\mathbf{h}_{(1)})\} &= \text{ave}_{S_\lambda}\{\mathcal{P}_2^{GCF(2)}(\mathbf{h}_{(1)})\} + \frac{\lambda}{2p_1} AU^C \\ &\quad + \frac{\lambda^2}{2p_1(p_1+2)} \{-(v_{b_1 b_2, b_3}^{g g g} + v_{b_1, b_2, b_3}^{g g g}) + 2v_{b_1 b_2, b_3}^{g g} + v_{b_1, b_2, b_3}^{g g}\} A_{b_4}^g \langle 3 \rangle v_{(11.2)}^{b_1, b_2} v_{(11.2)}^{b_3, b_4}, \\ \text{ave}_{S_\lambda}\{\mathcal{P}_{3,A}^{GCF(2)}(\mathbf{h}_{(1)})\} &= \text{ave}_{S_\lambda}\{\mathcal{P}_3^{GCF(2)}(\mathbf{h}_{(1)})\} + \frac{\lambda^2}{2p_1(p_1+2)} AU^C, \\ \text{ave}_{S_\lambda}\{\mathcal{P}_{v,A}^{GCF(2)}(\mathbf{h}_{(1)})\} &= \text{ave}_{S_\lambda}\{\mathcal{P}_v^{GCF(2)}(\mathbf{h}_{(1)})\}, \quad v = 4, 5 \end{aligned}$$

(as a special case, the $\text{ave}_{S_\lambda}\{\mathcal{P}_{v,A}^{GB(2)}(\mathbf{h}_{(1)})\}$'s are given by the $\text{ave}_{S_\lambda}\{\mathcal{P}_{v,A}^{GCF(2)}(\mathbf{h}_{(1)})\}$'s, with $C_{a_1 a_2 a_3}^{+g g g}(\cdot)$ replaced by $-(1/6)v_{a_1, a_2, a_3}^{g g g}(\cdot)$).

Here, the $\text{ave}_{S_\lambda}\{\mathcal{P}_v^{GCF(2)}(\mathbf{h}_{(1)})\}$'s for the ordinary case (see [20]) depend on the C-functions associated with $T_A^{(N)} \in \mathcal{T}_{N, 3, A}$ through

$C_{a_1 a_2 a_3}^{+g g g}(\cdot)$, $C_{a_1 a_2, k_1 k_2}^{g g}(\cdot) \mathcal{M}_{k_1 k_2, a_3 a_4}^{g g}(\cdot)$, and $C_{a_1 a_2, k_1 k_2}^{g g}(\cdot) C_{a_3 a_4, k_3 k_4}^{g g}(\cdot) \mathcal{M}_{k_1 k_2, k_3 k_4}(\cdot)$, where

$$\mathcal{M}_{j_1 j_2, j_3 j_4}(\cdot) = v_{j_1 j_2, j_3 j_4}(\cdot) - v_{j_1 j_2, k}(\cdot) v^{k, k'}(\cdot) v_{j_3 j_4, k'}(\cdot), \quad j_1, j_2, j_3, j_4 \in \{1, \dots, p\}.$$

Remark 6. (i) The third-order average local power of the GCF-(or GB-) adjusted test $T_A^{GCF(N)} > \chi_{p_1, \alpha}^2$ (or $T_A^{GB(N)} > \chi_{p_1, \alpha}^2$) is independent of the D-functions associated with $T_A^{(N)} \in \mathcal{T}_{N, 3, A}$.

(ii) Whenever $\mathcal{M}_{j_1 j_2, j_3 j_4}(\cdot) \equiv 0, j_1, j_2, j_3, j_4 \in \{1, \dots, p\}$, all GB-adjusted tests $T_A^{GB(N)} > \chi_{p_1, \alpha}^2$ for $T_A^{(N)} \in \mathcal{T}_{N, 3, A}$ are equivalent up to the third-order in terms of the average local power.

The following formula and its GB counterpart, which will form the basis for the results in Section 4, are obtained from Proposition 2, using the relation $xg_v(x; \omega^2) = v g_{v+2}(x; \omega^2) + \omega^2 g_{v+4}(x; \omega^2)$ (we derive (24) as a special case of (23), with $C^{+\frac{g}{a_1 a_2 a_3}}(\cdot)$ replaced by $-(1/6)v_{a_1, a_2, a_3}^{\frac{g}{g}}(\cdot)$):

$$\text{ave}_{S_\lambda}\{\pi_{\alpha, A}^{\text{GCF}(2)}(\mathbf{h}_{(1)})\} = \frac{\lambda}{2p_1} \{U_{1, \alpha, A}^C g_{p_1+2}(\chi_{p_1, \alpha}^2; 0) + U_{2, \alpha, A}^C g_{p_1+4}(\chi_{p_1, \alpha}^2; 0)\} + O(\lambda^2), \quad (23)$$

$$\text{ave}_{S_\lambda}\{\pi_{\alpha, A}^{\text{GB}(2)}(\mathbf{h}_{(1)})\} = \frac{\lambda}{2p_1} \{U_{1, \alpha, A}^C g_{p_1+2}(\chi_{p_1, \alpha}^2; 0) + (U_{210} + {}_A U) g_{p_1+4}(\chi_{p_1, \alpha}^2; 0)\} + O(\lambda^2) \quad (24)$$

(the coefficients $U_{1, \alpha}^C$ and $U_{2, \alpha}^C$ of the expressions (15) and (16) in [20] had minor errors, which should, respectively, read (23) and (24), with $\mathbf{A}(\cdot) \equiv \mathbf{O}_p$), where

$$U_{1, \alpha, A}^C = (U_{11}^C + U_{110} + {}_A U_1) + U_{12}^C \chi_{p_1, \alpha}^2, \\ U_{2, \alpha, A}^C = (U_{21}^C - U_{21} + U_{210} + {}_A U^C) + (U_{22}^C - U_{22}) \chi_{p_1, \alpha}^2$$

(U_{110} and U_{210} are common for any $T_A^{(N)} \in \mathcal{T}_{N, 3, A}$), with

$$U_{11}^C = v_{(11,2)}^{b_1, b_2} v_{(11,2)}^{b_3, b_4} C_{b_1 b_2, k_1 k_2}^{\frac{g}{g}} \mathcal{M}_{k_1 k_2, b_3 b_4}^{\frac{g}{g}}, \\ U_{12}^C = -\frac{1}{p_1 + 2} v_{(11,2)}^{b_1, b_2} v_{(11,2)}^{b_3, b_4} \langle 3 \rangle C_{b_1 b_2, k_1 k_2}^{\frac{g}{g}} C_{b_3 b_4, k_3 k_4}^{\frac{g}{g}} \mathcal{M}_{k_1 k_2, k_3 k_4}^{\frac{g}{g}}, \\ U_{21}^C = [3v_{(22)}^{r, r'} (v_{r', b_4}^{\frac{g}{g}} + v_{r, r'}^{\frac{g}{g}}) C_{b_1 b_2 b_3}^{+\frac{g}{g}} + 3C_{b_1 b_2 b}^{+\frac{g}{g}} v_{(11,2)}^{b, b'} v_{b_3 b_4, b'}^{\frac{g}{g}} \\ + 6C_{b_1 b_3 b}^{+\frac{g}{g}} v_{(11,2)}^{b, b'} \{2v_{b_2 b_4, b'}^{\frac{g}{g}} + v_{b_2, b_4, b'}^{\frac{g}{g}} - (2v_{b_2 b_4, b'}^{\frac{g}{g}} + v_{b_2, b_4, b'}^{\frac{g}{g}})\} \\ + 6\{(C_{b_1 b_2 b_3}^{+\frac{g}{g}})_{b_4} - (C_{b_1 b_2 b_3}^{+\frac{g}{g}})_{b_4}^{\frac{g}{g}}\} v_{(11,2)}^{b_1, b_2} v_{(11,2)}^{b_3, b_4}, \\ U_{22}^C = -\frac{1}{p_1 + 4} C_{b_1 b_2 b_3}^{+\frac{g}{g}} C_{b_4 b_5 b_6}^{+\frac{g}{g}} \langle 15 \rangle v_{(11,2)}^{b_1, b_2} v_{(11,2)}^{b_3, b_4} v_{(11,2)}^{b_5, b_6}$$

(we write $(C^{+\frac{g}{a_1 a_2 a_3}})_k = (\partial/\partial \theta_k) C^{+\frac{g}{a_1 a_2 a_3}}(\boldsymbol{\theta}^\dagger)$). Here, U_{2v} and ${}_A U$ are, respectively, defined as U_{2v}^C and ${}_A U^C$, with $C^{+\frac{g}{a_1 a_2 a_3}}(\cdot)$ replaced by $-(1/6)v_{a_1, a_2, a_3}^{\frac{g}{g}}(\cdot)$.

4. Main results

The following three points will be addressed for the class of the test statistics $\mathcal{T}_{N, 3, A}$, on the basis of $\text{ave}_{S_\lambda}\{\pi_{\alpha, A}^{\#(2)}(\mathbf{h}_{(1)})\}$ (see (23) and (24)), together with the identity (22):

- (i) Which Bartlett-type adjustment of the GCF and the GB is better?
- (ii) In the present setup, does the Rao test continue to be locally optimal?
- (iii) Can we choose $\mathbf{A}(\cdot)$ reasonably?

To deal with (i) and (ii), we prepare the following lemma (without proof).

Lemma 3. (i) We have

$$C_{b_1 b_2 b_3}^{+\frac{g}{g}}(\boldsymbol{\theta}) C_{b_4 b_5 b_6}^{+\frac{g}{g}}(\boldsymbol{\theta}) \langle 15 \rangle v_{(11,2)}^{b_1, b_2}(\boldsymbol{\theta}) v_{(11,2)}^{b_3, b_4}(\boldsymbol{\theta}) v_{(11,2)}^{b_5, b_6}(\boldsymbol{\theta}) \\ = \{9C_{b_1 b_2 b_5}^{+\frac{g}{g}}(\boldsymbol{\theta}) C_{b_3 b_4 b_6}^{+\frac{g}{g}}(\boldsymbol{\theta}) + 6C_{b_1 b_3 b_5}^{+\frac{g}{g}}(\boldsymbol{\theta}) C_{b_2 b_4 b_6}^{+\frac{g}{g}}(\boldsymbol{\theta})\} v_{(11,2)}^{b_1, b_2}(\boldsymbol{\theta}) v_{(11,2)}^{b_3, b_4}(\boldsymbol{\theta}) v_{(11,2)}^{b_5, b_6}(\boldsymbol{\theta}) \geq 0,$$

whose equality holds iff $C^{+\frac{g}{a_1 a_2 a_3}}(\cdot) \equiv 0$, $a_1, a_2, a_3 \in \{1, \dots, p_1\}$.

(ii) We have

$$v_{(11,2)}^{b_1, b_2}(\boldsymbol{\theta}) v_{(11,2)}^{b_3, b_4}(\boldsymbol{\theta}) \langle 3 \rangle C_{b_1 b_2, k_1 k_2}^{\frac{g}{g}}(\boldsymbol{\theta}) C_{b_3 b_4, k_3 k_4}^{\frac{g}{g}}(\boldsymbol{\theta}) \mathcal{M}_{k_1 k_2, k_3 k_4}^{\frac{g}{g}}(\boldsymbol{\theta}) \\ = \{C_{b_1 b_2, k_1 k_2}^{\frac{g}{g}}(\boldsymbol{\theta}) C_{b_3 b_4, k_3 k_4}^{\frac{g}{g}}(\boldsymbol{\theta}) + 2C_{b_1 b_3, k_1 k_2}^{\frac{g}{g}}(\boldsymbol{\theta}) C_{b_2 b_4, k_3 k_4}^{\frac{g}{g}}(\boldsymbol{\theta})\} v_{(11,2)}^{b_1, b_2}(\boldsymbol{\theta}) v_{(11,2)}^{b_3, b_4}(\boldsymbol{\theta}) E_\theta[\ell^{\perp k_1 k_2}(\mathbf{X}, \boldsymbol{\theta}) \ell^{\perp k_3 k_4}(\mathbf{X}, \boldsymbol{\theta})] \\ = E_\theta\left(\left[\text{tr}\{v_{(11,2)}^{-1}(\boldsymbol{\theta}) \mathbf{C}^{\frac{g}{g}, \perp}(\boldsymbol{\theta})\}^2\right] + 2E_\theta\left(\text{tr}\{v_{(11,2)}^{-1}(\boldsymbol{\theta}) \mathbf{C}^{\frac{g}{g}, \perp}(\boldsymbol{\theta})\}^2\right)\right) \geq 0,$$

where $\mathbf{C}^{\frac{g}{g}, \perp}(\boldsymbol{\theta}) = [C_{b_1 b_2, k_1 k_2}^{\frac{g}{g}}(\boldsymbol{\theta}) \ell_{k_1 k_2}^{\perp}(\mathbf{X}, \boldsymbol{\theta})]_{b_1, b_2 \in \{1, \dots, p_1\}}$, with

$$\ell_{j_1 j_2}^{\perp}(\mathbf{X}, \boldsymbol{\theta}) = \ell_{j_1 j_2}(\mathbf{X}, \boldsymbol{\theta}) - v_{j_1 j_2, k}(\boldsymbol{\theta}) v^{k, k'}(\boldsymbol{\theta}) \ell_{k'}(\mathbf{X}, \boldsymbol{\theta}), \quad j_1, j_2 \in \{1, \dots, p\}.$$

4.1. Comparison of two Bartlett-type adjustments: GCF and GB

We compare two Bartlett-type adjustments GCF and GB. Suppose there is at least one $(a'_1, a'_2, a'_3), a'_1, a'_2, a'_3 \in \{1, \dots, p_1\}$, such that $v_{a'_1, a'_2, a'_3}^{\mathcal{G}\mathcal{G}\mathcal{G}}(\cdot) + 6C_{a'_1, a'_2, a'_3}^{\mathcal{G}\mathcal{G}\mathcal{G}}(\cdot) \neq 0$ holds (otherwise, $T_A^{\text{GCF}(N)} = T_A^{\text{GB}(N)}$). Then, (23) and (24) immediately yield, for $T_A^{(N)} \in \mathcal{T}_{N,3,A}$,

$$\begin{aligned} & \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \text{ave}_{S_\lambda} \left[\lim_{N \rightarrow \infty} N \{ \pi_\alpha^{(N)}(T_A^{\text{GCF}(N)}; \mathbf{h}_{(1)}) - \pi_\alpha^{(N)}(T_A^{\text{GB}(N)}; \mathbf{h}_{(1)}) \} \right] \\ &= \frac{1}{p_1} \{ (U_{21}^C - U_{21} + AU^C - AU) + (U_{22}^C - U_{22}) \chi_{p_1, \alpha}^2 \} g_{p_1+4}(\chi_{p_1, \alpha}^2; 0), \end{aligned}$$

so that this limit is positive (negative) for sufficiently small α , if $U_{22}^C > U_{22}$ ($U_{22}^C < U_{22}$), depending on the C^+ -functions $C_{a_1, a_2, a_3}^{\mathcal{G}\mathcal{G}\mathcal{G}}(\cdot)$ (see (7)) associated with $T_A^{(N)} \in \mathcal{T}_{N,3,A}$; hereafter, “locally” means that both $\lambda > 0$ and $0 < \alpha < 1$ are sufficiently small, depending on the model to be tested and the dimension (p_1, p_2) . Although there is, in general, no uniform locally superiority and inferiority between the $T_A^{\text{GCF}(N)}$ -test and the $T_A^{\text{GB}(N)}$ -test, the definitive conclusion is obtained by letting $C_{a_1, a_2, a_3}^{\mathcal{G}\mathcal{G}\mathcal{G}}(\cdot) \equiv 0$ for $a_1, a_2, a_3 \in \{1, \dots, p_1\}$ (i.e., $U_{22}^C = 0$), as follows.

Theorem 4. If $v_{a_1, a_2, a_3}^{\mathcal{G}\mathcal{G}\mathcal{G}}(\cdot) \equiv 0, a_1, a_2, a_3 \in \{1, \dots, p_1\}$, then,

$$\text{ave}_{S_\lambda} \left[\lim_{N \rightarrow \infty} N \{ \pi_\alpha^{(N)}(T_{R,A}^{\text{GCF}(N)}; \mathbf{h}_{(1)}) - \pi_\alpha^{(N)}(T_{R,A}^{\text{GB}(N)}; \mathbf{h}_{(1)}) \} \right] = 0 \quad \text{for any } \alpha \in (0, 1) \text{ and } \lambda > 0.$$

On the other hand, if $v_{a'_1, a'_2, a'_3}^{\mathcal{G}\mathcal{G}\mathcal{G}}(\cdot) \neq 0$ for at least one $(a'_1, a'_2, a'_3), a'_1, a'_2, a'_3 \in \{1, \dots, p_1\}$, then, $-U_{22} > 0$ and

$$\begin{aligned} & \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \text{ave}_{S_\lambda} \left[\lim_{N \rightarrow \infty} N \{ \pi_\alpha^{(N)}(T_{R,A}^{\text{GCF}(N)}; \mathbf{h}_{(1)}) - \pi_\alpha^{(N)}(T_{R,A}^{\text{GB}(N)}; \mathbf{h}_{(1)}) \} \right] \\ &= -\frac{1}{p_1} \{ (U_{21} + AU) + U_{22} \chi_{p_1, \alpha}^2 \} g_{p_1+4}(\chi_{p_1, \alpha}^2; 0) > 0 \quad \text{for sufficiently small } \alpha. \end{aligned}$$

Hence, the $T_{R,A}^{\text{GCF}(N)}$ -test is locally superior to the $T_{R,A}^{\text{GB}(N)}$ -test.

4.2. Local optimality of the Rao test $T_{R,A}^{\text{GCF}(N)} > \chi_{p_1, \alpha}^2$

A further implication of (23) and (24) is that we have, for $T_A^{(N)} \in \mathcal{T}_{N,3,A}$,

$$\begin{aligned} & \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \text{ave}_{S_\lambda} \left[\lim_{N \rightarrow \infty} N \{ \pi_\alpha^{(N)}(T_{R,A}^{\text{GCF}(N)}; \mathbf{h}_{(1)}) - \pi_\alpha^{(N)}(T_A^{\text{GCF}(N)}; \mathbf{h}_{(1)}) \} \right] \\ &= -\frac{1}{p_1} [(U_{11}^C + U_{12}^C \chi_{p_1, \alpha}^2) g_{p_1+2}(\chi_{p_1, \alpha}^2; 0) + \{(U_{21}^C + AU^C) + U_{22}^C \chi_{p_1, \alpha}^2\} g_{p_1+4}(\chi_{p_1, \alpha}^2; 0)], \end{aligned} \quad (25)$$

$$\begin{aligned} & \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \text{ave}_{S_\lambda} \left[\lim_{N \rightarrow \infty} N \{ \pi_\alpha^{(N)}(T_{R,A}^{\text{GB}(N)}; \mathbf{h}_{(1)}) - \pi_\alpha^{(N)}(T_A^{\text{GB}(N)}; \mathbf{h}_{(1)}) \} \right] \\ &= -\frac{1}{p_1} (U_{11}^C + U_{12}^C \chi_{p_1, \alpha}^2) g_{p_1+2}(\chi_{p_1, \alpha}^2; 0). \end{aligned} \quad (26)$$

Consequently, the following local optimality of the Rao test $T_{R,A}^{\text{GCF}(N)} > \chi_{p_1, \alpha}^2$ can be established for the class $\mathcal{T}_{N,3,A}$, in which the C^+ -functions $C_{a_1, a_2, a_3}^{\mathcal{G}\mathcal{G}\mathcal{G}}(\cdot)$ (see (7)) and the functions $C_{a_1, a_2, j_1, j_2}^{\mathcal{G}\mathcal{G}}(\boldsymbol{\theta})$, associated with $T_A^{(N)} \in \mathcal{T}_{N,3,A}$, play a crucial role (thus, the previous result [20] for the subclass $\mathcal{T}_{N,3}^C$ ($\subset \mathcal{T}_{N,3}$) can be extended in this way, as a special case of $\mathbf{A}(\cdot) \equiv \mathbf{0}_p$).

Theorem 5. The GCF-adjusted (A-variant) Rao test $T_{R,A}^{\text{GCF}(N)} > \chi_{p_1, \alpha}^2$ (or the size-adjusted (A-variant) Rao test $T_{R,A}^{(N)} > q_{p_1, \alpha}^{R(N)}$) is locally optimal among the $T_A^{\text{GCF}(N)}$ -tests (or the size-adjusted $T_A^{(N)}$ -tests) and the $T_A^{\text{GB}(N)}$ -tests, for any $T_A^{(N)} \in \mathcal{T}_{N,3,A}$.

Proof. Recall that the coefficients of $\chi_{p_1, \alpha}^2 g_{p_1+2v}(\chi_{p_1, \alpha}^2; 0)$, $v = 1, 2$, in (25) are independent of $\mathbf{A}(\cdot)$ and nonnegative. That is, $-U_{12}^C \geq 0$ and $-U_{22}^C \geq 0$ (see Lemma 3). Thus, (25) is positive for sufficiently small α , if $-U_{12}^C > 0$ or $-U_{22}^C > 0$. It turns out that the $T_{R,A}^{\text{GCF}(N)}$ -test, with $-U_{12}^C = -U_{22}^C = 0$ (e.g., $C_{a_1, a_2, j_1, j_2}^{\mathcal{G}\mathcal{G}}(\cdot) = C_{a_1, a_2, j_1, j_2}^{\mathcal{G}\mathcal{G}\mathcal{G}}(\cdot) \equiv 0, a_1, a_2, a_3 \in \{1, \dots, p_1\}, j_1, j_2 \in \{1, \dots, p\}$), is locally optimal among the GCF-adjusted (A-variant) tests $T_A^{\text{GCF}(N)} > \chi_{p_1, \alpha}^2$ for $T_A^{(N)} \in \mathcal{T}_{N,3,A}$. From (22), this local optimality holds even in the class of the size-adjusted $T_A^{(N)}$ -tests $T_A^{(N)} > q_{p_1, \alpha}^{CD(N)}$ for $T_A^{(N)} \in \mathcal{T}_{N,3,A}$. Similarly,

(26) is positive for sufficiently small α , if $-U_{12}^C > 0$, so that the $T_{R,A}^{GB(N)}$ -test, with $-U_{12}^C = 0$ (e.g., $C_{a_1 a_2 j_1 j_2}^{\mathcal{G}}(\cdot) \equiv 0$, $a_1, a_2 \in \{1, \dots, p_1\}$, $j_1, j_2 \in \{1, \dots, p\}$), is locally optimal among the GB-adjusted (A -variant) tests $T_A^{GB(N)} > \chi_{p_1, \alpha}^2$ for $T_A^{(N)} \in \mathcal{T}_{N,3,A}$. Noting that the $T_{R,A}^{GCF(N)}$ -test is locally superior to the $T_{R,A}^{GB(N)}$ -test (Theorem 4), this completes the proof. \square

The essential point of Theorem 5 on the local optimality is that the original (unadjusted) Rao statistic (i.e., the expected A -variant Rao statistic $T_{R,A}^{(N)}$), belonging to the class $\mathcal{T}_{N,3,A}$, has the C^+ -functions $C_{a_1 a_2 a_3}^{\mathcal{G}}(\cdot) \equiv 0$, with $C_{a_1 a_2, k_1 k_2}^{\mathcal{G}}(\cdot) \equiv 0$ (these conditions are independent of the choice of $\mathbf{A}(\cdot)$). Note that the observed A -variant Rao/Wald statistics $T_{R^{ob}, A}^{(N)}, T_{W^{ob}, A}^{(N)}, T_{MR^{ob}, A}^{(N)}, T_{WR^{ob}, A}^{(N)} \in \mathcal{T}_{N,3,A}^{1/2}$ do not satisfy $C_{a_1 a_2, k_1 k_2}^{\mathcal{G}}(\cdot) \equiv 0$, in general.

Remark 7. (i) As described above, the GB-adjusted (A -variant) Rao test $T_{R,A}^{GB(N)} > \chi_{p_1, \alpha}^2$ is locally optimal among the $T_A^{GB(N)}$ -tests for any $T_A^{(N)} \in \mathcal{T}_{N,3,A}$ (the previous result [19] for the subclass $\mathcal{T}_{N,3}^c (\subset \mathcal{T}_{N,3})$ can be extended in this way, as a special case of $\mathbf{A}(\cdot) \equiv \mathbf{0}_p$). However, using the GB adjustment rather than the GCF adjustment, it is impossible to discriminate any test statistic belonging to the subclass $\mathcal{T}_{N,3,A}^c$, i.e., $T_{R,A}^{(N)}, T_{MR,A}^{(N)} \in \mathcal{T}_{N,3,A}^0$, or $T_{LR,A}^{(N)}, T_{R^{ob}, A}^{(N)}, T_{W^{ob}, A}^{(N)}, T_{MR^{ob}, A}^{(N)}, T_{MW^{ob}, A}^{(N)}, T_{grad,A}^{(N)} \in \mathcal{T}_{N,3,A}^{1/2}$, or $T_{W,A}^{(N)}, T_{MW,A}^{(N)} \in \mathcal{T}_{N,3,A}^1$ (see Remark 3). In fact, Proposition 2 reveals that

$$\text{ave} \left[\lim_{S_\lambda \rightarrow \infty} N \{ \pi_\alpha^{(N)}(T_{A,1}^{GB(N)}; \mathbf{h}_{(1)}) - \pi_\alpha^{(N)}(T_{A,2}^{GB(N)}; \mathbf{h}_{(1)}) \} \right] = 0 \quad \text{for any } \alpha \in (0, 1) \text{ and } \lambda > 0,$$

if the functions $C_{a_1 a_2 j_1 j_2}^{\mathcal{G}}(\cdot)$, $a_1, a_2 \in \{1, \dots, p_1\}$, $j_1, j_2 \in \{1, \dots, p\}$, associated with $T_{A,1}^{(N)} \in \mathcal{T}_{N,3,A}$, are the same as those of $T_{A,2}^{(N)} \in \mathcal{T}_{N,3,A}$.

(ii) Using the GCF adjustment, even if $\mathcal{M}_{j_1 j_2 j_3 j_4}(\cdot) \equiv 0$, $j_1, j_2, j_3, j_4 \in \{1, \dots, p\}$ (see also Remark 6(ii)), it is possible to discriminate any test statistic belonging to the subclass $\mathcal{T}_{N,3,A}^c$ (see Remark 3), in general. For example, (25) indicates that

$$\begin{aligned} & \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \text{ave} \left[\lim_{S_\lambda \rightarrow \infty} N \{ \pi_\alpha^{(N)}(T_{R,A}^{GCF(N)}; \mathbf{h}_{(1)}) - \pi_\alpha^{(N)}(T_{MR,A}^{GCF(N)}; \mathbf{h}_{(1)}) \} \right] \\ &= -\frac{1}{p_1} \{ (\text{MR}U_{21}^C + \text{MR}_A U^C) + \text{MR}U_{22}^C \chi_{p_1, \alpha}^2 \} g_{p_1+4}(\chi_{p_1, \alpha}^2; 0) > 0 \quad \text{for sufficiently small } \alpha, \end{aligned}$$

provided that $-\text{MR}U_{22}^C > 0$, i.e., there is at least one (a'_1, a'_2, a'_3) , $a'_1, a'_2, a'_3 \in \{1, \dots, p_1\}$, such that $\text{MR}C_{a_1 a_2 a_3}^{\mathcal{G}}(\cdot) = -\{(\langle 3 \rangle/3) v_{a'_1 a'_2 a'_3}^{\mathcal{G}}(\cdot) + (1/2) v_{a'_1 a'_2 a'_3}^{\mathcal{G}}(\cdot)\} \neq 0$ holds (note that

$$\text{ave} \left[\lim_{S_\lambda \rightarrow \infty} N \{ \pi_\alpha^{(N)}(T_{R,A}^{GCF(N)}; \mathbf{h}_{(1)}) - \pi_\alpha^{(N)}(T_{MR,A}^{GCF(N)}; \mathbf{h}_{(1)}) \} \right] = 0 \quad \text{for any } \alpha \in (0, 1) \text{ and } \lambda > 0,$$

if $(\langle 3 \rangle/3) v_{a_1 a_2 a_3}^{\mathcal{G}}(\cdot) + (1/2) v_{a_1 a_2 a_3}^{\mathcal{G}}(\cdot) \equiv 0$, $a_1, a_2, a_3 \in \{1, \dots, p_1\}$; see Proposition 2).

4.3. Ordinary versus A -variant inference

One may wish to compare two tests of $T_R^{GCF(N)} > \chi_{p_1, \alpha}^2$ and $T_{R,A}^{GCF(N)} > \chi_{p_1, \alpha}^2$. Actually, Theorem 5 demonstrates that the former is locally optimal among the $T_R^{GCF(N)}$ or $T_R^{GB(N)}$ -tests for any $T^{(N)} \in \mathcal{T}_{N,3}$, whereas the latter is locally optimal among the $T_A^{GCF(N)}$ or $T_A^{GB(N)}$ -tests for any $T_A^{(N)} \in \mathcal{T}_{N,3,A}$.

Now, we suppose that $A_{a'}^{\mathcal{G}}(\cdot) \neq 0$ for some $a' \in \{1, \dots, p_1\}$ (otherwise, we trivially have $T_{R,A}^{(N)} = T_R^{(N)}$ ($T_{LR,A}^{(N)} = T_{LR}^{(N)}$, $T_{W,A}^{(N)} = T_W^{(N)}$, $T_{MR,A}^{(N)} = T_{MR}^{(N)}$, $T_{MW,A}^{(N)} = T_{MW}^{(N)}$, $T_{R^{ob}, A}^{(N)} = T_{R^{ob}}^{(N)}$, $T_{W^{ob}, A}^{(N)} = T_{W^{ob}}^{(N)}$, $T_{MR^{ob}, A}^{(N)} = T_{MR^{ob}}^{(N)}$, $T_{MW^{ob}, A}^{(N)} = T_{MW^{ob}}^{(N)}$, and $T_{grad,A}^{(N)} = T_{grad}^{(N)}$), by definition). Recall that using the relation $xg_v(x; 0) = v g_{v+2}(x; 0)$, (23) yields

$$\begin{aligned} & \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \text{ave} \left[\lim_{S_\lambda \rightarrow \infty} N \{ \pi_\alpha^{(N)}(T_A^{GCF(N)}; \mathbf{h}_{(1)}) - \pi_\alpha^{(N)}(T_R^{GCF(N)}; \mathbf{h}_{(1)}) \} \right] \\ &= \frac{1}{p_1} \left(A U_1 + \frac{\chi_{p_1, \alpha}^2}{p_1 + 2} A U^C \right) g_{p_1+2}(\chi_{p_1, \alpha}^2; 0) \quad \text{for } T^{(N)} \in \mathcal{T}_{N,3} \text{ and } T_A^{(N)} \in \mathcal{T}_{N,3,A}. \end{aligned}$$

Therefore, by letting $C_{a_1 a_2 a_3}^{\mathcal{G}}(\cdot) \equiv 0$ for $a_1, a_2, a_3 \in \{1, \dots, p_1\}$ (i.e., $A U^C = 0$), there is, in general, no uniform superiority or inferiority between the $T_R^{GCF(N)}$ -test and the $T_{R,A}^{GCF(N)}$ -test. That is,

$$\lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \text{ave} \left[\lim_{S_\lambda \rightarrow \infty} N \{ \pi_\alpha^{(N)}(T_{R,A}^{GCF(N)}; \mathbf{h}_{(1)}) - \pi_\alpha^{(N)}(T_R^{GCF(N)}; \mathbf{h}_{(1)}) \} \right] = \frac{1}{p_1} A U_1 g_{p_1+2}(\chi_{p_1, \alpha}^2; 0)$$

is positive (negative) if $A U_1 > 0$ (< 0).

Remark 8. The analogous discussion under the global parameter orthogonality $\mathbf{v}_{(12)}(\cdot) \equiv \mathbf{O}_{p_1, p_2}$ (then, $\mathbf{v}_{(11,2)}^{a_1, a_2}(\cdot)$ is nothing but the (a_1, a_2) th element of $\mathbf{v}_{(11)}^{-1}(\cdot)$, denoted by $\mathbf{v}_{(11)}^{a_1, a_2}(\cdot)$, and $\mathbf{g}^\top(\cdot) \equiv [\mathbf{I}_{p_1} \mathbf{O}_{p_1, p_2}]$), corresponding to $T^{(N)} \in \mathcal{T}_{N,3}$ and $T_A^{(N)} \in \mathcal{T}_{N,3,A}$, is possible, on the basis of

$$\begin{aligned} & \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \text{ave}_{S_\lambda} \left[\lim_{N \rightarrow \infty} N \{ \pi_\alpha^{(N)}(T_A^{\text{GCF}(N)}; \mathbf{h}_{(1)}) - \pi_\alpha^{(N)}(T^{\text{GCF}(N)}; \mathbf{h}_{(1)}) \} \right] \\ &= \frac{1}{p_1} \left[\left\{ (2M_{b_1} - A_{b_1}) + \mathbf{v}_{(11)}^{b, b'} \left(\mathbf{v}_{bb', b_1} - \frac{6\chi_{p_1, \alpha}^2}{p_1 + 2} C_{bb' b_1}^+ \right) \right\} \mathbf{v}_{(11)}^{b_1, b_2} A_{b_2} \right] g_{p_1+2}(\chi_{p_1, \alpha}^2; 0). \end{aligned}$$

These results are substantial extensions of Mukerjee [25] under the global parameter orthogonality, who derived similar expressions for the LR, Rao, and Wald tests from the ordinary and conditional likelihood with the scalar parameter case ($p_1 = 1$).

5. Non-i.i.d. case

Although we have focused on the i.i.d. case, for notational simplicity, we arrive at the same conclusions even in a non-i.i.d. case where some regularity conditions are met for the log-likelihood derivatives according to the situations under consideration. For example, in the setting where $\mathbf{Y}_1, \dots, \mathbf{Y}_N$ are independent but non-identically distributed random vectors (taking values of \mathbb{R}^{d_Y}) according to a density $f^{[i]}(\mathbf{y}, \boldsymbol{\theta})$, $i = 1, \dots, N$, we have only to change the notation in Section 2, as follows. The log-likelihood and its derivative should read

$$\mathcal{L}^{(N)}(\boldsymbol{\theta}) = \sum_{i=1}^N \log f^{[i]}(\mathbf{Y}_i, \boldsymbol{\theta}) = \sum_{i=1}^N \ell^{[i]}(\mathbf{Y}_i, \boldsymbol{\theta}) \quad (\text{say})$$

and, for $I_R = j_1 \cdots j_R$,

$$\mathcal{L}_{I_R}^{(N)}(\boldsymbol{\theta}) = \frac{\partial}{\partial \theta_{j_1}} \cdots \frac{\partial}{\partial \theta_{j_R}} \mathcal{L}^{(N)}(\boldsymbol{\theta}) = \sum_{i=1}^N \ell_{I_R}^{[i]}(\mathbf{Y}_i, \boldsymbol{\theta}),$$

respectively. In that case, the cumulant $\mathbf{v}_{I_{R_1}, \dots, I_{R_v}}(\boldsymbol{\theta})$ (see (2)) should read the averaged cumulant

$$\frac{1}{N} \text{cum}_{\boldsymbol{\theta}}^{(N)}[\mathcal{L}_{I_{R_1}}^{(N)}(\boldsymbol{\theta}), \dots, \mathcal{L}_{I_{R_v}}^{(N)}(\boldsymbol{\theta})] = \frac{1}{N} \sum_{i=1}^N \text{cum}_{\boldsymbol{\theta}}^{[i]}[\ell_{I_{R_1}}^{[i]}(\mathbf{Y}_i, \boldsymbol{\theta}), \dots, \ell_{I_{R_v}}^{[i]}(\mathbf{Y}_i, \boldsymbol{\theta})] = \bar{\mathbf{v}}_{I_{R_1}, \dots, I_{R_v}}^{(N)}(\boldsymbol{\theta}) \quad (\text{say}),$$

provided that $\bar{\mathbf{v}}_{I_{R_1}, \dots, I_{R_v}}^{(N)}(\boldsymbol{\theta}) = O(1)$, where $\text{cum}_{\boldsymbol{\theta}}^{[i]}$ denotes the cumulant with respect to $\mathbf{Y}_i \sim f^{[i]}(\cdot, \boldsymbol{\theta})$.

Example 1. We assume that Y_1, \dots, Y_N are independent exponential random variables with

$$E[Y_i] = \exp(x_{i1}\theta_1 + \cdots + x_{ip}\theta_p) = \eta_i(\boldsymbol{\theta}) \quad (\text{say}), \quad (27)$$

where $\mathbf{X} = (x_{ij})_{i=1, \dots, N; j=1, \dots, p}$ is a fixed design matrix. The log-likelihood of Y_1, \dots, Y_N is given by

$$\mathcal{L}^{(N)}(\boldsymbol{\theta}) = \sum_{i=1}^N \left\{ -\log \eta_i(\boldsymbol{\theta}) - \frac{Y_i}{\eta_i(\boldsymbol{\theta})} \right\}.$$

For this exponential regression model, it is easy to see that

$$\bar{\mathbf{v}}_{j_1 j_2}^{(N)}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N x_{ij_1} x_{ij_2}, \quad \bar{\mathbf{v}}_{j_1 j_2}^{(N)}(\boldsymbol{\theta}) = -\frac{1}{N} \sum_{i=1}^N x_{ij_1} x_{ij_2}$$

(the expected Fisher information matrix (per observation) is $\mathbf{X}^\top \mathbf{X}/N$, for any $\boldsymbol{\theta} \in \mathbb{R}^p$, and hence, the two versions of the Rao/Wald statistics using the expected information are identical). In addition, we have

$$\begin{aligned} \bar{\mathbf{v}}_{j_1 j_2 j_3}^{(N)}(\boldsymbol{\theta}) &= \frac{2}{N} \sum_{i=1}^N x_{ij_1} x_{ij_2} x_{ij_3}, & \bar{\mathbf{v}}_{j_1 j_2 j_3}^{(N)}(\boldsymbol{\theta}) &= -\frac{1}{N} \sum_{i=1}^N x_{ij_1} x_{ij_2} x_{ij_3}, & \bar{\mathbf{v}}_{j_1 j_2 j_3}^{(N)}(\boldsymbol{\theta}) &= \frac{1}{N} \sum_{i=1}^N x_{ij_1} x_{ij_2} x_{ij_3}, \\ \bar{\mathbf{v}}_{j_1 j_2 j_3 j_4}^{(N)}(\boldsymbol{\theta}) &= \frac{6}{N} \sum_{i=1}^N x_{ij_1} x_{ij_2} x_{ij_3} x_{ij_4}, & \bar{\mathbf{v}}_{j_1 j_2 j_3 j_4}^{(N)}(\boldsymbol{\theta}) &= -\frac{2}{N} \sum_{i=1}^N x_{ij_1} x_{ij_2} x_{ij_3} x_{ij_4}, \\ \bar{\mathbf{v}}_{j_1 j_2 j_3 j_4}^{(N)}(\boldsymbol{\theta}) &= \bar{\mathbf{v}}_{j_1 j_2 j_3 j_4}^{(N)}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N x_{ij_1} x_{ij_2} x_{ij_3} x_{ij_4}, & \bar{\mathbf{v}}_{j_1 j_2 j_3 j_4}^{(N)}(\boldsymbol{\theta}) &= -\frac{1}{N} \sum_{i=1}^N x_{ij_1} x_{ij_2} x_{ij_3} x_{ij_4} \end{aligned}$$

(these quantities are free of the parameter).

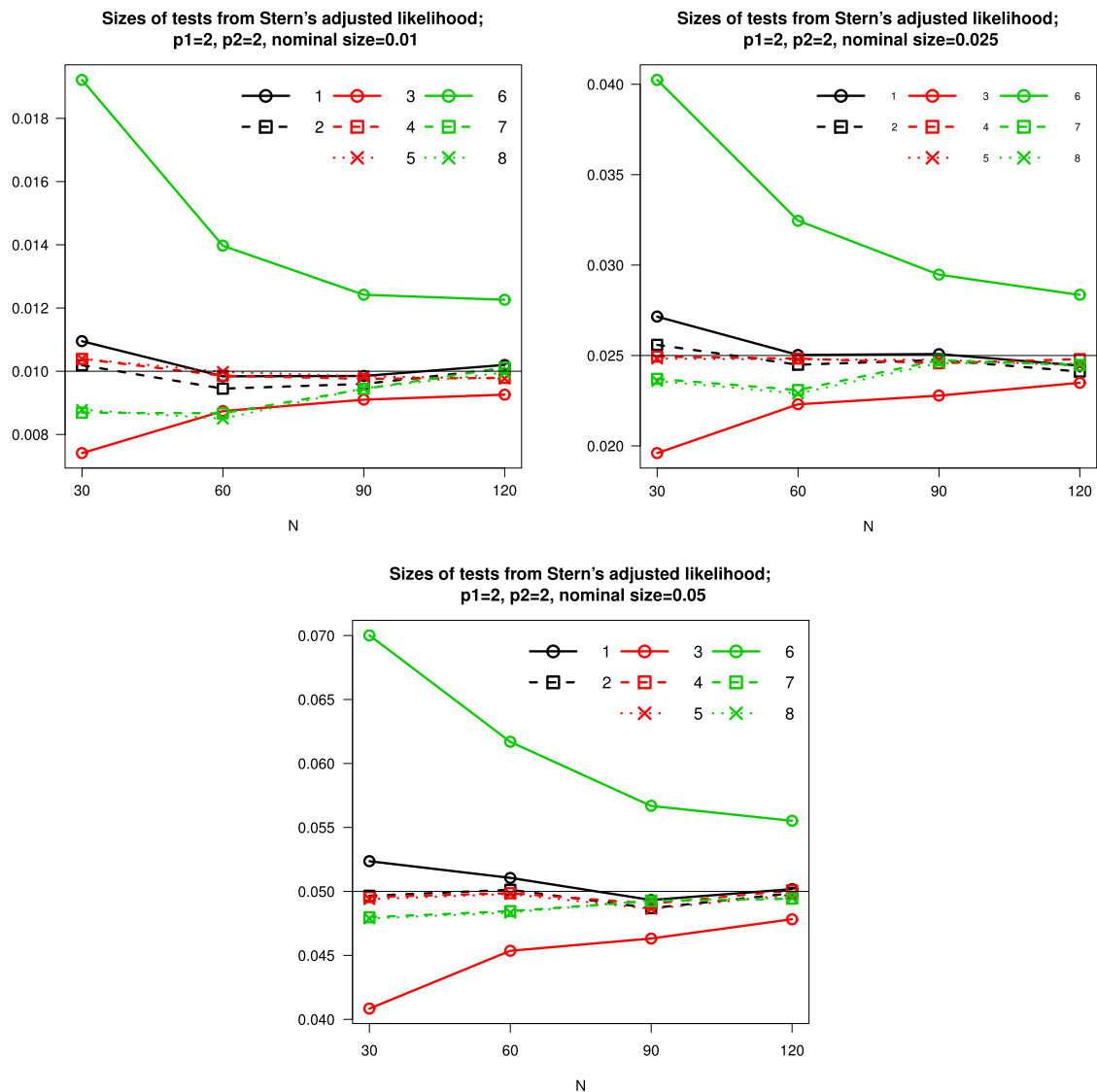


Fig. 1. Empirical sizes of “raw” tests or Bartlett(-type) adjusted tests from Stern's adjusted likelihood; 1: $T_{LR}^{AP(N)} > \chi_{p_1, \alpha}^2$, 2: $(T_{LR}^{AP})^{Bart(N)} > \chi_{p_1, \alpha}^2$, 3: $T_R^{AP(N)} > \chi_{p_1, \alpha}^2$, 4: $(T_R^{AP})^{CF_3(N)} > \chi_{p_1, \alpha}^2$, 5: $(T_R^{AP*})^{CF_2(N)} > \chi_{p_1, \alpha}^2$, 6: $T_W^{AP(N)} > \chi_{p_1, \alpha}^2$, 7: $(T_W^{AP})^{CF_3(N)} > \chi_{p_1, \alpha}^2$, and 8: $(T_W^{AP*})^{CF_2(N)} > \chi_{p_1, \alpha}^2$, where $N = 30, 60, 90, 120$, $(p_1, p_2) = (2, 2)$, and $\alpha = 0.01, 0.025, 0.05$.

Remark 9. For an exponential regression model with a fixed design matrix $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_{p_1}, \mathbf{1}_N)$, where \mathbf{X} is assumed to have full rank equal to $p_1 + 1 = p$ (say), if the null hypothesis to be tested is that the p_1 coefficients, except for the intercept, are all zero (i.e., $\theta_1 = \dots = \theta_{p_1} = 0$) then, Stern's adjusted profile likelihood (3) is the same as the (unadjusted) profile likelihood (1).

6. Simulation results

Given the sample size N and the full-rank design matrix $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_{p_1}, \mathbf{x}_{p_1+1}, \dots, \mathbf{x}_{p_1+p_2-1}, \mathbf{1}_N)$, we considered the hypothesis $H : \theta_{(1)} = \mathbf{0}_{p_1}$ against $A : \theta_{(1)} \neq \mathbf{0}_{p_1}$ about the exponential regression model (27), in the presence of the nuisance parameter $\theta_{(2)}$ (the true value for $\theta_{(2)}$ was set to $\mathbf{0}_{p_2}$), where the elements of \mathbf{x}_i were taken from the uniform distribution $\text{Unif}[-1/2, 1/2]$; these values were kept constant throughout the simulation (repetition 100,000).

First, the empirical sizes were examined for $N = 30, 60, 90, 120$ and $(p_1, p_2) = (2, 2)$. Figs. 1 and 2 show that the “raw” Rao test ($T_R^{AP(N)} > \chi_{p_1, \alpha}^2$ or $T_R^{(N)} > \chi_{p_1, \alpha}^2$) is conservative at each significance level $\alpha = 0.01, 0.025, 0.05$, whereas the “raw” LR test ($T_{LR}^{AP(N)} > \chi_{p_1, \alpha}^2$ or $T_{LR}^{(N)} > \chi_{p_1, \alpha}^2$) is slightly liberal, and the “raw” Wald test ($T_W^{AP(N)} > \chi_{p_1, \alpha}^2$ or $T_W^{(N)} > \chi_{p_1, \alpha}^2$) is seriously liberal (such a “raw” Wald test may be dangerous for a small sample size N , since, clearly, a fatal liberality implies

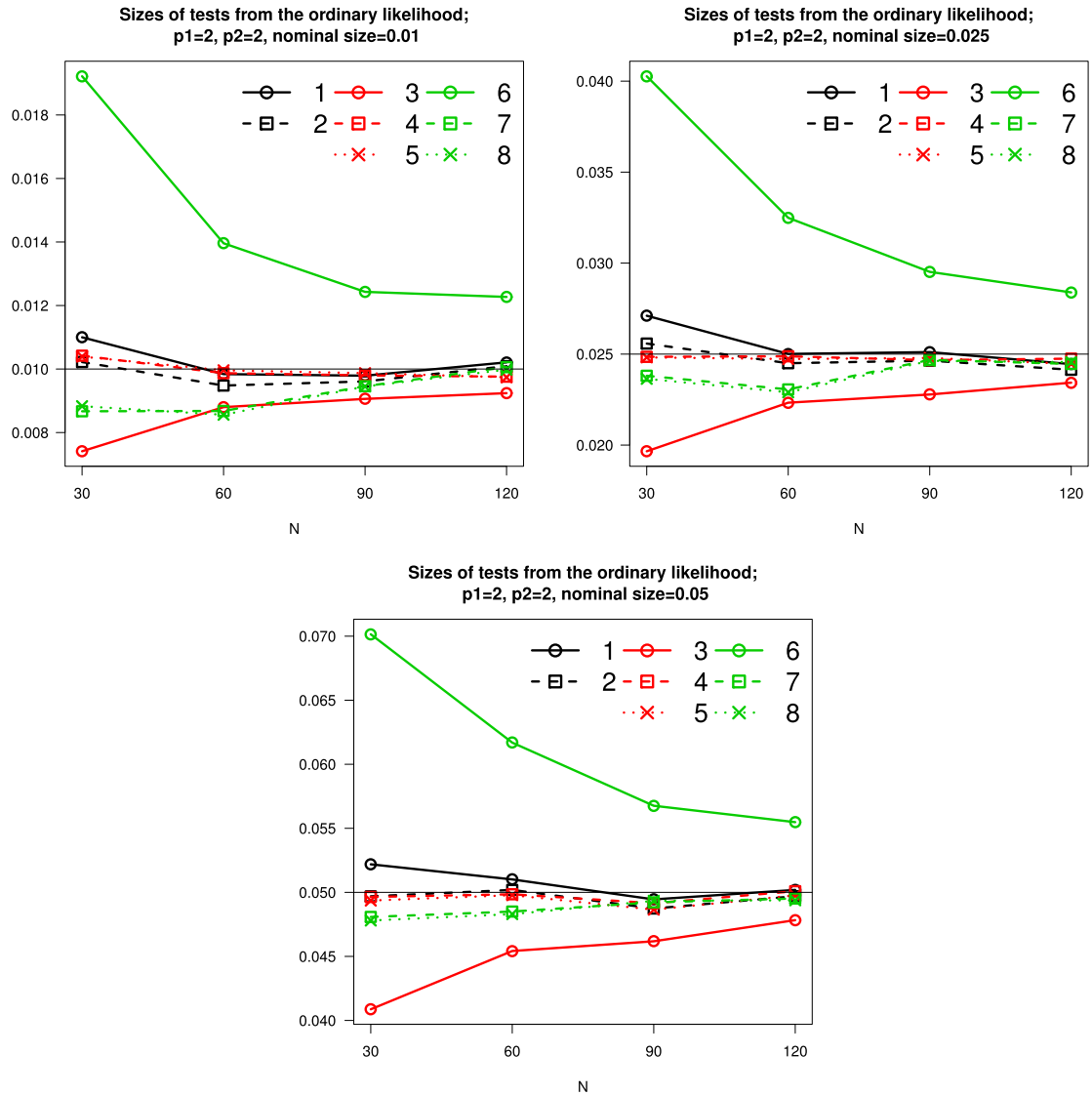


Fig. 2. Empirical sizes of “raw” tests or Bartlett(-type) adjusted tests from the ordinary likelihood; 1: $T_{LR}^{(N)} > \chi_{p_1, \alpha}^2$, 2: $T_{LR}^{Bart(N)} > \chi_{p_1, \alpha}^2$, 3: $T_R^{(N)} > \chi_{p_1, \alpha}^2$, 4: $T_R^{CF_3(N)} > \chi_{p_1, \alpha}^2$, 5: $(T_R^*)^{CF_2(N)} > \chi_{p_1, \alpha}^2$, 6: $T_W^{(N)} > \chi_{p_1, \alpha}^2$, 7: $T_W^{CF_3(N)} > \chi_{p_1, \alpha}^2$, and 8: $(T_W^*)^{CF_2(N)} > \chi_{p_1, \alpha}^2$, where $N = 30, 60, 90, 120$, $(p_1, p_2) = (2, 2)$, and $\alpha = 0.01, 0.025, 0.05$.

that the null hypothesis is likely to be rejected unfairly). Overall, as expected from asymptotic expansion, the size distortions by the “raw” LR, Rao, and Wald tests are improved by means of the Bartlett-type adjustments (we used (11) and (15)) for the Rao/Wald statistics, and the Bartlett adjustment (16) for the LR test statistic. Note that for the exponential regression model (27) we simulated, unfortunately, the simulation results for the Wald case may be unsatisfactory for a small sample size N , since, by Figs. 1 and 2 (see also the first row of each table at $\lambda = 0.0$), the Bartlett-type corrected Wald test still has the size distortion (in that case, the resulting test was conservative), for which Remark 4(ii) may be useful. However, the issue is not pursued further, since our main interest here is the numerical power analysis of the Bartlett-type corrected Rao test $(T_R^{AP})^{CF_3(N)} > \chi_{p_1, \alpha}^2$ or $T_R^{CF_3(N)} > \chi_{p_1, \alpha}^2$.

Next, the average power analysis was conducted for $N = 60$ and $(p_1, p_2) = (2, 2), (4, 4), (6, 6)$, as follows. In each repetition, we generated an exponential sample under $\theta = (\delta_{(1)}^\top, \mathbf{0}_{p_2}^\top)^\top$, where

$$\delta_{(1)} = \lambda^{1/2} (\mathbf{X}^\top \mathbf{X})_{(11:2)}^{-1/2} \left(\cos(U_1), \sin(U_1) \cos(U_2), \dots, \prod_{j=1}^{p_1-2} \sin(U_j) \cos(U_{p_1-1}), \prod_{j=1}^{p_1-1} \sin(U_j) \right)^\top$$

with independent $U_1 \sim \text{Unif}[0, \pi]$, \dots , $U_{p_1-2} \sim \text{Unif}[0, \pi]$, and $U_{p_1-1} \sim \text{Unif}[0, 2\pi]$. We counted the numbers of each test statistic exceeding $\chi_{p_1, \alpha}^2$, where

Table 1

Empirical average powers of Bartlett(-type) adjusted tests from Stern's adjusted likelihood; $(T_{LR}^{AP})^{\text{Bart}(N)} > \chi_{p_1, \alpha}^2$, $(T_R^{AP})^{\text{CF}_3(N)} > \chi_{p_1, \alpha}^2$, $(T_W^{AP})^{\text{CF}_3(N)} > \chi_{p_1, \alpha}^2$, $(T_R^{AP*})^{\text{CF}_2(N)} > \chi_{p_1, \alpha}^2$, and $(T_W^{AP*})^{\text{CF}_2(N)} > \chi_{p_1, \alpha}^2$, where $(N, p_1, p_2) = (60, 2, 2)$ and $\alpha = 0.01, 0.025, 0.05$. The bold-faced number indicates the largest average power in each row, except for $\lambda = 0.0$.

λ	$(T_{LR}^{AP})^{\text{Bart}(N)}$	$(T_R^{AP})^{\text{CF}_3(N)}$	$(T_R^{AP*})^{\text{CF}_2(N)}$	$(T_W^{AP})^{\text{CF}_3(N)}$	$(T_W^{AP*})^{\text{CF}_2(N)}$
(i) $\alpha = 0.01$					
0.0	0.0094	0.0099	0.0100	0.0087	0.0085
0.5	0.0226	0.0233	0.0227	0.0196	0.0197
1.0	0.0395	0.0392	0.0395	0.0333	0.0337
1.5	0.0583	0.0592	0.0593	0.0500	0.0504
2.0	0.0795	0.0806	0.0805	0.0677	0.0685
2.5	0.1050	0.1051	0.1055	0.0887	0.0893
3.0	0.1336	0.1326	0.1335	0.1137	0.1146
3.5	0.1640	0.1615	0.1626	0.1405	0.1415
4.0	0.1924	0.1900	0.1913	0.1656	0.1675
4.5	0.2258	0.2217	0.2234	0.1976	0.1989
5.0	0.2592	0.2528	0.2550	0.2274	0.2285
6.0	0.3285	0.3172	0.3201	0.2908	0.2927
7.0	0.3972	0.3816	0.3861	0.3570	0.3595
8.0	0.4650	0.4444	0.4501	0.4249	0.4258
(ii) $\alpha = 0.025$					
0.0	0.0245	0.0248	0.0248	0.0231	0.0229
0.5	0.0492	0.0499	0.0498	0.0452	0.0454
1.0	0.0784	0.0791	0.0787	0.0713	0.0718
1.5	0.1103	0.1117	0.1106	0.1002	0.1000
2.0	0.1439	0.1448	0.1444	0.1291	0.1300
2.5	0.1795	0.1809	0.1801	0.1627	0.1636
3.0	0.2180	0.2171	0.2179	0.1991	0.2000
3.5	0.2582	0.2554	0.2556	0.2358	0.2365
4.0	0.2959	0.2917	0.2931	0.2711	0.2719
4.5	0.3386	0.3313	0.3343	0.3113	0.3120
5.0	0.3778	0.3705	0.3721	0.3502	0.3512
6.0	0.4548	0.4438	0.4467	0.4281	0.4280
7.0	0.5286	0.5142	0.5174	0.5005	0.5017
8.0	0.5967	0.5780	0.5821	0.5682	0.5691
(iii) $\alpha = 0.05$					
0.0	0.0501	0.0499	0.0498	0.0485	0.0484
0.5	0.0878	0.0881	0.0880	0.0842	0.0836
1.0	0.1309	0.1315	0.1312	0.1231	0.1229
1.5	0.1754	0.1748	0.1751	0.1638	0.1646
2.0	0.2176	0.2185	0.2188	0.2047	0.2051
2.5	0.2640	0.2641	0.2644	0.2469	0.2476
3.0	0.3113	0.3084	0.3093	0.2927	0.2930
3.5	0.3578	0.3538	0.3554	0.3380	0.3385
4.0	0.3997	0.3958	0.3967	0.3790	0.3796
4.5	0.4463	0.4401	0.4414	0.4265	0.4267
5.0	0.4886	0.4808	0.4823	0.4684	0.4690
6.0	0.5677	0.5570	0.5590	0.5469	0.5471
7.0	0.6381	0.6253	0.6275	0.6185	0.6195
8.0	0.6980	0.6839	0.6865	0.6811	0.6814

- $(T_{\#}^{AP})^{\text{CF}_3(N)}$ or $T_{\#}^{\text{CF}_3(N)}$, $\# = R, W$, is the Bartlett-type corrected Rao/Wald statistic (11),
- $(T_{\#}^{AP*})^{\text{CF}_2(N)}$ or $(T_{\#}^*)^{\text{CF}_2(N)}$, $\# = R, W$, is the Bartlett-type corrected Rao/Wald statistic (15), and
- $(T_{LR}^{AP})^{\text{Bart}(N)}$ or $T_{LR}^{\text{Bart}(N)}$ is the Bartlett corrected LR test statistic (16).

From Tables 1 and 4, with $(p_1, p_2) = (2, 2)$, the following observations, in the average power sense, are made for the ordinary/Stern's adjusted profile likelihood inference. Firstly, each row in Tables 1 and 4 shows that, for small λ (not the whole λ), the Bartlett-type corrected Rao test $(T_R^{AP})^{\text{CF}_3(N)} > \chi_{p_1, \alpha}^2$ or $T_R^{\text{CF}_3(N)} > \chi_{p_1, \alpha}^2$ is superior to the other tests, as expected from the local optimality of Theorem 5 (here, "local" means that both α and λ are small enough). Secondly, for large λ , the Bartlett-corrected LR test $(T_{LR}^{AP})^{\text{Bart}(N)} > \chi_{p_1, \alpha}^2$ or $T_{LR}^{\text{Bart}(N)} > \chi_{p_1, \alpha}^2$ is superior to the other tests (we had no theoretical evidence). Thirdly, the Bartlett-type corrected Wald test $(T_W^{AP})^{\text{CF}_2(N)} > \chi_{p_1, \alpha}^2$ or $T_W^{\text{CF}_2(N)} > \chi_{p_1, \alpha}^2$ outperforms the Bartlett-type corrected Wald test $(T_W^{AP})^{\text{CF}_3(N)} > \chi_{p_1, \alpha}^2$ or $T_W^{\text{CF}_3(N)} > \chi_{p_1, \alpha}^2$ (we had its theoretical evidence, being opposite to Theorem 4 for the Rao case, on the basis of the inequality ${}_W U_{22}^C < {}_W U_{22}$ (see Section 4.1) that can be checked for the exponential regression model (27)).

Table 2

Empirical average powers of Bartlett(-type) adjusted tests from Stern's adjusted likelihood; $(T_{LR}^{AP})^{Bart(N)} > \chi_{p_1, \alpha}^2$, $(T_R^{AP})^{CF_3(N)} > \chi_{p_1, \alpha}^2$, $(T_W^{AP})^{CF_3(N)} > \chi_{p_1, \alpha}^2$, $(T_R^{AP*})^{CF_2(N)} > \chi_{p_1, \alpha}^2$, and $(T_W^{AP*})^{CF_2(N)} > \chi_{p_1, \alpha}^2$, where $(N, p_1, p_2) = (60, 4, 4)$ and $\alpha = 0.01, 0.025, 0.05$. The bold-faced number indicates the largest average power in each row, except for $\lambda = 0.0$.

λ	$(T_{LR}^{AP})^{Bart(N)}$	$(T_R^{AP})^{CF_3(N)}$	$(T_R^{AP*})^{CF_2(N)}$	$(T_W^{AP})^{CF_3(N)}$	$(T_W^{AP*})^{CF_2(N)}$
(i) $\alpha = 0.01$					
0.0	0.0101	0.0099	0.0100	0.0078	0.0075
0.5	0.0176	0.0181	0.0180	0.0131	0.0126
1.0	0.0275	0.0277	0.0281	0.0194	0.0187
1.5	0.0384	0.0399	0.0396	0.0270	0.0263
2.0	0.0520	0.0530	0.0537	0.0362	0.0353
2.5	0.0684	0.0696	0.0704	0.0471	0.0457
3.0	0.0854	0.0863	0.0873	0.0586	0.0572
3.5	0.1038	0.1041	0.1056	0.0713	0.0704
4.0	0.1253	0.1256	0.1283	0.0865	0.0849
4.5	0.1478	0.1464	0.1504	0.1033	0.1021
5.0	0.1714	0.1684	0.1720	0.1192	0.1180
6.0	0.2221	0.2167	0.2207	0.1596	0.1589
7.0	0.2754	0.2668	0.2722	0.2030	0.2008
8.0	0.3324	0.3182	0.3257	0.2505	0.2491
(ii) $\alpha = 0.025$					
0.0	0.0248	0.0249	0.0250	0.0220	0.0210
0.5	0.0402	0.0412	0.0410	0.0333	0.0322
1.0	0.0584	0.0591	0.0593	0.0470	0.0460
1.5	0.0771	0.0791	0.0787	0.0609	0.0601
2.0	0.0997	0.1026	0.1026	0.0785	0.0774
2.5	0.1252	0.1282	0.1287	0.0978	0.0964
3.0	0.1504	0.1531	0.1542	0.1184	0.1173
3.5	0.1784	0.1795	0.1810	0.1417	0.1394
4.0	0.2083	0.2086	0.2098	0.1653	0.1629
4.5	0.2389	0.2377	0.2391	0.1911	0.1894
5.0	0.2681	0.2665	0.2696	0.2163	0.2146
6.0	0.3330	0.3294	0.3320	0.2734	0.2713
7.0	0.3959	0.3869	0.3914	0.3297	0.3282
8.0	0.4577	0.4459	0.4517	0.3889	0.3866
(iii) $\alpha = 0.05$					
0.0	0.0501	0.0498	0.0495	0.0463	0.0450
0.5	0.0750	0.0756	0.0757	0.0659	0.0644
1.0	0.1018	0.1051	0.1046	0.0888	0.0872
1.5	0.1299	0.1315	0.1312	0.1112	0.1096
2.0	0.1620	0.1648	0.1657	0.1378	0.1369
2.5	0.1957	0.1982	0.1989	0.1667	0.1653
3.0	0.2289	0.2310	0.2319	0.1956	0.1934
3.5	0.2620	0.2641	0.2661	0.2255	0.2239
4.0	0.2994	0.2999	0.3023	0.2582	0.2562
4.5	0.3350	0.3339	0.3357	0.2909	0.2891
5.0	0.3689	0.3685	0.3705	0.3226	0.3209
6.0	0.4425	0.4376	0.4398	0.3904	0.3884
7.0	0.5058	0.4973	0.5013	0.4527	0.4516
8.0	0.5690	0.5571	0.5620	0.5159	0.5137

We conducted other simulations with $(p_1, p_2) = (4, 4), (6, 6)$ (see [Tables 2, 3, 5](#) and [6](#)), in order to illustrate what happens if the dimension p_1 of the interest parameter increases or the dimension p_2 of the nuisance parameter increases. Interestingly, [Tables 1–3](#) (or [Tables 4–6](#)) indicate that, as the dimension p_1 of the interest parameter increases, the region of λ , such that the Bartlett-type corrected Rao test $(T_R^{AP})^{CF_2(N)} > \chi_{p_1, \alpha}^2$ or $T_R^{CF_2(N)} > \chi_{p_1, \alpha}^2$ outperforms the Bartlett-corrected LR test in the average power sense, is spread out.

7. Concluding remarks

Stern [\[33\]](#) also proposed, instead of [\(3\)](#),

$$\mathcal{L}^{AP_2(N)}(\theta_{(1)}) = \mathcal{L}^{AP(N)}(\theta_{(1)}) - \frac{1}{2N} \check{M}_{b_1 b_2}^{\check{g} \check{g}}(\theta_{(1)}) \prod_{i=1}^2 [\check{v}_{(11.2)}^{-1}(\theta_{(1)}) \check{Z}_{(1)}^{(N)}(\theta_{(1)})]_{b_i} \quad (28)$$

([\[10\]](#) corrected a typo for the matrix $[\check{v}_{(11.2)}^{a,b}(\cdot) \check{g}_{k,b}(\cdot) \check{M}_{kk'}(\cdot) \check{g}_{k',b'}(\cdot) \check{v}_{(11.2)}^{b',a'}(\cdot)]_{a,a' \in \{1, \dots, p_1\}}$), which enables us to adjust not only the score bias, but also the information bias. Naturally, there may be a question about the third-order average local power

Table 3

Empirical average powers of Bartlett(-type) adjusted tests from Stern's adjusted likelihood; $(T_{LR}^{AP})^{\text{Bart}(N)} > \chi_{p_1, \alpha}^2$, $(T_R^{AP})^{\text{CF}_3(N)} > \chi_{p_1, \alpha}^2$, $(T_W^{AP})^{\text{CF}_3(N)} > \chi_{p_1, \alpha}^2$, $(T_R^{AP*})^{\text{CF}_2(N)} > \chi_{p_1, \alpha}^2$, and $(T_W^{AP*})^{\text{CF}_2(N)} > \chi_{p_1, \alpha}^2$, where $(N, p_1, p_2) = (60, 6, 6)$ and $\alpha = 0.01, 0.025, 0.05$. The bold-faced number indicates the largest average power in each row, except for $\lambda = 0.0$.

λ	$(T_{LR}^{AP})^{\text{Bart}(N)}$	$(T_R^{AP})^{\text{CF}_3(N)}$	$(T_R^{AP*})^{\text{CF}_2(N)}$	$(T_W^{AP})^{\text{CF}_3(N)}$	$(T_W^{AP*})^{\text{CF}_2(N)}$
(i) $\alpha = 0.01$					
0.0	0.0096	0.0097	0.0098	0.0044	0.0039
0.5	0.0150	0.0155	0.0159	0.0064	0.0057
1.0	0.0216	0.0219	0.0223	0.0087	0.0080
1.5	0.0305	0.0314	0.0324	0.0119	0.0109
2.0	0.0402	0.0414	0.0420	0.0162	0.0151
2.5	0.0511	0.0522	0.0529	0.0199	0.0185
3.0	0.0647	0.0645	0.0660	0.0249	0.0233
3.5	0.0782	0.0779	0.0800	0.0307	0.0289
4.0	0.0926	0.0927	0.0937	0.0369	0.0349
4.5	0.1103	0.1098	0.1122	0.0446	0.0423
5.0	0.1285	0.1281	0.1304	0.0521	0.0493
6.0	0.1668	0.1636	0.1696	0.0719	0.0689
7.0	0.2070	0.2021	0.2073	0.0937	0.0899
8.0	0.2533	0.2454	0.2513	0.1196	0.1152
(ii) $\alpha = 0.025$					
0.0	0.0252	0.0255	0.0253	0.0146	0.0135
0.5	0.0368	0.0372	0.0370	0.0205	0.0190
1.0	0.0489	0.0502	0.0505	0.0270	0.0255
1.5	0.0648	0.0670	0.0673	0.0358	0.0338
2.0	0.0811	0.0828	0.0835	0.0447	0.0423
2.5	0.1000	0.1009	0.1008	0.0544	0.0521
3.0	0.1207	0.1224	0.1230	0.0674	0.0646
3.5	0.1402	0.1424	0.1436	0.0782	0.0751
4.0	0.1614	0.1633	0.1653	0.0923	0.0885
4.5	0.1877	0.1914	0.1926	0.1068	0.1029
5.0	0.2127	0.2135	0.2161	0.1232	0.1192
6.0	0.2639	0.2637	0.2663	0.1582	0.1533
7.0	0.3146	0.3092	0.3133	0.1945	0.1900
8.0	0.3706	0.3639	0.3694	0.2359	0.2298
(iii) $\alpha = 0.05$					
0.0	0.0509	0.0509	0.0514	0.0353	0.0335
0.5	0.0689	0.0704	0.0702	0.0470	0.0445
1.0	0.0898	0.0923	0.0919	0.0603	0.0574
1.5	0.1124	0.1149	0.1158	0.0745	0.0712
2.0	0.1357	0.1384	0.1378	0.0907	0.0874
2.5	0.1612	0.1643	0.1644	0.1093	0.1056
3.0	0.1898	0.1930	0.1938	0.1284	0.1248
3.5	0.2156	0.2207	0.2216	0.1463	0.1421
4.0	0.2425	0.2456	0.2475	0.1668	0.1629
4.5	0.2759	0.2785	0.2797	0.1889	0.1845
5.0	0.3055	0.3079	0.3096	0.2132	0.2085
6.0	0.3650	0.3645	0.3667	0.2610	0.2553
7.0	0.4217	0.4184	0.4218	0.3088	0.3036
8.0	0.4812	0.4744	0.4777	0.3617	0.3562

of the resulting test statistic. It is not difficult to see that the LR, Rao, Wald, and gradient test statistics arising from (28) are members of the class $\mathcal{T}_{N,3,M}$, with the expressions for ${}_M D_{j_1 j_2}$ being different from those arising from $\mathcal{L}^{AP(N)}(\theta_{(1)}) = \mathcal{L}_M^{(N)}(\theta_{(1)})$. Noting that the third-order average local power of the GCF-(or GB-) adjusted test for $T^{(N)} \in \mathcal{T}_{N,3,M}$ is independent of the D -functions (see Remark 6(i)), we can say that such an information bias adjustment is not essential if an interest lies in the third-order average local power.

Instead of analytical size corrections using the Bartlett-type adjustments (see Definitions 1 and 2) or Cornish–Fisher's type asymptotic expansion for the percentile (see (22)), using the bootstrap method is an alternative. However, this is beyond the scope of this paper, because the (parametric) bootstrap approach would be computer-intensive (of course, the theory of this approach itself is implicitly related to the validity of the higher-order asymptotic expansion). The author has reported “bootstrap-based Bartlett-type adjustment” at the 2014 Japanese Joint Statistical Meeting (2014, September) and autumn meeting (2014) of the Mathematical Society of Japan. This topic will be discussed further elsewhere.

Acknowledgments

The author thanks the editor and referees for their comments on an earlier draft of the paper.

Table 4

Empirical average powers of Bartlett(-type) adjusted tests from the ordinary likelihood; $T_{LR}^{Bart(N)} > \chi_{p_1, \alpha}^2, T_R^{CF_3(N)} > \chi_{p_1, \alpha}^2, T_W^{CF_3(N)} > \chi_{p_1, \alpha}^2, (T_R^*)^{CF_2(N)} > \chi_{p_1, \alpha}^2$, and $(T_W^*)^{CF_2(N)} > \chi_{p_1, \alpha}^2$, where $(N, p_1, p_2) = (60, 2, 2)$ and $\alpha = 0.01, 0.025, 0.05$. The bold-faced number indicates the largest average power in each row, except for $\lambda = 0.0$.

λ	$T_{LR}^{Bart(N)}$	$T_R^{CF_3(N)}$	$(T_R^*)^{CF_2(N)}$	$T_W^{CF_3(N)}$	$(T_W^*)^{CF_2(N)}$
(i) $\alpha = 0.01$					
0.0	0.0095	0.0099	0.0100	0.0087	0.0086
0.5	0.0231	0.0232	0.0231	0.0198	0.0199
1.0	0.0394	0.0392	0.0395	0.0334	0.0336
1.5	0.0584	0.0592	0.0593	0.0501	0.0504
2.0	0.0811	0.0820	0.0818	0.0691	0.0699
2.5	0.1066	0.1066	0.1068	0.0912	0.0914
3.0	0.1342	0.1326	0.1338	0.1146	0.1151
3.5	0.1641	0.1615	0.1626	0.1406	0.1416
4.0	0.1943	0.1914	0.1922	0.1682	0.1701
4.5	0.2258	0.2216	0.2234	0.1977	0.1991
5.0	0.2594	0.2528	0.2551	0.2275	0.2285
6.0	0.3286	0.3171	0.3203	0.2911	0.2927
7.0	0.3973	0.3813	0.3861	0.3573	0.3596
8.0	0.4650	0.4442	0.4499	0.4250	0.4259
(ii) $\alpha = 0.025$					
0.0	0.0245	0.0249	0.0247	0.0231	0.0229
0.5	0.0500	0.0500	0.0499	0.0459	0.0460
1.0	0.0785	0.0791	0.0787	0.0713	0.0718
1.5	0.1103	0.1116	0.1106	0.1002	0.1000
2.0	0.1453	0.1447	0.1452	0.1308	0.1316
2.5	0.1816	0.1810	0.1809	0.1650	0.1653
3.0	0.2187	0.2173	0.2183	0.2001	0.2006
3.5	0.2583	0.2552	0.2557	0.2359	0.2364
4.0	0.2985	0.2941	0.2947	0.2733	0.2745
4.5	0.3387	0.3312	0.3344	0.3114	0.3121
5.0	0.3779	0.3706	0.3721	0.3503	0.3513
6.0	0.4548	0.4435	0.4466	0.4282	0.4281
7.0	0.5286	0.5139	0.5173	0.5008	0.5017
8.0	0.5967	0.5778	0.5821	0.5684	0.5692
(iii) $\alpha = 0.05$					
0.0	0.0502	0.0498	0.0498	0.0485	0.0483
0.5	0.0895	0.0898	0.0893	0.0848	0.0844
1.0	0.1309	0.1314	0.1314	0.1230	0.1231
1.5	0.1754	0.1746	0.1751	0.1639	0.1645
2.0	0.2188	0.2196	0.2198	0.2061	0.2064
2.5	0.2654	0.2647	0.2641	0.2496	0.2496
3.0	0.3119	0.3095	0.3105	0.2934	0.2938
3.5	0.3580	0.3537	0.3555	0.3379	0.3384
4.0	0.4025	0.3976	0.3995	0.3828	0.3831
4.5	0.4463	0.4399	0.4414	0.4265	0.4269
5.0	0.4885	0.4806	0.4820	0.4685	0.4691
6.0	0.5679	0.5569	0.5589	0.5471	0.5471
7.0	0.6381	0.6253	0.6276	0.6188	0.6196
8.0	0.6981	0.6838	0.6865	0.6814	0.6813

Appendix

We give the outlined proof of Proposition 1. By definition, we can rewrite the GB case (10) (also (9)) as

$$T_A^{GB\Delta(N)} = T_A^{*(N)} + \frac{2}{N} \sum_{R=2,4} \tilde{F}_{b_1 \dots b_R}^{GB} \prod_{i=1}^R [\tilde{v}_{(11.2)}^{-1} \tilde{z}_{(1)}^{(N)}]_{b_i} = (T_A^*)^{GCF(N)} \quad (\text{A.1})$$

(we set $\Gamma_{a_1 \dots a_6}(\cdot) \equiv 0$ in Definition 1), where

$$\begin{aligned} T_A^{*\Delta(N)} = & T_A^{(N)} + \frac{2}{N^{1/2}} \tilde{F}_{b_1 b_2 b_3}^C \prod_{i=1}^3 [\tilde{v}_{(11.2)}^{-1} \tilde{z}_{(1)}^{(N)}]_{b_i} \\ & + \frac{2}{N} \left(\tilde{\Delta}_{b_1 b_2 b_3 b_4}^{GB} \prod_{i=1}^4 [\tilde{v}_{(11.2)}^{-1} \tilde{z}_{(1)}^{(N)}]_{b_i} + \tilde{\Delta}_{b_1 b_2 b_3, k_1 k_2}^{GB} \tilde{z}_{k_1 k_2}^{(N)} \prod_{i=1}^3 [\tilde{v}_{(11.2)}^{-1} \tilde{z}_{(1)}^{(N)}]_{b_i} \right). \end{aligned}$$

Table 5

Empirical average powers of Bartlett(-type) adjusted tests from the ordinary likelihood; $T_{LR}^{\text{Bart}(N)} > \chi_{p_1, \alpha}^2, T_R^{\text{CF}_3(N)} > \chi_{p_1, \alpha}^2, T_W^{\text{CF}_3(N)} > \chi_{p_1, \alpha}^2, (T_R^*)^{\text{CF}_2(N)} > \chi_{p_1, \alpha}^2, (T_W^*)^{\text{CF}_2(N)} > \chi_{p_1, \alpha}^2$, and $(T_W^*)^{\text{CF}_2(N)} > \chi_{p_1, \alpha}^2$, where $(N, p_1, p_2) = (60, 4, 4)$ and $\alpha = 0.01, 0.025, 0.05$. The bold-faced number indicates the largest average power in each row, except for $\lambda = 0.0$.

λ	$T_{LR}^{\text{Bart}(N)}$	$T_R^{\text{CF}_3(N)}$	$(T_R^*)^{\text{CF}_2(N)}$	$T_W^{\text{CF}_3(N)}$	$(T_W^*)^{\text{CF}_2(N)}$
(i) $\alpha = 0.01$					
0.0	0.0101	0.0098	0.0099	0.0078	0.0075
0.5	0.0174	0.0176	0.0178	0.0129	0.0121
1.0	0.0265	0.0272	0.0280	0.0192	0.0187
1.5	0.0381	0.0390	0.0395	0.0266	0.0257
2.0	0.0514	0.0527	0.0534	0.0353	0.0343
2.5	0.0673	0.0685	0.0694	0.0458	0.0448
3.0	0.0853	0.0854	0.0866	0.0574	0.0567
3.5	0.1041	0.1043	0.1060	0.0707	0.0695
4.0	0.1240	0.1240	0.1259	0.0858	0.0841
4.5	0.1454	0.1439	0.1469	0.1021	0.1006
5.0	0.1681	0.1654	0.1693	0.1197	0.1179
6.0	0.2187	0.2122	0.2172	0.1565	0.1556
7.0	0.2719	0.2630	0.2686	0.1997	0.1976
8.0	0.3278	0.3147	0.3231	0.2467	0.2449
(ii) $\alpha = 0.025$					
0.0	0.0248	0.0251	0.0250	0.0221	0.0210
0.5	0.0398	0.0405	0.0405	0.0330	0.0320
1.0	0.0573	0.0588	0.0584	0.0464	0.0452
1.5	0.0781	0.0796	0.0791	0.0621	0.0607
2.0	0.1003	0.1015	0.1023	0.0786	0.0769
2.5	0.1241	0.1259	0.1262	0.0980	0.0963
3.0	0.1496	0.1506	0.1513	0.1182	0.1167
3.5	0.1763	0.1765	0.1782	0.1404	0.1387
4.0	0.2049	0.2046	0.2067	0.1638	0.1622
4.5	0.2353	0.2337	0.2362	0.1891	0.1872
5.0	0.2658	0.2638	0.2665	0.2156	0.2128
6.0	0.3287	0.3241	0.3286	0.2705	0.2682
7.0	0.3943	0.3848	0.3898	0.3277	0.3249
8.0	0.4561	0.4437	0.4494	0.3859	0.3836
(iii) $\alpha = 0.05$					
0.0	0.0501	0.0497	0.0497	0.0464	0.0448
0.5	0.0740	0.0755	0.0752	0.0664	0.0649
1.0	0.1015	0.1026	0.1023	0.0881	0.0864
1.5	0.1302	0.1323	0.1320	0.1120	0.1108
2.0	0.1605	0.1624	0.1626	0.1378	0.1364
2.5	0.1922	0.1937	0.1946	0.1653	0.1639
3.0	0.2265	0.2272	0.2285	0.1945	0.1932
3.5	0.2607	0.2628	0.2634	0.2256	0.2233
4.0	0.2966	0.2964	0.2987	0.2568	0.2545
4.5	0.3319	0.3315	0.3341	0.2883	0.2863
5.0	0.3684	0.3659	0.3688	0.3208	0.3188
6.0	0.4392	0.4344	0.4366	0.3864	0.3841
7.0	0.5056	0.4970	0.5014	0.4514	0.4486
8.0	0.5677	0.5571	0.5614	0.5134	0.5110

In fact, compared with the original form (6), the $C^{\mathcal{G}}, D^{\mathcal{G}}$ -functions associated with $T_A^{*\Delta(N)} \in \mathcal{T}_{N,3,A}$ should read as $C_{a_1 a_2 a_3}^{\mathcal{G}}(\cdot) + \Gamma_{a_1 a_2 a_3}^C(\cdot), D_{a_1 a_2 a_3 a_4}^{\mathcal{G}}(\cdot) + \Delta_{a_1 a_2 a_3 a_4}^{GB}(\cdot)$, and $D_{a_1 a_2 a_3, k_1 k_2}^{\mathcal{G}}(\cdot) + \Delta_{a_1 a_2 a_3, k_1 k_2}^{GB}(\cdot)$, while the others do not change. Thus, the GB adjustment (Definition 2) can be treated as the GCF adjustment (Definition 1). We consider the GCF case only.

Recall that, for the case $\mathbf{A}(\cdot) \equiv \mathbf{0}_p$, Kakizawa [16,17,20] treated

$$U_a^{T_{2,4,6}^{(N)}} = [\mathbf{Z}_{(1)}^{0(N)}]_a + \frac{1}{N^{1/2}} U_a^{C(N)} + \frac{1}{N} \left(U_a^{CD(N)} + \sum_{R=2,4,6} \Gamma_{b_1 \dots b_{R-1} a} \prod_{i=1}^{R-1} [\mathbf{v}_{(11-2)}^{-1} \mathbf{Z}_{(1)}^{0(N)}]_{b_i} \right), \quad (\text{A.2})$$

where $a = 1, \dots, p_1$. Note that $U_a^{C(N)}$ is a certain quadratic polynomial in $\mathbf{Z}_{(1)}^{0(N)}, \rho^{0(N)}$, and $[\mathbf{Z}_{j_1 j_2}^{\perp(N)}]_{j_1, j_2 \in \{1, \dots, p\}}$, and that $U_a^{CD(N)}$ is a certain cubic polynomial in $\mathbf{Z}_{(1)}^{0(N)}, \rho^{0(N)}$, and $[\mathbf{Z}_{j_1 j_2}^{\perp(N)}, \mathbf{Z}_{j_1 j_2 j_3}^{\perp(N)}]_{j_1, j_2, j_3 \in \{1, \dots, p\}}$, where $\mathbf{Z}_{(1)}^{0(N)} = \mathcal{G}^T \mathbf{Z}^{(N)}, \rho^{0(N)} = \mathbf{v}_{(22)}^{-1} \mathbf{Z}_{(2)}^{(N)}$, and $\mathbf{Z}_{j_1 \dots j_R}^{\perp(N)} = \mathbf{Z}_{j_1 \dots j_R}^{(N)} - \mathbf{v}_{j_1 \dots j_R, k} \mathbf{v}^{k, k'} \mathbf{Z}_{k'}^{(N)}, R = 2, 3$. In the present framework (see (6) and (8)), introducing

$$AD_{a_1 a_2}^{+\mathcal{G}}(\cdot) = AD_{a_1 a_2}^{\mathcal{G}}(\cdot) + \frac{\langle 2 \rangle}{2} AD_{a_1, k_1 k_2}^{\mathcal{G}}(\cdot) \mathbf{v}_{k_1 k_2, a_2}^{\mathcal{G}}(\cdot),$$

Table 6

Empirical average powers of Bartlett(-type) adjusted tests from the ordinary likelihood; $T_{LR}^{Bart(N)} > \chi^2_{p_1, \alpha}, T_R^{CF_3(N)} > \chi^2_{p_1, \alpha}, T_W^{CF_3(N)} > \chi^2_{p_1, \alpha}, (T_R^*)^{CF_2(N)} > \chi^2_{p_1, \alpha}, (T_W^*)^{CF_2(N)} > \chi^2_{p_1, \alpha}$, and $(T_W^*)^{CF_2(N)} > \chi^2_{p_1, \alpha}$, where $(N, p_1, p_2) = (60, 6, 6)$ and $\alpha = 0.01, 0.025, 0.05$. The bold-faced number indicates the largest average power in each row, except for $\lambda = 0.0$.

λ	$T_{LR}^{Bart(N)}$	$T_R^{CF_3(N)}$	$(T_R^*)^{CF_2(N)}$	$T_W^{CF_3(N)}$	$(T_W^*)^{CF_2(N)}$
(i) $\alpha = 0.01$					
0.0	0.0096	0.0098	0.0097	0.0046	0.0039
0.5	0.0155	0.0156	0.0157	0.0065	0.0058
1.0	0.0224	0.0227	0.0233	0.0090	0.0082
1.5	0.0305	0.0314	0.0315	0.0120	0.0109
2.0	0.0399	0.0410	0.0421	0.0155	0.0144
2.5	0.0511	0.0523	0.0530	0.0196	0.0183
3.0	0.0638	0.0642	0.0660	0.0248	0.0227
3.5	0.0780	0.0778	0.0803	0.0305	0.0282
4.0	0.0942	0.0937	0.0963	0.0372	0.0345
4.5	0.1108	0.1103	0.1130	0.0449	0.0416
5.0	0.1289	0.1270	0.1299	0.0534	0.0498
6.0	0.1673	0.1638	0.1680	0.0719	0.0682
7.0	0.2103	0.2030	0.2091	0.0951	0.0903
8.0	0.2552	0.2465	0.2519	0.1212	0.1160
(ii) $\alpha = 0.025$					
0.0	0.0252	0.0255	0.0254	0.0148	0.0134
0.5	0.0364	0.0373	0.0375	0.0206	0.0192
1.0	0.0491	0.0506	0.0510	0.0272	0.0256
1.5	0.0640	0.0661	0.0664	0.0354	0.0332
2.0	0.0814	0.0825	0.0835	0.0442	0.0418
2.5	0.1003	0.1023	0.1033	0.0550	0.0517
3.0	0.1204	0.1223	0.1241	0.0667	0.0633
3.5	0.1421	0.1440	0.1448	0.0794	0.0757
4.0	0.1651	0.1661	0.1677	0.0933	0.0894
4.5	0.1886	0.1887	0.1912	0.1083	0.1042
5.0	0.2136	0.2129	0.2155	0.1240	0.1196
6.0	0.2646	0.2634	0.2652	0.1587	0.1538
7.0	0.3180	0.3124	0.3168	0.1973	0.1914
8.0	0.3733	0.3641	0.3694	0.2394	0.2327
(iii) $\alpha = 0.05$					
0.0	0.0509	0.0508	0.0512	0.0357	0.0334
0.5	0.0691	0.0705	0.0704	0.0471	0.0444
1.0	0.0896	0.0919	0.0913	0.0602	0.0574
1.5	0.1128	0.1148	0.1157	0.0752	0.0722
2.0	0.1373	0.1400	0.1404	0.0914	0.0881
2.5	0.1638	0.1671	0.1670	0.1091	0.1051
3.0	0.1907	0.1932	0.1937	0.1277	0.1234
3.5	0.2186	0.2210	0.2217	0.1477	0.1435
4.0	0.2468	0.2493	0.2509	0.1688	0.1642
4.5	0.2761	0.2781	0.2798	0.1913	0.1863
5.0	0.3058	0.3065	0.3084	0.2146	0.2093
6.0	0.3670	0.3650	0.3672	0.2635	0.2572
7.0	0.4264	0.4212	0.4245	0.3136	0.3074
8.0	0.4841	0.4751	0.4798	0.3652	0.3594

it is shown, with some extra algebra, that

$$T_A^{GCF(N)} = U_{a,A}^{F_{2,4,6}^{(N)}} v_{(11,2)}^{a,b} U_{b,A}^{F_{2,4,6}^{(N)}} + \frac{1}{N^{3/2}} o_{\theta^\dagger}^{(N)} (1 + \min(\delta/2, \xi), \max(9/2, \beta)),$$

where

$$\begin{aligned}
 U_{a,A}^{F_{2,4,6}^{(N)}} &= [Z_{(1)}^{0(N)}]_a + \frac{1}{N^{1/2}} (U_a^{C(N)} + A_a^g) \\
 &+ \frac{1}{N} \left[U_a^{CD(N)} + \left\{ (A_a^g)_r + \frac{1}{2} A_b^g v_{(11,2)}^{b,b'} v_{a,b'/r}^{g,g} \right\} [\rho^{0(N)}]_r \right. \\
 &+ (A^{D+\frac{g}{2}}_{b_1 a} - C^{+\frac{g}{2}}_{b_1 a b} v_{(11,2)}^{b,b'} A_{b'}^g) [v_{(11,2)}^{-1} Z_{(1)}^{0(N)}]_{b_1} + (A^{D\frac{g}{2}}_{a b, k_1 k_2} - C^{g\frac{g}{2}}_{a b, k_1 k_2} v_{(11,2)}^{b,b'} A_{b'}^g) Z_{k_1 k_2}^{\perp(N)} \\
 &\left. + \sum_{R=2,4,6} \Gamma_{b_1 \dots b_{R-1} a} \prod_{i=1}^{R-1} [v_{(11,2)}^{-1} Z_{(1)}^{0(N)}]_{b_i} \right], \tag{A.3}
 \end{aligned}$$

$a = 1, \dots, p_1$, with $v_{a_1 a_2 / k}^{\mathcal{G} \mathcal{G}} = \mathcal{G}_{j_1, a_1} \mathcal{G}_{j_2, a_2} (\partial / \partial \theta_k) v_{j_1 j_2}(\boldsymbol{\theta}^\dagger) = -(\partial / \partial \theta_k) v_{(11-2) a_1, a_2}(\boldsymbol{\theta}^\dagger)$. In view of (A.2) and (A.3), we observe that $U_{a,A}^{F_{2,4,6}^{(N)}} - U_a^{F_{2,4,6}^{(N)}}$ is the sum of $N^{-1/2} A_a^{\mathcal{G}}$ (nonrandom) and the random linear N^{-1} -terms, and hence, the cumulants of $U_{a,A}^{F_{2,4,6}^{(N)}}$ are almost the same as those of $U_a^{F_{2,4,6}^{(N)}}$; see [20]. Here, we omit the detailed derivation for the third-order asymptotic expansion of the null/non-null distribution of $U_{a,A}^{F_{2,4,6}^{(N)}} v_{(11-2) a, b}^{a, b} U_{b,A}^{F_{2,4,6}^{(N)}}$ via the valid Edgeworth expansion of $[U_{a,A}^{F_{2,4,6}^{(N)}}]_{a=1, \dots, p_1}$ in the sense of Bhattacharya and Ghosh [2] (see also [4,22]).

References

- [1] M.S. Bartlett, Properties of sufficiency and statistical tests, *Proc. R. Soc. Lond. Ser. A* 160 (1937) 268–282.
- [2] R.N. Bhattacharya, J.K. Ghosh, On the validity of the formal Edgeworth expansion, *Ann. Statist.* 6 (1978) 434–451. Correction: (1980). 8 1399.
- [3] P.J. Bickel, J.K. Ghosh, A decomposition for the likelihood ratio statistic and the Bartlett correction – A Bayesian argument, *Ann. Statist.* 18 (1990) 1070–1090.
- [4] R.N. Bhattacharya, R.R. Rao, *Normal Approximation and Asymptotic Expansions*, Wiley, New York, 1976.
- [5] T.K. Chandra, R. Mukerjee, Bartlett-type modification for Rao's efficient score statistic, *J. Multivariate Anal.* 36 (1991) 103–112.
- [6] G.M. Cordeiro, S.L.P. Ferrari, A modified score test statistic having chi-squared distribution to order n^{-1} , *Biometrika* 78 (1991) 573–582.
- [7] D.R. Cox, Some aspects of conditional and asymptotic inference: A review, *Sankhyā A* 50 (1988) 314–337.
- [8] D.R. Cox, N. Reid, Parameter orthogonality and approximate conditional inference (with discussion), *J. R. Stat. Soc. Ser. B* 49 (1987) 1–39.
- [9] T.J. DiCiccio, S.E. Stern, Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods, *J. R. Stat. Soc. Ser. B* 56 (1994) 397–408.
- [10] S.L.P. Ferrari, F. Lucambio, F. Cribari-Neto, Improved profile likelihood inference, *J. Statist. Plann. Inference* 134 (2005) 373–391.
- [11] J.K. Ghosh, R. Mukerjee, Adjusted versus conditional likelihood: Power properties and Bartlett-type adjustment, *J. R. Stat. Soc. Ser. B* 56 (1994) 185–188.
- [12] T. Hayakawa, The likelihood ratio criterion for a composite hypothesis under a local alternative, *Biometrika* 62 (1975) 451–460.
- [13] T. Hayakawa, The likelihood ratio criterion and the asymptotic expansion of its distribution, *Ann. Inst. Statist. Math.* 29 (1977) 359–378. Correction: (1987). 39 681.
- [14] Y. Kakizawa, Higher order monotone Bartlett-type adjustment for some multivariate test statistics, *Biometrika* 83 (1996) 923–927.
- [15] Y. Kakizawa, Comparison of Bartlett-type adjusted tests in the multiparameter case, *J. Multivariate Anal.* 101 (2010) 1638–1655.
- [16] Y. Kakizawa, Improved chi-squared tests for a composite hypothesis, *J. Multivariate Anal.* 107 (2012) 141–161.
- [17] Y. Kakizawa, Generalized Cordeiro–Ferrari Bartlett-type adjustment, *Statist. Probab. Lett.* 82 (2012) 2008–2016.
- [18] Y. Kakizawa, Second-order powers of a class of tests in the presence of a nuisance parameter, *Comm. Statist. Theory Methods* 41 (2012) 3676–3691.
- [19] Y. Kakizawa, Third-order local power properties of tests for a composite hypothesis, *J. Multivariate Anal.* 114 (2013) 303–317.
- [20] Y. Kakizawa, Third-order local power properties of tests for a composite hypothesis, II, *J. Multivariate Anal.* 140 (2015) 99–112.
- [21] D.N. Lawley, A general method for approximating to the distribution of likelihood ratio criteria, *Biometrika* 43 (1956) 295–303.
- [22] M.A. Magdalinos, Stochastic expansions and asymptotic approximations, *Econometric Theory* 8 (1992) 343–367.
- [23] R. Mukerjee, Comparison of tests in the presence of a nuisance parameter, in: Y. Dodge (Ed.), *Statistical Data Analysis and Inference*, North-Holland, Amsterdam, 1989, pp. 131–139.
- [24] R. Mukerjee, Existence of a Bartlett-type modification for a conditional version of Rao's statistic, *J. Quant. Econ.* 7 (1991) 231–238.
- [25] R. Mukerjee, Conditional likelihood and power: Higher-order asymptotics, *Proc. R. Soc. Lond. Ser. A* 438 (1992) 433–446.
- [26] R. Mukerjee, An extension of the conditional likelihood ratio test to the general multiparameter case, *Ann. Inst. Statist. Math.* 45 (1993) 759–771.
- [27] R. Mukerjee, Rao's score test: Recent asymptotic results, in: G.S. Maddala, C.R. Rao, H.D. Vinod (Eds.), *Handbook of Statistics*, Vol. 11, North-Holland, Amsterdam, 1993, pp. 363–379.
- [28] R. Mukerjee, Comparison of tests in their original forms, *Sankhyā A* 56 (1994) 118–127.
- [29] R. Mukerjee, T.K. Chandra, Bartlett-type adjustment for the conditional likelihood ratio statistic of Cox and Reid, *Biometrika* 78 (1991) 365–372.
- [30] C.R. Rao, R. Mukerjee, Comparison of Bartlett-type adjustments for the efficient score statistic, *J. Statist. Plann. Inference* 46 (1995) 137–146.
- [31] C.R. Rao, R. Mukerjee, Comparison of LR, score, and Wald tests in a non-iid setting, *J. Multivariate Anal.* 60 (1997) 99–110.
- [32] T.A. Severini, *Likelihood Methods in Statistics*, Oxford University Press, Oxford, 2000.
- [33] S.E. Stern, A second-order adjustment to the profile likelihood in the case of a multidimensional parameter of interest, *J. R. Stat. Soc. Ser. B* 59 (1997) 653–665.
- [34] K. Tamaki, Second order asymptotic properties of a class of test statistics under the existence of nuisance parameters, *Sci. Math. Jpn.* 61 (2005) 119–144.
- [35] M. Taniguchi, Third-order asymptotic properties of a class of test statistics under a local alternative, *J. Multivariate Anal.* 37 (1991) 223–238.
- [36] G.R. Terrell, The gradient statistic, *Comput. Sci. Statist.* 34 (2002) 206–215.