Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 99 (2008) 2265–2284
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Preliminary test estimators and phi-divergence measures in generalized linear models with binary data

M.L. Menéndez[a,*], L. Pardo[b], M.C. Pardo[b]

[a] *Department of Applied Mathematics, E.T.S.A.M., Technical University of Madrid, Spain*
[b] *Department of Statistics and O.R. I, Complutense University of Madrid, Spain*

Received 12 February 2007
Available online 12 February 2008

Abstract

We consider the problem of estimation of the parameters in Generalized Linear Models (GLM) with binary data when it is suspected that the parameter vector obeys some exact linear restrictions which are linearly independent with some degree of uncertainty. Based on minimum ϕ-divergence estimation ($M\phi E$), we consider some estimators for the parameters of the GLM: Unrestricted $M\phi E$, restricted $M\phi E$, Preliminary $M\phi E$, Shrinkage $M\phi E$, Shrinkage preliminary $M\phi E$, James–Stein $M\phi E$, Positive-part of Stein-Rule $M\phi E$ and Modified preliminary $M\phi E$. Asymptotic bias as well as risk with a quadratic loss function are studied under contiguous alternative hypotheses. Some discussion about dominance among the estimators studied is presented. Finally, a simulation study is carried out.

AMS 2000 subject classifications: 62J12; 62J07; 62F12; 62F30

Keywords: ϕ-divergence measures; Minimum ϕ-divergence estimator; ϕ-divergence statistics; Preliminary test estimator; Contiguous alternative hypotheses; Asymptotic bias; Asymptotic quadratic risk

1. Introduction

Let $Y_i, i = 1, \ldots, I$, independent binomial random variables with parameters π_i and n_i, $i = 1, \ldots, I$. We shall assume that the parameters $\pi_i = \Pr(Y_i = 1), i = 1, \ldots, I$, depend on the unknown parameters $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)^{\mathrm{T}}$ and explanatory variables $\mathbf{x}_i^{\mathrm{T}} = (x_{i0}, \ldots, x_{ik})\,, x_{i0} = 1, i = 1, \ldots, I$ through, the linear predictor

* Corresponding author.
E-mail address: ml.menendez@upm.es (M.L. Menéndez).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2008.02.011

$$\eta_i \equiv g(\pi_i) = \sum_{j=0}^{k} x_{ij}\beta_j, \quad i = 1, \ldots, I. \tag{1}$$

Here g is the link function. Unless restrictions are imposed on $\boldsymbol{\beta}$, we have $-\infty < \eta_i < \infty, i = 1, \ldots, I$. We denote by $\mathbf{X}$ the $I \times (k+1)$ matrix with rows $\mathbf{x}_i^{\mathrm{T}}, i = 1, \ldots, I$. We also shall assume that rank($\mathbf{X}$) $= k+1$. The function g maps the unit interval onto the whole real line $(-\infty, \infty)$. So we consider Generalized Linear Models (GLM) with binary data. Link functions $\eta_i = g(\pi_i)$ can be any monotonic, differentiable function; however, in practice, only a small set of link functions are actually used. In particular, links are chosen such that the inverse link $\pi_i = g^{-1}(\eta_i)$ is easily computed. Some link functions can be seen in McCullagh and Nelder [10] in p. 108. In the following, we denote $\pi_i \equiv \pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)$.

Let y_i be the number of "successes" associated with the binomial random variable $Y_i, i = 1, \ldots, I$. The maximum likelihood estimator (MLE), $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(Y_1, \ldots, Y_I)$, of the true value of the parameter $\boldsymbol{\beta}_0$ maximizes the expression

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{I} \log\left(\pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)^{y_i}\left(1 - \pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)\right)^{n_i - y_i}\right),$$

i.e.,

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}\in\Theta} l(\boldsymbol{\beta}), \tag{2}$$

where

$$\Theta = \{\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k) : \beta_i \in (-\infty, \infty),\ i = 0, \ldots, k\}.$$

If we denote

$$\widehat{\mathbf{p}}_i = \left(\frac{y_i}{n_i}, \frac{n_i - y_i}{n_i}\right)^{\mathrm{T}} \quad \text{and} \quad \boldsymbol{\pi}_i(\boldsymbol{\beta}) = \left(\pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right), 1 - \pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)\right)^{\mathrm{T}}, \tag{3}$$

$i = 1, \ldots, I$, we have

$$l(\boldsymbol{\beta}) = -\sum_{i=1}^{I} n_i D_{\mathrm{Kull}}\left(\widehat{\mathbf{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\beta})\right) - \sum_{i=1}^{I} n_i H\left(\widehat{\mathbf{p}}_i\right),$$

where $D_{\mathrm{Kull}}\left(\widehat{\mathbf{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\beta})\right)$ is the Kullback–Leibler divergence between the probability vectors $\widehat{\mathbf{p}}_i$ and $\boldsymbol{\pi}_i(\boldsymbol{\beta})$, defined in (3), and $H\left(\widehat{\mathbf{p}}_i\right)$ is the Shannon entropy associated with the probability vector $\widehat{\mathbf{p}}_i$. Their expressions are

$$D_{\mathrm{Kull}}\left(\widehat{\mathbf{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\beta})\right) = \frac{y_i}{n_i}\log\frac{\frac{y_i}{n_i}}{\pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)} + \frac{n_i - y_i}{n_i}\log\frac{\frac{n_i - y_i}{n_i}}{1 - \pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)}$$

and

$$H\left(\widehat{\mathbf{p}}_i\right) = -\frac{y_i}{n_i}\log\frac{y_i}{n_i} - \frac{n_i - y_i}{n_i}\log\frac{n_i - y_i}{n_i},$$

respectively. Therefore the MLE, defined in (2), can be alternatively defined by

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\Theta} \sum_{i=1}^{I} n_i D_{\mathrm{Kull}}\left(\widehat{\mathbf{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\beta})\right), \tag{4}$$

because $H\left(\widehat{\mathbf{p}}_i\right)$ does not depend on $\boldsymbol{\beta}, i = 1, \ldots, I$. For more details about Kullback–Leibler divergence measure, see Kullback [7].

A new class of estimators can be obtained if we replace the Kullback–Leibler divergence in (4) for a general family of divergence measures. One of the most known generalizations of the Kullback–Leibler divergence is the ϕ-divergence measure introduced by Csiszár [4] and Ali and Silvey [1], simultaneously. The ϕ-divergence measure between the probability vectors $\widehat{\mathbf{p}}_i$ and $\boldsymbol{\pi}_i(\boldsymbol{\beta})$ is given by

$$D_\phi\left(\widehat{\mathbf{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\beta})\right) \equiv \pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)\phi\left(\frac{y_i}{\pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)n_i}\right) + \left(1 - \pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)\right)\phi\left(\frac{n_i - y_i}{\left(1 - \pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)\right)n_i}\right),$$

$\phi \in \Phi^*$, Φ^* is the class of all convex functions $\phi(x), x > 0$, such that at $x = 1, \phi(1) = \phi'(1) = 0, \phi''(1) > 0$. In the following, we shall assume the conventions $0\phi(0/0) = 0$ and $0\phi(p/0) = p \lim_{u\to\infty} \phi(u)/u$. For a systematic study of ϕ-divergences see Pardo [11] and Vajda [18].

As a natural extension of the maximum likelihood estimator, given in (4) it is possible to consider the minimum ϕ-divergence estimator given by

$$\widehat{\boldsymbol{\beta}}_\phi \equiv \arg\min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}} \sum_{i=1}^{I} n_i D_\phi\left(\widehat{\mathbf{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\beta})\right).$$

In the following, we refer to $\widehat{\boldsymbol{\beta}}_\phi$ as the *unrestricted minimum ϕ-divergence estimator* of the true value of the parameter $\boldsymbol{\beta}_0$.

Now we assume that non-sample prior information on the value of $\boldsymbol{\beta}_0$ is available, either from previous studies or from the practical experience of the researchers or experts. Let the non-sample prior information be expressed by the subset Θ_0 of Θ defined by

$$\Theta_0 = \left\{\boldsymbol{\beta} \in \Theta / \mathbf{K}^{\mathrm{T}}\boldsymbol{\beta} = \mathbf{m}\right\},$$

where $\mathbf{K}^{\mathrm{T}}$ is any matrix of r rows and $k+1$ columns and $\mathbf{m}$ is a vector, of order r of specified constants; we can define the minimum ϕ-divergence estimator restricted to Θ_0 by

$$\widehat{\boldsymbol{\beta}}_\phi^{H_0} \equiv \arg\min_{\boldsymbol{\beta}\in\Theta_0} \sum_{i=1}^{I} n_i D_\phi\left(\widehat{\mathbf{p}}_i, \boldsymbol{\pi}_i(\boldsymbol{\beta})\right).$$

We refer to it as the *restricted minimum ϕ-divergence estimator*. There is only the limitation on $\mathbf{K}^{\mathrm{T}}$ in the sense that it must have full row rank, i.e., rank$(\mathbf{K}^{\mathrm{T}}) = r$.

If we know that $\boldsymbol{\beta}_0 \in \Theta$, without any additional information, we shall estimate $\boldsymbol{\beta}_0$ using the unrestricted minimum ϕ-divergence estimator $\widehat{\boldsymbol{\beta}}_\phi$, and if we are completely sure about $\boldsymbol{\beta}_0 \in \Theta_0$ we shall estimate $\boldsymbol{\beta}_0$ using the restricted minimum ϕ-divergence estimator, $\widehat{\boldsymbol{\beta}}_\phi^{H_0}$. If we have some "uncertainty" about if $\boldsymbol{\beta}_0 \in \Theta_0$, a better procedure for estimating $\boldsymbol{\beta}_0$ will be to use a "preliminary test estimator". It is well known that preliminary test estimation of parameters was introduced in the literature to estimate parameters of a model when it is suspected that some "uncertain prior information" on the parameter of interest is available. In this paper we introduce preliminary test estimators about the parameter $\boldsymbol{\beta}$ of the GLM when it is a priori suspected that the GLM parameters belong to the subspace of the parameter space determined by $\mathbf{K}^{\mathrm{T}}\boldsymbol{\beta} = \mathbf{m}$. Preliminary test estimators involve a statistical hypothesis test of the "uncertain prior information", in our case $H_0 : \mathbf{K}^{\mathrm{T}}\boldsymbol{\beta} = \mathbf{m}$, based on an appropriate test statistic. We shall consider for testing

$H_0 : \mathbf{K}^{\mathrm{T}}\boldsymbol{\beta} = \mathbf{m}$, a family of test statistics based on ϕ-divergence measures that in some sense generalizes the likelihood ratio test. On the basis of this test a decision on using the unrestricted minimum ϕ-divergence estimator or the restricted minimum ϕ-divergence estimator or both of them simultaneously will be taken. Preliminary test estimators were introduced by Bancroft [2] and studied later for many different authors in different problems. Preliminary test estimators of the Stein-type were introduced in Stein [17] and James and Stein [6] and expanded by Saleh and Sen [13,14] and Sen and Saleh [16] in the nonparametric context. A very nice state of the art about this type of estimators in many different problems can be seen in Saleh [15].

Section 2 is devoted to introducing a very wide family of preliminary test estimators as alternatives to $\widehat{\boldsymbol{\beta}}_\phi$ and $\widehat{\boldsymbol{\beta}}_\phi^{H_0}$. In Section 3, we obtain the asymptotic bias of them under the null hypothesis as well as under contiguous alternative hypotheses. Their asymptotic quadratic risk and different relations among them are studied in Section 4. Finally, a simulation study is carried out in Section 5 in order to analyze the behavior of the different preliminary test estimators for small samples.

2. Alternative estimators

We denote $N = \sum_{i=1}^{I} n_i$,

$$\mathbf{W}_N = \operatorname{diag}\left(\left(\frac{n_i}{N\pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0\right)\left(1-\pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0\right)\right)}\left(\frac{\partial\pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0\right)}{\partial\eta_i}\right)^2\right)_{i=1,\ldots,I}\right)$$

and by $\widehat{\mathbf{p}}$ and $\mathbf{p}\left(\boldsymbol{\beta}_0\right)$ the probability vectors

$$\widehat{\mathbf{p}} = \left(\frac{y_1}{N}, \frac{n_1 - y_1}{N}, \frac{y_2}{N}, \frac{n_2 - y_2}{N}, \ldots, \frac{y_I}{N}, \frac{n_I - y_I}{N}\right)^{\mathrm{T}},$$

$$\mathbf{p}\left(\boldsymbol{\beta}_0\right) \equiv \left(\pi\left(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0\right)\frac{n_1}{N}, \left(1-\pi\left(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0\right)\right)\frac{n_1}{N}, \ldots, \pi\left(\mathbf{x}_I^{\mathrm{T}}\boldsymbol{\beta}_0\right)\frac{n_I}{N}, \left(1-\pi\left(\mathbf{x}_I^{\mathrm{T}}\boldsymbol{\beta}_0\right)\right)\frac{n_I}{N}\right)^{\mathrm{T}}.$$

Under the assumptions that π has continuous second partial derivatives in a neighborhood of the true value of the parameter $\boldsymbol{\beta}_0$, and $\phi(t) \in \Phi^*$ is twice differentiable at $t > 0$. The minimum ϕ-divergence estimator of $\boldsymbol{\beta}_0$, for the GLM, given in (1), verifies

$$\widehat{\boldsymbol{\beta}}_\phi = \boldsymbol{\beta}_0 + \left(\mathbf{X}^{\mathrm{T}}\mathbf{W}_N\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{D}_{\left(\mathbf{C}_i^0\right)^{\mathrm{T}}_{i=1,\ldots,I}}\mathbf{D}^{-1/2}_{\mathbf{p}(\boldsymbol{\beta}_0)}\left(\widehat{\mathbf{p}} - \mathbf{p}\left(\boldsymbol{\beta}_0\right)\right) + \left\|\widehat{\mathbf{p}} - \mathbf{p}\left(\boldsymbol{\beta}_0\right)\right\|\boldsymbol{\alpha}_1\left(\widehat{\mathbf{p}}; \widehat{\mathbf{p}} - \mathbf{p}\left(\boldsymbol{\beta}_0\right)\right), \tag{5}$$

where

$$\mathbf{C}_i^0 = \left(\frac{n_i}{N}\right)^{1/2}\frac{\partial\pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0\right)}{\partial\eta_i}\begin{pmatrix}\pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0\right)^{-1/2}\\ -\left(1-\pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0\right)\right)^{-1/2}\end{pmatrix}, \quad i = 1, \ldots, I,$$

and the function $\boldsymbol{\alpha}_1 : \mathbb{R}^{2I} \to \mathbb{R}^{k+1}$ verifies $\boldsymbol{\alpha}_1\left(\mathbf{p}; \mathbf{p}-\mathbf{p}\left(\boldsymbol{\beta}_0\right)\right) \to \mathbf{0}$ as $\mathbf{p}\to\mathbf{p}\left(\boldsymbol{\beta}_0\right)$. By $\mathbf{D}_{\mathbf{a}}$, we are denoting the diagonal matrix with elements $\mathbf{a}$.

Pardo and Pardo, [12] established that

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_\phi - \boldsymbol{\beta}_0\right) \underset{N\to\infty}{\overset{L}{\to}} \mathcal{N}\left(\mathbf{0}, \left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)^{-1}\right), \tag{6}$$

where $\mathbf{W} = \lim_{N\to\infty} \mathbf{W}_N$.

In relation to the restricted minimum ϕ-divergence estimator, $\widehat{\boldsymbol{\beta}}_\phi^{H_0}$, we have the following expansion:

$$\widehat{\boldsymbol{\beta}}_\phi^{H_0} = \boldsymbol{\beta}_0 + \mathbf{H}_N\left(\boldsymbol{\beta}_0\right)\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}_N\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{D}_{\left(\mathbf{C}_i^0\right)^{\mathrm{T}}_{i=1,\ldots,I}}\mathbf{D}_{\mathbf{p}(\boldsymbol{\beta}_0)}^{-1/2}\left(\widehat{\mathbf{p}} - \mathbf{p}\left(\boldsymbol{\beta}_0\right)\right)$$
$$+ \left\|\widehat{\mathbf{p}} - \mathbf{p}\left(\boldsymbol{\beta}_0\right)\right\| \boldsymbol{\alpha}_2\left(\widehat{\mathbf{p}}; \widehat{\mathbf{p}} - \mathbf{p}\left(\boldsymbol{\beta}_0\right)\right), \tag{7}$$

where

$$\mathbf{H}_N\left(\boldsymbol{\beta}_0\right) = \mathbf{I} - \left(\mathbf{X}^{\mathrm{T}}\mathbf{W}_N\mathbf{X}\right)^{-1}\mathbf{K}\left(\mathbf{K}^{\mathrm{T}}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}_N\mathbf{X}\right)^{-1}\mathbf{K}\right)^{-1}\mathbf{K}^{\mathrm{T}}$$

and the function $\boldsymbol{\alpha}_2 : \mathbb{R}^{2I} \to \mathbb{R}^{k+1}$, verifies $\boldsymbol{\alpha}_2\left(\mathbf{p}; \mathbf{p} - \mathbf{p}\left(\boldsymbol{\beta}_0\right)\right) \to \mathbf{0}$ as $\mathbf{p} \to \mathbf{p}\left(\boldsymbol{\beta}_0\right)$.

For more details about obtaining Eqs. (5)–(7) see Pardo and Pardo, [12].

Based on (5) and (7) it is not difficult to establish that

$$\widehat{\boldsymbol{\beta}}_\phi^{H_0} - \widehat{\boldsymbol{\beta}}_\phi = -\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}_N\mathbf{X}\right)^{-1}\mathbf{K}\left(\mathbf{K}^{\mathrm{T}}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}_N\mathbf{X}\right)^{-1}\mathbf{K}\right)^{-1}\left(\mathbf{K}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_\phi - \mathbf{m}\right)$$
$$+ \left\|\widehat{\mathbf{p}} - \mathbf{p}\left(\boldsymbol{\beta}_0\right)\right\|\left(\boldsymbol{\alpha}_2\left(\widehat{\mathbf{p}}; \widehat{\mathbf{p}} - \mathbf{p}\left(\boldsymbol{\beta}_0\right)\right) - \boldsymbol{\alpha}_1\left(\widehat{\mathbf{p}}; \widehat{\mathbf{p}} - \mathbf{p}\left(\boldsymbol{\beta}_0\right)\right)\right). \tag{8}$$

To test the compatibility of the restricted and unrestricted minimum ϕ-divergence estimators $\widehat{\boldsymbol{\beta}}_\phi^{H_0}$ and $\widehat{\boldsymbol{\beta}}_\phi$, we can consider the family of ϕ-divergence statistics

$$T_N^{\phi_1,\phi_2} = \frac{2}{\phi_1''(1)}\sum_{i=1}^{I} n_i D_{\phi_1}\left(\boldsymbol{\pi}_i\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right), \boldsymbol{\pi}_i\left(\widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}\right)\right), \tag{9}$$

where $\boldsymbol{\pi}_i\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right)$ and $\boldsymbol{\pi}_i\left(\widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}\right)$ are obtained from (3) replacing $\boldsymbol{\beta}$ by $\widehat{\boldsymbol{\beta}}_{\phi_2}$ and $\widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}$, respectively.

In fact, we consider two ϕ-divergence measures, D_{ϕ_1} associated with the ϕ_1-divergence test statistic and D_{ϕ_2} associated with the minimum ϕ_2-divergence estimator. If the linear hypotheses, $\mathbf{K}^{\mathrm{T}}\boldsymbol{\beta} = \mathbf{m}$, are correct, $\left(H_0 : \mathbf{K}^{\mathrm{T}}\boldsymbol{\beta} = \mathbf{m}\right)$, the asymptotic distribution of $T_N^{\phi_1,\phi_2}$, given in (9), is chi-squared with r degrees of freedom (see Theorem 1).

If we choose in (9) $\phi_2(x) = x\log x - x + 1$ and $\phi_1(x) = \frac{1}{2}(x-1)^2$, we get the classical Pearson test statistic for testing

$$H_0 : \mathbf{K}^{\mathrm{T}}\boldsymbol{\beta} = \mathbf{m} \text{ versus } H_1 : \mathbf{K}^{\mathrm{T}}\boldsymbol{\beta} \neq \mathbf{m} \tag{10}$$

and for $\phi_1(x) = \phi_2(x) = x\log x - x + 1$, we get $T_N^{\phi_1,\phi_2} = LR + o_p(1)$, where LR is the classical likelihood ratio test. Now we use the test statistic (9) as well as the sample information to define alternative estimators of $\boldsymbol{\beta}_0$, to $\widehat{\boldsymbol{\beta}}_{\phi_2}$ and $\widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}$.

In the case where we do not have enough evidence about if $\boldsymbol{\beta}_0 \in \Theta_0$, we propose to consider the family of estimators based on the test statistic defined in (9),

$$\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{h} = \widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0} + \left(1 - h\left(T_N^{\phi_1,\phi_2}\right)\right)\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}\right).$$

The election of different functions h give some well known estimators. If we choose $h(x) = 0, \forall x$ we get the unrestricted minimum ϕ_2-divergence estimator, $\widehat{\boldsymbol{\beta}}^h_{\phi_1,\phi_2} \equiv \widehat{\boldsymbol{\beta}}_{\phi_2}$. For $h(x) = 1, \forall x$, we get the restricted minimum ϕ_2-divergence estimator, $\widehat{\boldsymbol{\beta}}^h_{\phi_1,\phi_2} \equiv \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2}$. For $h(x) = 1 - a, \forall x, a \in (0, 1)$, the Shrinkage minimum divergence estimator, $\widehat{\boldsymbol{\beta}}^h_{\phi_1,\phi_2} \equiv \widehat{\boldsymbol{\beta}}^{SRE}_{\phi_2}$. For $h(x) = I_{(0,\chi^2_{r,\alpha})}(x)$, the preliminary minimum divergence estimator, $\widehat{\boldsymbol{\beta}}^h_{\phi_1,\phi_2} \equiv \widehat{\boldsymbol{\beta}}^{PTE}_{\phi_1,\phi_2}$. For $h(x) = aI_{(0,\chi^2_{r,\alpha})}(x), a \in (0, 1)$, the Shrinkage preliminary minimum ϕ_2-divergence estimator, $\widehat{\boldsymbol{\beta}}^h_{\phi_1,\phi_2} \equiv \widehat{\boldsymbol{\beta}}^{SPT}_{\phi_1,\phi_2}$. For $h(x) = (r-2)x^{-1}, (r > 2)$, the James–Stein minimum divergence estimator, $\widehat{\boldsymbol{\beta}}^h_{\phi_1,\phi_2} \equiv \widehat{\boldsymbol{\beta}}^{S}_{\phi_1,\phi_2}$. For $h(x) = 1 - \left(1-(r-2)x^{-1}\right) I_{(r-2,\infty)}(x), (r > 2)$, the positive-part of Stein-Rule minimum divergence estimator, $\widehat{\boldsymbol{\beta}}^h_{\phi_1,\phi_2} \equiv \widehat{\boldsymbol{\beta}}^{S_+}_{\phi_1,\phi_2}$. For $h(x) = 1 - \left(1-(r-2)x^{-1}\right) I_{[\chi^2_{r,\alpha},\infty)}(x) \quad (r > 2)$ the modified preliminary minimum divergence estimator, $\widehat{\boldsymbol{\beta}}^h_{\phi_1,\phi_2} \equiv \widehat{\boldsymbol{\beta}}^{PTE_+}_{\phi_1,\phi_2}$.

If we restrict ourselves to the maximum likelihood estimator and the classical likelihood ratio test some of the above estimators were considered by Matin and Saleh ([9,8]) in the particular case of the logistic regression model, not in the context of generalized linear models, considered in this paper. Also the parameters are restricted to the equation $\mathbf{K}^{\mathrm{T}}\boldsymbol{\beta} = \boldsymbol{\beta}_0$ with $\mathbf{K}^{\mathrm{T}} = \mathbf{I}$ and $\boldsymbol{\beta}_0 = \left(\beta_0^0, \beta_1^0, \ldots, \beta_k^0\right)$ is a fixed value of $\boldsymbol{\beta}$ which is a particular case of our equation $\mathbf{K}^{\mathrm{T}}\boldsymbol{\beta} = \mathbf{m}$. In the following Section, we shall obtain the asymptotic bias of $\widehat{\boldsymbol{\beta}}^h_{\phi_1,\phi_2}$ as well as some properties of them.

3. Asymptotic bias of $\widehat{\boldsymbol{\beta}}^h_{\phi_1,\phi_2}$ under contiguous alternative hypotheses

First we are going to get the asymptotic distribution of $T_N^{\phi_1,\phi_2}$ if some or all the hypotheses are incorrect as well as other results that will be necessary to get the asymptotic bias as well as the asymptotic distributional quadratic risk.

Let $\boldsymbol{\beta}_N \in \Theta - \Theta_0$ be a given alternative and let $\boldsymbol{\beta}$ be the element in Θ_0 closest to $\boldsymbol{\beta}_N$ in the Euclidean distance sense. A possibility to introduce contiguous alternatives is to relax the condition $\mathbf{f}(\boldsymbol{\beta}) = \mathbf{K}^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{m} = \mathbf{0}$ defining Θ_0. Let $\boldsymbol{\delta} \in \mathbb{R}^r$ and consider the following sequence, $\boldsymbol{\beta}_N$, of parameters approaching Θ_0 according to

$$H_{1,N} : \mathbf{f}\left(\boldsymbol{\beta}_N\right) = N^{-1/2}\boldsymbol{\delta}.$$

It is clear, under $H_{1,N}$, that

$$\mathbf{S}_N \equiv \sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \boldsymbol{\beta}_N\right) \xrightarrow[N\to\infty]{L} \mathbf{S}, \tag{11}$$

where $\mathbf{S}$ is a normal random vector with mean vector $\boldsymbol{\mu}_{\mathbf{S}} = \mathbf{0}$ and variance–covariance matrix

$$\boldsymbol{\Sigma}_{\mathbf{S}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)^{-1} \quad \text{where } \mathbf{W} = \lim_{N\to\infty} \mathbf{W}_N.$$

The following theorem presents the asymptotic distribution of $T_N^{\phi_1,\phi_2}$, given in (9) under $H_{1,N}$.

Theorem 1. *Under $H_{1,N}$, the asymptotic distribution of $T_N^{\phi_1,\phi_2}$ is a noncentral chi-square with r degrees of freedom and noncentrality parameter*

$$\lambda = \boldsymbol{\delta}^{\mathrm{T}}\left(\mathbf{K}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\right)^{-1}\boldsymbol{\delta} \tag{12}$$

and under H_0 given in (10) is a chi-square with r degrees of freedom.

Proof. A second Taylor expansion gives

$$T_N^{\phi_1,\phi_2} = \mathbf{Y}^{\mathrm{T}}\mathbf{Y} + o_P(1)$$

where

$$\mathbf{Y} = \sqrt{N}\mathbf{D}_{\mathbf{p}(\boldsymbol{\beta}_0)}^{-1/2}\left(\mathbf{p}(\widehat{\boldsymbol{\beta}}_{\phi_2}) - \mathbf{p}(\widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0})\right).$$

But

$$\mathbf{p}(\boldsymbol{\beta}) - \mathbf{p}(\boldsymbol{\beta}_N) = \mathbf{F}_{2I\times(k+1)}(\boldsymbol{\beta})\left(\boldsymbol{\beta} - \boldsymbol{\beta}_N\right) + \left\|\boldsymbol{\beta} - \boldsymbol{\beta}_N\right\|\boldsymbol{\alpha}^*(\boldsymbol{\beta}; \boldsymbol{\beta} - \boldsymbol{\beta}_N)$$

where $\boldsymbol{\alpha}^* : \mathbb{R}^{k+1} \longrightarrow \mathbb{R}^{2I}$ verifies $\boldsymbol{\alpha}^*(\boldsymbol{\beta}; \boldsymbol{\beta} - \boldsymbol{\beta}_N) \longrightarrow \mathbf{0}$ as $\boldsymbol{\beta} \longrightarrow \boldsymbol{\beta}_N$ and

$$\mathbf{F}_{2I\times(k+1)}(\boldsymbol{\beta}) = \operatorname{diag}\begin{pmatrix} \frac{n_i}{N}\pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)^{-1/2}\left(1 - \pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)^{-1/2}\right) \\ -\frac{n_i}{N}\pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)^{-1/2}\left(1 - \pi\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)^{-1/2}\right)\end{pmatrix}_{i=1,\ldots,I}$$

Taking into account that

$$\mathbf{X}^{\mathrm{T}}\mathbf{F}_{2I\times(k+1)}\left(\boldsymbol{\beta}_0\right)^{\mathrm{T}}\mathbf{D}_{\mathbf{p}(\boldsymbol{\beta}_0)}^{-1/2}\mathbf{D}_{\mathbf{p}(\boldsymbol{\beta}_0)}^{-1/2}\mathbf{F}_{2I\times(k+1)}\left(\boldsymbol{\beta}_0\right)\mathbf{X} = \mathbf{X}^{\mathrm{T}}\mathbf{W}_N\mathbf{X}$$

we have

$$T_N^{\phi_1,\phi_2} = \sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}\right)^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{W}_N\mathbf{X}\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}\right) + o_p(1).$$

Based on (8) we get,

$$T_N^{\phi_1,\phi_2} = \sqrt{N}\left(\mathbf{K}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\phi_2} - \mathbf{m}\right)^{\mathrm{T}}\left(\mathbf{K}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\right)^{-1}\sqrt{N}\left(\mathbf{K}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\phi_2} - \mathbf{m}\right) + o_p(1).$$

A Taylor series expansion of $\mathbf{f}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right)$ around $\boldsymbol{\beta}_N$ yields

$$\mathbf{f}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right) = \mathbf{f}\left(\boldsymbol{\beta}_N\right) + \mathbf{K}^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \boldsymbol{\beta}_N\right) + \left\|\widehat{\boldsymbol{\beta}}_{\phi_2} - \boldsymbol{\beta}_N\right\|\boldsymbol{\alpha}_3\left(\widehat{\boldsymbol{\beta}}_{\phi_2}; \widehat{\boldsymbol{\beta}}_{\phi_2} - \boldsymbol{\beta}_N\right),$$

where $\boldsymbol{\alpha}_3$:$\mathbb{R}^{k+1} \to \mathbb{R}^r$ verifies $\boldsymbol{\alpha}_3\left(\boldsymbol{\beta}; \boldsymbol{\beta} - \boldsymbol{\beta}_N\right) \to \mathbf{0}$ as $\boldsymbol{\beta} \to \boldsymbol{\beta}_N$. We know that $\mathbf{f}\left(\boldsymbol{\beta}_N\right) = N^{-1/2}\boldsymbol{\delta}$, and therefore

$$\mathbf{f}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right) = N^{-1/2}\boldsymbol{\delta} + \mathbf{K}^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \boldsymbol{\beta}_N\right) + \left\|\widehat{\boldsymbol{\beta}}_{\phi_2} - \boldsymbol{\beta}_N\right\|\boldsymbol{\alpha}_3\left(\widehat{\boldsymbol{\beta}}_{\phi_2}; \widehat{\boldsymbol{\beta}}_{\phi_2} - \boldsymbol{\beta}_N\right).$$

As $N^{1/2}\left\|\widehat{\boldsymbol{\beta}}_{\phi_2} - \boldsymbol{\beta}_N\right\|$ is bounded in probability and by (11), we have

$$\sqrt{N}\mathbf{f}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right) \xrightarrow[N\to\infty]{L} \mathcal{N}\left(\boldsymbol{\delta}, \mathbf{K}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\right).$$

Now the result follows from lemma in page 63 of Ferguson [5].

Under H_0, $\boldsymbol{\delta} = \mathbf{0}$, therefore, the asymptotic distribution of $T_N^{\phi_1,\phi_2}$ is a chi-square with r degrees of freedom. ■

Proposition 2. *Under $H_{1,N}$ the random vectors,*

$$\mathbf{Y}_N = \sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}\right) \quad \textit{and} \quad \mathbf{Z}_N = \sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0} - \boldsymbol{\beta}_N\right)$$

are asymptotically independent and

$$\mathbf{Y}_N \xrightarrow[N\to\infty]{L} \mathbf{Y}, \qquad \mathbf{Z}_N \xrightarrow[N\to\infty]{L} \mathbf{Z},$$

where $\mathbf{Y}_N$ *and* $\mathbf{Z}_N$ *are normal random vectors with mean vectors* $\boldsymbol{\mu}_{\mathbf{Y}} = \mathbf{L}\boldsymbol{\delta}$ *and* $\boldsymbol{\mu}_{\mathbf{Z}} = -\mathbf{L}\boldsymbol{\delta}$, *respectively and variance–covariance matrices*

$$\boldsymbol{\Sigma}_{\mathbf{Y}} = \mathbf{L}\mathbf{K}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}} \quad \textit{and} \quad \boldsymbol{\Sigma}_{\mathbf{Z}} = \boldsymbol{\Sigma}_{\mathbf{S}} - \mathbf{L}\mathbf{K}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}},$$

respectively, where

$$\mathbf{L} = \boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\left(\mathbf{K}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\right)^{-1}.$$

Proof. The result for $\mathbf{Y}_N$ is obtained from (8) and (6) and the result for $\mathbf{Z}_N$ is obtained from (7) and (6). The random vectors $\mathbf{Y}$ and $\mathbf{Z}$ are asymptotically independent, because it is not difficult to establish that

$$\lim_{N\to\infty} \mathrm{Cov}\left(\mathbf{Y}_N, \mathbf{Z}_N\right) = \mathbf{0}. \quad \blacksquare$$

Remark 3. We have that

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}\right) - \mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}\right) = \boldsymbol{\Sigma}_{\mathbf{Y}}.$$

Therefore the difference of the variance–covariance matrices is a positive semi-definite matrix, and hence it can be concluded that the restricted minimum ϕ_2-divergence estimator has a smaller sampling variance than the unrestricted minimum ϕ_2-divergence estimator.

Let $\widetilde{\boldsymbol{\beta}}^*$ be a suitable estimator of $\boldsymbol{\beta}$, and we denote by $F_{\widetilde{\boldsymbol{\beta}}^*}$ the asymptotic distribution of $\sqrt{N}\left(\widetilde{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_N\right)$. The asymptotic bias of $\widetilde{\boldsymbol{\beta}}^*$ is defined by

$$B(\widetilde{\boldsymbol{\beta}}^*) = \int \mathbf{x}\,\mathrm{d}F_{\widetilde{\boldsymbol{\beta}}^*}\left(\mathbf{x}\right).$$

In the following theorem, we present the expression of $B(\widehat{\boldsymbol{\beta}}^h_{\phi_1,\phi_2})$.

Theorem 4. *The asymptotic bias of*

$$\widehat{\boldsymbol{\beta}}^h_{\phi_1,\phi_2} = \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2} + \left(1 - h\left(T_N^{\phi_1,\phi_2}\right)\right)\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2}\right)$$

under $H_{1,N}$ *is given by*

$$B(\widehat{\boldsymbol{\beta}}^h_{\phi_1,\phi_2}) = -\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\left(\mathbf{K}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\right)^{-1}\boldsymbol{\delta}E\left[h\left(\chi^2_{r+2}\left(\lambda\right)\right)\right].$$

Proof. We have,

$$\begin{aligned}\sqrt{N}\left(\widehat{\boldsymbol{\beta}}^h_{\phi_1,\phi_2} - \boldsymbol{\beta}_N\right) &= \sqrt{N}\left(\widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2} - \boldsymbol{\beta}_N\right) + \left(1 - h\left(T_N^{\phi_1,\phi_2}\right)\right)\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2}\right)\\ &= \sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \boldsymbol{\beta}_N\right) - \sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2}\right)h\left(T_N^{\phi_1,\phi_2}\right).\end{aligned}$$

Under $H_{1,N}$, $\lim_{N\to\infty} E\left[\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}-\boldsymbol{\beta}_N\right)\right]=\mathbf{0}$ and $\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}-\widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}\right)$ can be written as

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2}-\widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}\right)=\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}_N\mathbf{X}\right)^{-1}\mathbf{K}\left(\mathbf{K}^{\mathrm{T}}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}_N\mathbf{X}\right)^{-1}\mathbf{K}\right)^{-1}\sqrt{N}\left(\mathbf{K}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\phi_2}-\mathbf{m}\right)$$
$$+\sqrt{N}\left\|\widehat{\mathbf{p}}-\mathbf{p}\left(\boldsymbol{\beta}_0\right)\right\|\left(\boldsymbol{\alpha}_2\left(\widehat{\mathbf{p}};\widehat{\mathbf{p}}-\mathbf{p}\left(\boldsymbol{\beta}_0\right)\right)-\boldsymbol{\alpha}_1\left(\widehat{\mathbf{p}};\widehat{\mathbf{p}}-\mathbf{p}\left(\boldsymbol{\beta}_0\right)\right)\right).$$

Therefore,

$$B(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{h})=-\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\left(\mathbf{K}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\right)^{-1/2}E\left[\boldsymbol{\xi}h\left(\boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{\xi}\right)\right],$$

where $\boldsymbol{\xi}$ is a normal random vector with mean vector

$$\boldsymbol{\mu}_{\boldsymbol{\xi}}=\left(\mathbf{K}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\right)^{-1/2}\boldsymbol{\delta}$$

and variance–covariance matrix the identity $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}=\mathbf{I}$. Applying Theorem 6 in Saleh [15], we have

$$B(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{h})=-\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\left(\mathbf{K}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\right)^{-1}\boldsymbol{\delta}E\left[h\left(\chi_{r+2}^{2}(\lambda)\right)\right]$$

where $\lambda=\boldsymbol{\delta}^{\mathrm{T}}\left(\mathbf{K}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\right)^{-1}\boldsymbol{\delta}$. ■

Remark 5. From the above theorem, we can get the asymptotic bias for the different estimators considered in Section 2 using the corresponding expressions of h. The estimator $\widehat{\boldsymbol{\beta}}_{\phi_2}$ is asymptotically unbiased and $\widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}$, $\widehat{\boldsymbol{\beta}}_{\phi_2}^{SRE}$, $\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{PTE}$, $\widehat{\boldsymbol{\beta}}_{\phi_2}^{SPT}$, $\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{S}$, $\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{S_+}$, $\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{PTE_+}$ are biased. Under the null hypothesis $H_0:\mathbf{K}^{\mathrm{T}}\boldsymbol{\beta}=\mathbf{m}$, the bias of $\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{h}$, independently of h, is zero.

The previous results are not in a scalar form, and in order to be able to do comparisons we can consider the asymptotic quadratic bias of them

$$B^*(\widehat{\boldsymbol{\beta}}^*)=B(\widehat{\boldsymbol{\beta}}^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}}^{-1}B(\widehat{\boldsymbol{\beta}}^*).$$

This is given by

$$B^*(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{h})=\left(E\left[h\left(\chi_{r+2}^{2}(\lambda)\right)\right]\right)^2\boldsymbol{\delta}^{\mathrm{T}}\left(\mathbf{K}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\right)^{-1}\boldsymbol{\delta}=\left(E\left[h\left(\chi_{r+2}^{2}(\lambda)\right)\right]\right)^2\lambda, \quad (13)$$

where λ was defined in (12).

From (13), the asymptotic quadratic bias for the different estimators considered in this paper is given by $B^*(\widehat{\boldsymbol{\beta}}_{\phi_2})=0$, $B^*(\widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0})=\lambda$, $B^*(\widehat{\boldsymbol{\beta}}_{\phi_2}^{SRE})=(1-a)^2\lambda$, $B^*(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{PTE})=\lambda G_{r+2}^{2}\left(\chi_{r,\alpha}^{2};\lambda\right)$,
$B^*(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{SPT})=\lambda a^2 G_{r+2}^{2}\left(\chi_{r,\alpha}^{2};\lambda\right)$, $B^*(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{S})=(r-2)^2\lambda\left(E\left[\chi_{r+2}^{-2}(\lambda)\right]\right)^2$,
$B^*(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{S_+})=\lambda\left\{G_{r+2}(r-2;\lambda)+(r-2)E\left[\chi_{r+2}^{-2}(\lambda)\right]\right.$
$\left.-(r-2)E\left[\chi_{r+2}^{-2}(\lambda)I_{(0,r-2)}\left(\chi_{r+2}^{2}(\lambda)\right)\right]\right\}^2$ and $B^*(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{PTE_+})=\lambda\left\{G_{r+2}\left(\chi_{r,\alpha}^{2};\lambda\right)\right.$
$\left.+(r-2)E\left[\chi_{r+2}^{-2}(\lambda)\right]-(r-2)E\left[\chi_{r+2}^{-2}(\lambda)I_{(0,\chi_{r,\alpha}^{2})}\left(\chi_{r+2}^{2}(\lambda)\right)\right]\right\}^2$.

By $G_r(x;\lambda)$, we are denoting the distribution function of a non-central chi-square with r degrees of freedom and noncentrality parameter λ evaluated at x.

Theorem 6. *The asymptotic quadratic bias of the estimators* $\widehat{\boldsymbol{\beta}}_{\phi_2}$, $\widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}$, $\widehat{\boldsymbol{\beta}}_{\phi_2}^{SRE}$, $\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{PTE}$, $\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{SPT}$, $\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{S}$, $\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{S_+}$ *and* $\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{PTE_+}$ *can be ordered as follows:*

(a) $B^*(\widehat{\boldsymbol{\beta}}_{\phi_2}) \leq B^*(\widehat{\boldsymbol{\beta}}^{SRE}_{\phi_2}) \leq B^*(\widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2})$

(b) $B^*(\widehat{\boldsymbol{\beta}}_{\phi_2}) \leq B^*(\widehat{\boldsymbol{\beta}}^{SPT}_{\phi_1,\phi_2}) \leq B^*(\widehat{\boldsymbol{\beta}}^{PTE}_{\phi_1,\phi_2})$

(c) $B^*(\widehat{\boldsymbol{\beta}}^{S_+}_{\phi_1,\phi_2}) \leq B^*(\widehat{\boldsymbol{\beta}}^{S}_{\phi_1,\phi_2})$ *and* $B^*(\widehat{\boldsymbol{\beta}}^{PTE_+}_{\phi_1,\phi_2}) \leq B^*(\widehat{\boldsymbol{\beta}}^{PTE}_{\phi_1,\phi_2})$

(d) $B^*(\widehat{\boldsymbol{\beta}}^{S}_{\phi_1,\phi_2}) \leq B^*(\widehat{\boldsymbol{\beta}}^{PTE}_{\phi_1,\phi_2})$ *iff* $G_{r+2}\left(\chi^2_{r,\alpha};\lambda\right) \geq (r-2)\, E\left[\chi^{-2}_{r+2}(\lambda)\right]$ *for all* α *and* λ.

(e) $B^*(\widehat{\boldsymbol{\beta}}^{S_+}_{\phi_1,\phi_2}) \leq B^*(\widehat{\boldsymbol{\beta}}^{PTE}_{\phi_1,\phi_2})$ *iff* $G_{r+2}\left(\chi^2_{r,\alpha};\lambda\right) \geq E\Big[1-\Big(1-(r-2)\,\chi^{-2}_{r+2}(\lambda)\Big)$
$\times I_{(r-2,\infty)}\left(\chi^2_{r+2}(\lambda)\right)\Big]$ *for all* α *and* λ.

Proof. Parts (a), (b), (d) and (e) are immediate.

(c) It is clear that $l = B^*(\widehat{\boldsymbol{\beta}}^{S}_{\phi_1,\phi_2}) - B^*(\widehat{\boldsymbol{\beta}}^{S_+}_{\phi_1,\phi_2})$, can be written as

$$l = \lambda\,(r-2)^2 \left\{E\left[\chi^{-2}_{r+2}(\lambda)\, I_{(0,r-2)}\left(\chi^2_{r+2}(\lambda)\right)\right] - \frac{1}{(r-2)} G^2_{r+2}(r-2;\lambda)\right\}$$
$$\left\{2E\left[\chi^{-2}_{r+2}(\lambda)\right] + \frac{1}{(r-2)} G_{r+2}(r-2;\lambda) - E\left[\chi^{-2}_{r+2}(\lambda)\, I_{(0,r-2)}\left(\chi^2_{r+2}(\lambda)\right)\right]\right\}.$$

Now applying (2.2.1a), (2.2.13a), and (2.213g) in Saleh [15], we have $l \geq 0$. On the other hand, we have

$$E\left[\chi^{-2}_{r+2}(\lambda)\right] - E\left[\chi^{-2}_{r+2}(\lambda)\, I_{(0,\chi^2_{r,\alpha})}\left(\chi^2_{r+2}(\lambda)\right)\right] \geq 0.$$

Therefore

$$\begin{aligned} B^*\left(\widehat{\boldsymbol{\beta}}^{PTE_+}_{\phi_1,\phi_2}\right) &= \lambda\left\{G_{r+2}\left(\chi^2_{r,\alpha};\lambda\right) + (r-2)E\left[\chi^{-2}_{r+2}(\lambda)\right]\right. \\ &\quad \left. -(r-2)E\left[\chi^{-2}_{r+2}(\lambda)\, I_{(0,r-2)}\left(\chi^2_{r+2}(\lambda)\right)\right]\right\}^2 \\ &\geq \lambda G_{r+2}\left(\chi^2_{r,\alpha};\lambda\right)^2 = B^*\left(\widehat{\boldsymbol{\beta}}^{PTE}_{\phi_1,\phi_2}\right). \quad \blacksquare \end{aligned}$$

4. Asymptotic quadratic risk of $\widehat{\boldsymbol{\beta}}^{h}_{\phi_1,\phi_2}$ under contiguous alternative hypotheses: Performance

Let $\widetilde{\boldsymbol{\beta}}^*$ be a suitable estimator of $\boldsymbol{\beta}$ and $F_{\widetilde{\boldsymbol{\beta}}^*}$ the asymptotic distribution of $\sqrt{N}\left(\widetilde{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_N\right)$. Given a positive semidefinite matrix $\mathbf{M}$, we define the asymptotic distributional quadratic risk (ADQR) of $\widetilde{\boldsymbol{\beta}}^*$ by

$$R\left(\widetilde{\boldsymbol{\beta}}^*;\mathbf{M}\right) = \int \mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{x}\,\mathrm{d}F_{\widetilde{\boldsymbol{\beta}}^*}(\mathbf{x}).$$

Theorem 7. *The asymptotic distributional quadratic risk of*

$$\widehat{\boldsymbol{\beta}}^{h}_{\phi_1,\phi_2} = \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2} + \left(1 - h\left(T_N^{\phi_1,\phi_2}\right)\right)\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2}\right)$$

is given by

$$R\left(\widehat{\boldsymbol{\beta}}^{h}_{\phi_1,\phi_2};\mathbf{M}\right)$$
$$= \operatorname{trace}(\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{S}}) - \operatorname{trace}(\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{M})\left\{2E\left[h\left(\chi^2_{r+2}(\lambda)\right)\right] - E\left[h\left(\chi^2_{r+2}(\lambda)\right)^2\right]\right\}$$
$$+\boldsymbol{\delta}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{M}\mathbf{L}\boldsymbol{\delta}\left\{E\left[h\left(\chi^2_{r+4}(\lambda)\right)^2\right] - 2E\left[h\left(\chi^2_{r+4}(\lambda)\right)\right] + 2E\left[h\left(\chi^2_{r+2}(\lambda)\right)\right]\right\}.$$

Proof. We have

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}^{h}_{\phi_1,\phi_2} - \boldsymbol{\beta}_N\right) = \sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \boldsymbol{\beta}_N\right) - \sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2}\right) h\left(T_N^{\phi_1,\phi_2}\right).$$

Therefore,

$$\begin{aligned}
R\left(\widehat{\boldsymbol{\beta}}^{h}_{\phi_1,\phi_2};\mathbf{M}\right) &= E\left[\lim_{N\to\infty}\left(\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \boldsymbol{\beta}_N\right)^{\mathrm{T}} - \sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2}\right)^{\mathrm{T}} h\left(T_N^{\phi_1,\phi_2}\right)\right)\mathbf{M}\right.\\
&\qquad\left.\left(\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \boldsymbol{\beta}_N\right) - \sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2}\right) h\left(T_N^{\phi_1,\phi_2}\right)\right)\right]\\
&= E\left[\lim_{N\to\infty}\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \boldsymbol{\beta}_N\right)^{\mathrm{T}}\mathbf{M}\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \boldsymbol{\beta}_N\right)\right]\\
&\quad - E\left[\lim_{N\to\infty}\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \boldsymbol{\beta}_N\right)^{\mathrm{T}}\mathbf{M}\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2}\right) h\left(T_N^{\phi_1,\phi_2}\right)\right]\\
&\quad - E\left[\lim_{N\to\infty} h\left(T_N^{\phi_1,\phi_2}\right)\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2}\right)^{\mathrm{T}}\mathbf{M}\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \boldsymbol{\beta}_N\right)\right]\\
&\quad + E\left[\lim_{N\to\infty} h\left(T_N^{\phi_1,\phi_2}\right)^2\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2}\right)^{\mathrm{T}}\mathbf{M}\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2}\right)\right]\\
&= A - B - C + D.
\end{aligned}$$

Now we are going to get the expressions of A, B, C and D. By (11), we have

$$A = E\left[\mathbf{S}^{\mathrm{T}}\mathbf{M}\mathbf{S}\right] = \operatorname{trace}(\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{S}}).$$

We know

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{\phi_2} - \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2}\right) = \sqrt{N}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}_N\mathbf{X}\right)^{-1}\mathbf{K}\left(\mathbf{K}^{\mathrm{T}}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}_N\mathbf{X}\right)^{-1}\mathbf{K}\right)^{-1/2}\boldsymbol{\xi}_N,$$

where $\sqrt{N}\boldsymbol{\xi}_N = \left(\mathbf{K}^{\mathrm{T}}\left(\mathbf{X}^{\mathrm{T}}\mathbf{W}_N\mathbf{X}\right)^{-1}\mathbf{K}\right)^{-1/2}\sqrt{N}\left(\mathbf{K}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\phi_2} - \mathbf{m}\right)$ and

$$\sqrt{N}\boldsymbol{\xi}_N \xrightarrow[N\to\infty]{L} \boldsymbol{\xi},$$

being $\boldsymbol{\xi}$ a random normal vector with mean $\boldsymbol{\mu}_{\boldsymbol{\xi}} = \left(\mathbf{K}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\right)^{-1/2}\boldsymbol{\delta}$ and variance–covariance matrix the identity $\boldsymbol{\Sigma}_{\boldsymbol{\xi}} = \mathbf{I}$.

By Theorem 8 in Saleh [15], we have

$$\begin{aligned}
D &= E\left[h\left(\boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{\xi}\right)^2\boldsymbol{\xi}^{\mathrm{T}}\left(\mathbf{K}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\right)^{-1/2}\mathbf{K}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\left(\mathbf{K}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{K}\right)^{-1/2}\boldsymbol{\xi}\right]\\
&= \operatorname{trace}(\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{M})E\left[h\left(\chi^2_{r+2}(\lambda)\right)^2\right] + E\left[h\left(\chi^2_{r+4}(\lambda)\right)^2\right]\boldsymbol{\delta}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{M}\mathbf{L}\boldsymbol{\delta}.
\end{aligned}$$

In relation with B, we have the following:

$$\begin{aligned} B &= E\left[\lim_{N\to\infty}\sqrt{N}\left(\widehat{\boldsymbol\beta}_{\phi_2}-\boldsymbol\beta_N\right)^{\mathrm T}\mathbf M\sqrt{N}\left(\widehat{\boldsymbol\beta}_{\phi_2}-\widehat{\boldsymbol\beta}^{H_0}_{\phi_2}\right)h\left(T_N^{\phi_1,\phi_2}\right)\right]\\ &= E\left[\lim_{N\to\infty}\mathbf S_N^{\mathrm T}\mathbf M h\left(T_N^{\phi_1,\phi_2}\right)\left(\mathbf X^{\mathrm T}\mathbf W_N\mathbf X\right)^{-1}\mathbf K\left(\mathbf K^{\mathrm T}\left(\mathbf X^{\mathrm T}\mathbf W_N\mathbf X\right)^{-1}\mathbf K\right)^{-1/2}\sqrt{N}\boldsymbol\xi_N\right]\\ &= E\left[\mathbf S^{\mathrm T}\mathbf M h\left(\boldsymbol\xi^{\mathrm T}\boldsymbol\xi\right)\boldsymbol\Sigma_{\mathbf S}\mathbf K\left(\mathbf K^{\mathrm T}\boldsymbol\Sigma_{\mathbf S}\mathbf K\right)^{-1/2}\boldsymbol\xi\right]\\ &= E\left[h\left(\boldsymbol\xi^{\mathrm T}\boldsymbol\xi\right)E\left[\mathbf S^{\mathrm T}/\boldsymbol\xi\right]\mathbf M\boldsymbol\Sigma_{\mathbf S}\mathbf K\left(\mathbf K^{\mathrm T}\boldsymbol\Sigma_{\mathbf S}\mathbf K\right)^{-1/2}\boldsymbol\xi\right]. \end{aligned}$$

It is not difficult to establish that

$$E\left[\mathbf S/\boldsymbol\xi=\mathbf x\right]=\boldsymbol\Sigma_{\mathbf S}\mathbf K\left(\mathbf K^{\mathrm T}\boldsymbol\Sigma_{\mathbf S}\mathbf K\right)^{-1/2}\mathbf x-\mathbf L\boldsymbol\delta.$$

Therefore we have,

$$\begin{aligned} C &= E\left[h\left(\boldsymbol\xi^{\mathrm T}\boldsymbol\xi\right)\boldsymbol\xi^{\mathrm T}\left(\mathbf K^{\mathrm T}\boldsymbol\Sigma_{\mathbf S}\mathbf K\right)^{-1/2}\mathbf K^{\mathrm T}\boldsymbol\Sigma_{\mathbf S}\mathbf M\boldsymbol\Sigma_{\mathbf S}\mathbf K\left(\mathbf K^{\mathrm T}\boldsymbol\Sigma_{\mathbf S}\mathbf K\right)^{-1/2}\boldsymbol\xi\right]\\ &\quad - E\left[h\left(\boldsymbol\xi^{\mathrm T}\boldsymbol\xi\right)\boldsymbol\delta^{\mathrm T}\mathbf L^{\mathrm T}\mathbf M\boldsymbol\Sigma_{\mathbf S}\mathbf K\left(\mathbf K^{\mathrm T}\boldsymbol\Sigma_{\mathbf S}\mathbf K\right)^{-1/2}\boldsymbol\xi\right]\\ &= \operatorname{trace}(\boldsymbol\Sigma_{\mathbf Y}\mathbf M)E\left[h\left(\chi^2_{r+2}(\lambda)\right)\right]+\boldsymbol\delta^{\mathrm T}\mathbf L^{\mathrm T}\mathbf M\mathbf L\boldsymbol\delta E\left[h\left(\chi^2_{r+4}(\lambda)\right)\right]\\ &\quad -\boldsymbol\delta^{\mathrm T}\mathbf L^{\mathrm T}\mathbf M\mathbf L\boldsymbol\delta E\left[h\left(\chi^2_{r+2}(\lambda)\right)\right]\\ &= \operatorname{trace}(\boldsymbol\Sigma_{\mathbf Y}\mathbf M)E\left[h\left(\chi^2_{r+2}(\lambda)\right)\right]\\ &\quad +\boldsymbol\delta^{\mathrm T}\mathbf L^{\mathrm T}\mathbf M\mathbf L\boldsymbol\delta\left\{E\left[h\left(\chi^2_{r+4}(\lambda)\right)\right]-E\left[h\left(\chi^2_{r+2}(\lambda)\right)\right]\right\}. \end{aligned}$$

Now the result follows. ■

First we are going to see the relation between the asymptotic quadratic risk of $\widehat{\boldsymbol\beta}_{\phi_2}$, $\widehat{\boldsymbol\beta}^{H_0}_{\phi_2}$, $\widehat{\boldsymbol\beta}^{SRE}_{\phi_2}$, $\widehat{\boldsymbol\beta}^{PTE}_{\phi_1,\phi_2}$, $\widehat{\boldsymbol\beta}^{SPT}_{\phi_2}$, $\widehat{\boldsymbol\beta}^{S}_{\phi_1,\phi_2}$, $\widehat{\boldsymbol\beta}^{S_+}_{\phi_1,\phi_2}$ and $\widehat{\boldsymbol\beta}^{PTE_+}_{\phi_1,\phi_2}$ under the null hypothesis $H_0:\mathbf K^{\mathrm T}\boldsymbol\beta=\mathbf m$.

Theorem 8. *Under the null hypothesis $H_0:\mathbf K^{\mathrm T}\boldsymbol\beta=\mathbf m$, the asymptotic distributional quadratic risk of the estimators $\widehat{\boldsymbol\beta}_{\phi_2}$, $\widehat{\boldsymbol\beta}^{H_0}_{\phi_2}$, $\widehat{\boldsymbol\beta}^{SRE}_{\phi_2}$, $\widehat{\boldsymbol\beta}^{PTE}_{\phi_1,\phi_2}$, $\widehat{\boldsymbol\beta}^{SPT}_{\phi_1,\phi_2}$, $\widehat{\boldsymbol\beta}^{S}_{\phi_1,\phi_2}$, $\widehat{\boldsymbol\beta}^{S_+}_{\phi_1,\phi_2}$ and $\widehat{\boldsymbol\beta}^{PTE_+}_{\phi_1,\phi_2}$ can be ordered as follows:*

(a) $R\left(\widehat{\boldsymbol\beta}^{H_0}_{\phi_2};\mathbf M\right)\le R\left(\widehat{\boldsymbol\beta}^{PTE}_{\phi_1,\phi_2};\mathbf M\right)\le R\left(\widehat{\boldsymbol\beta}^{SPT}_{\phi_1,\phi_2};\mathbf M\right)\le R\left(\widehat{\boldsymbol\beta}_{\phi_2};\mathbf M\right)$

(b) *If* $r>2$,

$$R\left(\widehat{\boldsymbol\beta}^{H_0}_{\phi_2};\mathbf M\right)\le R\left(\widehat{\boldsymbol\beta}^{S_+}_{\phi_1,\phi_2};\mathbf M\right)\le R\left(\widehat{\boldsymbol\beta}^{S}_{\phi_1,\phi_2};\mathbf M\right)\le R\left(\widehat{\boldsymbol\beta}_{\phi_2};\mathbf M\right)$$

and

$$R\left(\widehat{\boldsymbol\beta}^{H_0}_{\phi_2};\mathbf M\right)\le R\left(\widehat{\boldsymbol\beta}^{PTE_+}_{\phi_1,\phi_2};\mathbf M\right)\le R\left(\widehat{\boldsymbol\beta}^{PTE}_{\phi_1,\phi_2};\mathbf M\right)\le R\left(\widehat{\boldsymbol\beta}_{\phi_2};\mathbf M\right).$$

(c) $R\left(\widehat{\boldsymbol\beta}^{H_0}_{\phi_2};\mathbf M\right)<R\left(\widehat{\boldsymbol\beta}^{SRE}_{\phi_2};\mathbf M\right)<R\left(\widehat{\boldsymbol\beta}_{\phi_2};\mathbf M\right)$

(d) *If* $G_{r+2}(\chi^2_{r,\alpha};0) \geq (r-2)/2$,

$$R\left(\widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2};\mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}^{PTE}_{\phi_1,\phi_2};\mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}^{S}_{\phi_1,\phi_2};\mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_{\phi_2};\mathbf{M}\right).$$

Proof. Under the null hypothesis $H_0 : \mathbf{K}^{\mathrm{T}}\boldsymbol{\beta} = \mathbf{m}$ by Theorem 7, we have

$$R\left(\widehat{\boldsymbol{\beta}}^{h}_{\phi_1,\phi_2};\mathbf{M}\right) = \operatorname{trace}(\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{Z}}) + \operatorname{trace}(\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{M})E\left[\left(1-h\left(\chi^2_{r+2}(0)\right)\right)^2\right].$$

Therefore the ADQR is an increasing function of

$$E\left[\left(1-h\left(\chi^2_{r+2}(0)\right)\right)^2\right]. \tag{14}$$

For this reason, we are going to analyze this expression for the different functions h associated with $\widehat{\boldsymbol{\beta}}_{\phi_2}, \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2}, \widehat{\boldsymbol{\beta}}^{SRE}_{\phi_2}, \widehat{\boldsymbol{\beta}}^{PTE}_{\phi_1,\phi_2}, \widehat{\boldsymbol{\beta}}^{SPT}_{\phi_1,\phi_2}, \widehat{\boldsymbol{\beta}}^{S}_{\phi_1,\phi_2}, \widehat{\boldsymbol{\beta}}^{S_+}_{\phi_1,\phi_2}$ and $\widehat{\boldsymbol{\beta}}^{PTE_+}_{\phi_1,\phi_2}$, respectively. It is not difficult to establish that $E\left[\left(1-h\left(\chi^2_{r+2}(0)\right)\right)^2\right]$ is equal $1, 0, a^2$, $1-G_{r+2}(\chi^2_{r,\alpha};0)$, $1-a(2-a)G_{r+2}(\chi^2_{r,\alpha};0)$, $\int_0^\infty(1-(r-2)x^{-1})^2\mathrm{d}G_{r+2}(x;0)$, $\int_{r-2}^\infty(1-(r-2)x^{-1})^2\mathrm{d}G_{r+2}(x;0)$ and $\int_{\chi^2_{r,\alpha}}^\infty(1-(r-2)x^{-1})^2\mathrm{d}G_{r+2}(x;0)$ for the functions h associated with $\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2}, \widehat{\boldsymbol{\beta}}^{SRE}_{\phi_2}, \widehat{\boldsymbol{\beta}}^{PTE}_{\phi_1,\phi_2}, \widehat{\boldsymbol{\beta}}^{SPT}_{\phi_1,\phi_2}, \widehat{\boldsymbol{\beta}}^{S}_{\phi_1,\phi_2}, \widehat{\boldsymbol{\beta}}^{S_+}_{\phi_1,\phi_2}$ and $\widehat{\boldsymbol{\beta}}^{PTE_+}_{\phi_1,\phi_2}$, respectively. Now it is clear that

$$R\left(\widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2};\mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}^{PTE}_{\phi_1,\phi_2};\mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}^{SPT}_{\phi_1,\phi_2};\mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_{\phi_2};\mathbf{M}\right).$$

We consider $r > 2$. The expression (14) for $\widehat{\boldsymbol{\beta}}^{S}_{\phi_1,\phi_2}$ is $1-\frac{r-2}{r}$, and for $\widehat{\boldsymbol{\beta}}_{\phi_2}$ it is 1. Therefore we have,

$$R\left(\widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2};\mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}^{S_+}_{\phi_1,\phi_2};\mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}^{S}_{\phi_1,\phi_2};\mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_{\phi_2};\mathbf{M}\right).$$

The inequality between $R\left(\widehat{\boldsymbol{\beta}}^{S_+}_{\phi_1,\phi_2};\mathbf{M}\right)$ and $R\left(\widehat{\boldsymbol{\beta}}^{S}_{\phi_1,\phi_2};\mathbf{M}\right)$ follows, because

$$\int_{r-2}^\infty(1-(r-2)x^{-1})^2\mathrm{d}G_{r+2}(x;0) \leq \int_0^\infty(1-(r-2)x^{-1})^2\mathrm{d}G_{r+2}(x;0).$$

We have,

$$\begin{aligned}\int_{\chi^2_{r,\alpha}}^\infty(1-(r-2)x^{-1})^2\mathrm{d}G_{r+2}(x;0) &= 1-G_{r+2}(\chi^2_{r,\alpha};0)+\frac{r-2}{r}\left\{-\left(1-G_r(\chi^2_{r,\alpha};0)\right)\right.\\ &\quad \left.-\left(G_{r-2}(\chi^2_{r,\alpha};0)-G_r(\chi^2_{r,\alpha};0)\right)\right\}.\end{aligned}$$

and

$$\begin{aligned}&R\left(\widehat{\boldsymbol{\beta}}^{PTE_+}_{\phi_1,\phi_2};\mathbf{M}\right) - R\left(\widehat{\boldsymbol{\beta}}^{PTE}_{\phi_1,\phi_2};\mathbf{M}\right)\\ &\quad = \frac{r-2}{r}\left\{-\left(1-G_r(\chi^2_{r,\alpha};0)\right)-\left(G_{r-2}(\chi^2_{r,\alpha};0)-G_r(\chi^2_{r,\alpha};0)\right)\right\}.\end{aligned}$$

Therefore, $R\left(\widehat{\boldsymbol{\beta}}^{PTE_+}_{\phi_1,\phi_2};\mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}^{PTE}_{\phi_1,\phi_2};\mathbf{M}\right)$, because $G_{r-2}(\chi^2_{r,\alpha};0) > G_r(\chi^2_{r,\alpha};0)$.

Now it is clear that

$$R\left(\widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2};\mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}^{PTE_+}_{\phi_1,\phi_2};\mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}^{PTE}_{\phi_1,\phi_2};\mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_{\phi_2};\mathbf{M}\right).$$

It is also immediate to see that $R\left(\widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0};\mathbf{M}\right) < R\left(\widehat{\boldsymbol{\beta}}_{\phi_2}^{SRE};\mathbf{M}\right) < R\left(\widehat{\boldsymbol{\beta}}_{\phi_2};\mathbf{M}\right)$.

The result in (d) follows because

$$\int_0^\infty (1-(r-2)x^{-1})^2 \mathrm{d}G_{r+2}(x;0) = 1 - \frac{r-2}{r}.$$

This last result has been obtained taking into account that

$$E\left[\chi_{r+2}^2(0)^k\right] = 2^k \Gamma\left(\frac{r+2}{2}+k\right)\Gamma\left(\frac{r+2}{2}\right)^{-1}. \quad \blacksquare$$

The following theorems give some relations among the ADQRs of the estimators $\widehat{\boldsymbol{\beta}}_{\phi_2}, \widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}, \widehat{\boldsymbol{\beta}}_{\phi_2}^{SRE}, \widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{PTE}, \widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{SPT}, \widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{S}, \widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{S+}$ and $\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{PTE+}$ under contiguous alternative hypotheses. In the following, we shall denote by $Ch_{\max}(\mathbf{A})$ and $Ch_{\min}(\mathbf{A})$ the largest and the smallest eigenvalues of the matrix $\mathbf{M}$.

Theorem 9. *Under contiguous alternative hypotheses $H_{1,N} : \mathbf{f}(\boldsymbol{\beta}_N) = N^{-1/2}\boldsymbol{\delta}$, $r > 2$ and assuming that $\mathbf{M}$ verifies*

$$\operatorname{trace}(\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{M}) \geq \frac{r+2}{2} Ch_{\max}(\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{M}), \tag{15}$$

we have $R\left(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{S+};\mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{S};\mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_{\phi_2};\mathbf{M}\right)$, $\forall\lambda$. If $\lambda \to 0$, $R\left(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{S};\mathbf{M}\right) \to R\left(\widehat{\boldsymbol{\beta}}_{\phi_2};\mathbf{M}\right)$.

Proof. From Theorem 7 we have,

$$\begin{aligned} &R\left(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{S+};\mathbf{M}\right) - R\left(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{S};\mathbf{M}\right) \\ &\quad = -\operatorname{trace}(\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{M})E\left[\left(1-(r-2)\chi_{r+2}^{-2}(\lambda)\right)^2 I_{(0,r-2)}\left(\chi_{r+2}^2(\lambda)\right)\right] \\ &\qquad - \boldsymbol{\delta}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{ML}\boldsymbol{\delta}E\left[\left(1-(r-2)\chi_{r+4}^{-2}(\lambda)\right)^2 I_{(0,r-2)}\left(\chi_{r+4}^2(\lambda)\right)\right] \\ &\qquad - 2\boldsymbol{\delta}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{ML}\boldsymbol{\delta}E\left[\left((r-2)\chi_{r+2}^{-2}(\lambda)-1\right) I_{(0,r-2)}\left(\chi_{r+2}^2(\lambda)\right)\right]. \end{aligned}$$

Therefore, $R\left(\widehat{\boldsymbol{\beta}}_{\phi_2}^{S+};\mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{S};\mathbf{M}\right)$, because

$$\begin{aligned} &E\left[\left(1-(r-2)\chi_{r+2}^{-2}(\lambda)\right) I_{(0,r-2)}\left(\chi_{r+2}^2(\lambda)\right)\right] \\ &\quad = \int_0^{r-2}\left(1-(r-2)x^{-1}\right)\mathrm{d}G_{r+2}(x;\lambda) < 0. \end{aligned}$$

By Courant's Theorem, it is a simple exercise to establish that $\boldsymbol{\delta}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{ML}\boldsymbol{\delta} \leq \lambda Ch_{\max}(\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{M})$. By Theorem 7 we get,

$$\begin{aligned} R\left(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{S};\mathbf{M}\right) - R\left(\widehat{\boldsymbol{\beta}}_{\phi_2};\mathbf{M}\right) = &-(r-2)\operatorname{trace}(\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{Y}})\left\{(r-2)E\left[\chi_{r+2}^{-4}(\lambda)\right]\right. \\ &\left.+\left[1-\frac{\boldsymbol{\delta}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{ML}\boldsymbol{\delta}(r+2)}{2\lambda\operatorname{trace}(\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{Y}})}\right]2\lambda E\left[\chi_{r+4}^{-4}(\lambda)\right]\right\} \leq 0. \end{aligned}$$

The last inequality follows by (15). ■

Theorem 10. *Under contiguous alternative hypotheses $H_{1,N}$: $\mathbf{f}\left(\boldsymbol{\beta}_N\right) = N^{-1/2}\boldsymbol{\delta}$, we have the following relations for the ADQR,*

(a) $R\left(\widehat{\boldsymbol{\beta}}_{\phi_2}^{SRE}; \mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_{\phi_2}; \mathbf{M}\right)$ *iff*

$$\lambda \leq \left(1 - a^2\right) \operatorname{trace}(\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{Y}})\,(1-a)^{-2}\,(Ch_{\max}(\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{M}))^{-1}$$

and $R\left(\widehat{\boldsymbol{\beta}}_{\phi_2}; \mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_{\phi_2}^{SRE}; \mathbf{M}\right)$ *iff*

$$\lambda \geq \left(1 - a^2\right) \operatorname{trace}(\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{Y}})\,(1-a)^{-2}\,(Ch_{\min}(\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{M}))^{-1}$$

(b) $R\left(\widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}; \mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_{\phi_2}; \mathbf{M}\right)$ *iff* $\lambda \leq \operatorname{trace}(\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{Y}})\,(Ch_{\max}(\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{M}))^{-1}$ *and* $R\left(\widehat{\boldsymbol{\beta}}_{\phi_2}; \mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}; \mathbf{M}\right)$ *iff* $\lambda \geq \operatorname{trace}(\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{Y}})\,(Ch_{\min}(\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{M}))^{-1}$.

(c) $R\left(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{SPT}; \mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{PTE}; \mathbf{M}\right)$ *iff*

$$\lambda \geq (1-a)G_{r+2}\left(\chi^2_{r;\alpha};\lambda\right) \operatorname{trace}(\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{Y}})\,(Ch_{\min}(\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{M}))^{-1} \times \left[2G_{r+2}\left(\chi^2_{r;\alpha};\lambda\right) - (1-a)G_{r+4}\left(\chi^2_{r;\alpha};\lambda\right)\right]^{-1}$$

and $R\left(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{PTE}; \mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{SPT}; \mathbf{M}\right)$ *iff*

$$\lambda \leq (1-a)G_{r+2}\left(\chi^2_{r;\alpha};\lambda\right) \operatorname{trace}(\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{Y}})\,(Ch_{\max}(\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{M}))^{-1} \times \left[2G_{r+2}\left(\chi^2_{r;\alpha};\lambda\right) - (1-a)G_{r+4}\left(\chi^2_{r;\alpha};\lambda\right)\right]^{-1}.$$

(d) $R\left(\widehat{\boldsymbol{\beta}}_{\phi_2}; \mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{PTE}; \mathbf{M}\right)$ *iff*

$$\lambda \geq \operatorname{trace}(\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{Y}})G_{r+2}\left(\chi^2_{r;\alpha};\lambda\right)(Ch_{\min}(\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{M}))^{-1} \times \left[2G_{r+2}\left(\chi^2_{r;\alpha};\lambda\right) - G_{r+4}\left(\chi^2_{r;\alpha};\lambda\right)\right]^{-1}$$

and $R\left(\widehat{\boldsymbol{\beta}}_{\phi_2}; \mathbf{M}\right) \geq R\left(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{PTE}; \mathbf{M}\right)$ *iff*

$$\lambda \leq \operatorname{trace}(\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{Y}})G_{r+2}\left(\chi^2_{r;\alpha};\lambda\right)(Ch_{\max}(\boldsymbol{\Sigma}_{\mathbf{S}}\mathbf{M}))^{-1} \times \left[2G_{r+2}\left(\chi^2_{r;\alpha};\lambda\right) - G_{r+4}\left(\chi^2_{r;\alpha};\lambda\right)\right]^{-1}.$$

If $\lambda \to \infty$ *or* $\alpha \to 1$, $R\left(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{PTE}; \mathbf{M}\right) \to R\left(\widehat{\boldsymbol{\beta}}_{\phi_2}; \mathbf{M}\right)$.

(e) $R\left(\widehat{\boldsymbol{\beta}}_{\phi_2}^{H_0}; \mathbf{M}\right) \leq R\left(\widehat{\boldsymbol{\beta}}_{\phi_1,\phi_2}^{PTE}; \mathbf{M}\right)$ *iff*

$$\lambda \leq \left(1 - G_{r+2}\left(\chi^2_{r;\alpha};\lambda\right)\right) \operatorname{trace}(\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{Y}})\,(Ch_{\max}(\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{M}))^{-1} \times \left[1 - 2G_{r+2}\left(\chi^2_{r;\alpha};\lambda\right) + G_{r+4}\left(\chi^2_{r;\alpha};\lambda\right)\right]^{-1}$$

and $R\left(\widehat{\boldsymbol{\beta}}^{PTE}_{\phi_1,\phi_2};\mathbf{M}\right) \le R\left(\widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2};\mathbf{M}\right)$ *iff*

$$\lambda \ge \left(1 - G_{r+2}\left(\chi^2_{r;\alpha};\lambda\right)\right)\operatorname{trace}(\mathbf{M}\boldsymbol{\Sigma}_{\mathbf{Y}})\left(Ch_{\min}(\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{M})\right)^{-1} \times \left[1 - 2G_{r+2}\left(\chi^2_{r;\alpha};\lambda\right) + G_{r+4}\left(\chi^2_{r;\alpha};\lambda\right)\right]^{-1}.$$

If $\alpha \to 0$, $R\left(\widehat{\boldsymbol{\beta}}^{PTE}_{\phi_1,\phi_2};\mathbf{M}\right) \to R\left(\widehat{\boldsymbol{\beta}}^{H_0}_{\phi_2};\mathbf{M}\right)$.

(f) $R\left(\widehat{\boldsymbol{\beta}}^{S_+}_{\phi_1,\phi_2};\mathbf{M}\right) \le R\left(\widehat{\boldsymbol{\beta}}^{PTE_+}_{\phi_1,\phi_2};\mathbf{M}\right)$ *if* $\chi^2_{r,\alpha} < r-2$

(g) *Assuming that* $\mathbf{M} = \boldsymbol{\Sigma}_S^{-1}$, *we have* $R\left(\widehat{\boldsymbol{\beta}}^{PTE_+}_{\phi_1,\phi_2};\mathbf{M}\right) \le R\left(\widehat{\boldsymbol{\beta}}^{PTE}_{\phi_1,\phi_2};\mathbf{M}\right)$.

Proof. The results follow by Theorem 7, inequalities (2.2.18d)–(2.2.13h) in Saleh [15], and using Courant's Theorem. ■

5. Simulation results

The small-sample properties of the preliminary test estimators based on ϕ-divergence measures, $\widehat{\boldsymbol{\beta}}^{\mathrm{h}}_{\phi_1,\phi_2}$ are studied under a null hypothesis as well as under contiguous alternative hypotheses using a Monte Carlo experiment. In order to carry out the experiment, we are going to consider the parametric family of ϕ-divergence measures based on

$$\phi_\lambda(x) = \begin{cases} \dfrac{x^{\lambda+1} - x - \lambda(x-1)}{\lambda(\lambda+1)}, & \lambda \ne 0, -1 \\ x\log x - x + 1, & \lambda = 0 \\ \log x + x - 1, & \lambda = -1, \end{cases}$$

which was introduced and studied by Cressie and Read [3]. That it is to say, we consider for the study the preliminary test estimators

$$\widehat{\boldsymbol{\beta}}^{h}_{\lambda_1,\lambda_2} \equiv \widehat{\boldsymbol{\beta}}^{\mathrm{h}}_{\phi_{\lambda_1},\phi_{\lambda_2}} = \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_{\lambda_2}} + \left(1 - h(T_N^{\phi_{\lambda_1},\phi_{\lambda_2}})\right)\left(\widehat{\boldsymbol{\beta}}_{\phi_{\lambda_2}} - \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_{\lambda_2}}\right),$$

for $\lambda_1 = 0, 2/3$ and 1, $\lambda_2 = 0, 2/3$ and 1 and the choices of function h as in Section 2. Note that $\widehat{\boldsymbol{\beta}}^{SRE}_{\lambda_1,\lambda_2}$ and $\widehat{\boldsymbol{\beta}}^{PTE}_{\lambda_1,\lambda_2}$ depend on the parameter $a \in (0,1)$ which we take as 0.5 for our study.

We consider a logistic regression model consisting of a dichotomous dependent variable and four normally distributed, with zero mean and unit variance, explanatory variables. We generated 10 000 samples of different sample sizes $\mathbf{n} = (n_1,\ldots,n_N)^{\mathrm{T}} \in \mathcal{N} = \{n^1, n^2\}$ with $n_i^1 = 30$, $i = 1,\ldots,8$, $n^2 = (25, 25, 25, 25, 10, 10, 10, 10)$. The regression coefficients $\boldsymbol{\beta}^{\mathrm{T}} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ were generated from a uniform over $(0, 2)$.

To have a general idea about overall performance of each of the estimators, the summed mean squared error (SMSE) is computed under the null hypothesis

$$H_0 : \mathbf{K}^{\mathrm{T}}\boldsymbol{\beta} = \mathbf{m},$$

where $\mathbf{K}^{\mathrm{T}} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$ and $\mathbf{m} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ as well as the contiguous alternative hypotheses

$$H_{\delta,N} : \mathbf{K}^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{m} = N^{-1/2}\boldsymbol{\delta}$$

for $\boldsymbol{\delta} = (1, 1, 1)$ and $\boldsymbol{\delta} = (-1, 10, -5)$.

Table 1
SMSE of the estimates for $\boldsymbol{\delta} = (0, 0, 0)$, $\mathbf{n} = n^1$

λ_1	0			2/3			1		
λ_2	0	2/3	1	0	2/3	1	0	2/3	1
$\widehat{\boldsymbol{\beta}}_{\phi_{\lambda_2}}$	0.3905	0.3647	0.3578	0.3905	0.3647	0.3578	0.3905	0.3647	0.3578
$\widehat{\boldsymbol{\beta}}^{H_0}_{\phi_{\lambda_2}}$	0.0659	0.0596	0.0590	0.0659	0.0596	0.0590	0.0659	0.0596	0.0590
$\widehat{\boldsymbol{\beta}}^{SRE}_{\lambda_1,\lambda_2}$	0.1484	0.1365	0.1338	0.1484	0.1365	0.1338	0.1484	0.1365	0.1338
$\widehat{\boldsymbol{\beta}}^{PTE}_{\lambda_1,\lambda_2}$	0.1297	0.1161	0.1143	0.1278	0.1130	0.1104	0.1275	0.1117	0.1085
$\widehat{\boldsymbol{\beta}}^{SPT}_{\lambda_1,\lambda_2}$	0.1961	0.1789	0.1754	0.1947	0.1765	0.1725	0.1944	0.1755	0.1711
$\widehat{\boldsymbol{\beta}}^{S}_{\lambda_1,\lambda_2}$	0.2788	0.2659	0.2563	0.2781	0.2645	0.2546	0.2780	0.2641	0.2541
$\widehat{\boldsymbol{\beta}}^{S+}_{\lambda_1,\lambda_2}$	0.2458	0.2256	0.2139	0.2451	0.2241	0.2190	0.2450	0.2236	0.2185
$\widehat{\boldsymbol{\beta}}^{PTE+}_{\lambda_1,\lambda_2}$	0.1178	0.1055	0.1039	0.1162	0.1028	0.1005	0.1160	0.1017	0.0990

Table 2
SMSE of the estimates for $\boldsymbol{\delta} = (1, 1, 1)$, $\mathbf{n} = n^1$

λ_1	0			2/3			1		
λ_2	0	2/3	1	0	2/3	1	0	2/3	1
$\widehat{\boldsymbol{\beta}}_{\phi_{\lambda_2}}$	0.3989	0.3715	0.3643	0.3989	0.3715	0.3643	0.3989	0.3715	0.3643
$\widehat{\boldsymbol{\beta}}^{H_0}_{\phi_{\lambda_2}}$	0.0816	0.0732	0.0721	0.0816	0.0732	0.0721	0.0816	0.0732	0.0721
$\widehat{\boldsymbol{\beta}}^{SRE}_{\lambda_1,\lambda_2}$	0.1543	0.1414	0.1385	0.1543	0.1414	0.1385	0.1543	0.1414	0.1385
$\widehat{\boldsymbol{\beta}}^{PTE}_{\lambda_1,\lambda_2}$	0.1588	0.1427	0.1390	0.1559	0.1378	0.1341	0.1559	0.1366	0.1320
$\widehat{\boldsymbol{\beta}}^{SPT}_{\lambda_1,\lambda_2}$	0.2134	0.1946	0.1901	0.2111	0.1909	0.1862	0.2111	0.1900	0.1846
$\widehat{\boldsymbol{\beta}}^{S}_{\lambda_1,\lambda_2}$	0.2924	0.2650	0.2610	0.2913	0.2632	0.2589	0.2911	0.2625	0.2581
$\widehat{\boldsymbol{\beta}}^{S+}_{\lambda_1,\lambda_2}$	0.2621	0.2406	0.2356	0.2609	0.2386	0.2334	0.2607	0.2380	0.2326
$\widehat{\boldsymbol{\beta}}^{PTE+}_{\lambda_1,\lambda_2}$	0.1442	0.1294	0.1262	0.1417	0.1252	0.1219	0.1416	0.1241	0.1202

From Tables 1 and 4, it is clear that $\widehat{\boldsymbol{\beta}}^{H_0}_{\phi_{\lambda_2}} \prec \widehat{\boldsymbol{\beta}}^{PTE+}_{\lambda_1,\lambda_2} \prec \widehat{\boldsymbol{\beta}}^{PTE}_{\lambda_1,\lambda_2} \prec \widehat{\boldsymbol{\beta}}^{SRE}_{\lambda_1,\lambda_2} \prec \widehat{\boldsymbol{\beta}}^{SPT}_{\lambda_1,\lambda_2} \prec \widehat{\boldsymbol{\beta}}^{S+}_{\lambda_1,\lambda_2} \prec \widehat{\boldsymbol{\beta}}^{S}_{\lambda_1,\lambda_2} \prec \widehat{\boldsymbol{\beta}}_{\phi_{\lambda_2}}$ where '$\prec$' means 'prefer to' the same relation as we prove in Theorem 8. Therefore, we can conclude that the asymptotic results of Theorem 8 are also valid for small and moderate sample sizes.

For the alternative corresponding to $\boldsymbol{\delta} = (1, 1, 1)$ and $\mathbf{n} = n^1$, it can be seen in Table 2 that the above relations among the estimators hold. A little change happens when $\mathbf{n} = n^2$ (Table 5), since the order of $\widehat{\boldsymbol{\beta}}^{SRE}_{\lambda_1,\lambda_2}$ and $\widehat{\boldsymbol{\beta}}^{PTE}_{\lambda_1,\lambda_2}$ is reverse. This means that the result of Theorem 8 is true when we move away a little bit from the null hypothesis.

From Table 3, the arrangement of the estimators is $\widehat{\boldsymbol{\beta}}^{SRE}_{\lambda_1,\lambda_2} \prec \widehat{\boldsymbol{\beta}}^{S+}_{\lambda_1,\lambda_2} \prec \widehat{\boldsymbol{\beta}}^{S}_{\lambda_1,\lambda_2} \prec \widehat{\boldsymbol{\beta}}^{PTE+}_{\lambda_1,\lambda_2} \prec \widehat{\boldsymbol{\beta}}_{\phi_{\lambda_2}} \prec \widehat{\boldsymbol{\beta}}^{SPT}_{\lambda_1,\lambda_2} \prec \widehat{\boldsymbol{\beta}}^{PTE}_{\lambda_1,\lambda_2} \prec \widehat{\boldsymbol{\beta}}^{H_0}_{\phi_{\lambda_2}}$ for $\boldsymbol{\delta} = (-1, 10, -5)$ and $\mathbf{n} = n^1$. For $\boldsymbol{\delta} = (-1, 10, -5)$ and $\mathbf{n} = n^2$ (Table 6), the only difference is the behavior of $\widehat{\boldsymbol{\beta}}^{H_0}_{\phi_{\lambda_2}}$ and $\widehat{\boldsymbol{\beta}}_{\phi_{\lambda_2}}$. Therefore, under

Table 3
SMSE of the estimates for $\boldsymbol{\delta} = (-1, 10, -5)$, $\mathbf{n} = n^1$

λ_1	0			2/3			1		
λ_2	0	2/3	1	0	2/3	1	0	2/3	1
$\widehat{\boldsymbol{\beta}}_{\phi_{\lambda_2}}$	0.4848	0.4422	0.4308	0.4848	0.4422	0.4308	0.4848	0.4422	0.4308
$\widehat{\boldsymbol{\beta}}^{H_0}_{\phi_{\lambda_2}}$	0.6355	0.6709	0.6941	0.6355	0.6709	0.6941	0.6355	0.6709	0.6941
$\widehat{\boldsymbol{\beta}}^{SRE}_{\lambda_1,\lambda_2}$	0.2866	0.2881	0.2936	0.2866	0.2881	0.2936	0.2866	0.2881	0.2936
$\widehat{\boldsymbol{\beta}}^{PTE}_{\lambda_1,\lambda_2}$	0.4854	0.4428	0.4313	0.4854	0.4428	0.4314	0.4854	0.4428	0.4314
$\widehat{\boldsymbol{\beta}}^{SPT}_{\lambda_1,\lambda_2}$	0.4850	0.4424	0.4309	0.4850	0.4424	0.4310	0.4850	0.4424	0.4310
$\widehat{\boldsymbol{\beta}}^{S}_{\lambda_1,\lambda_2}$	0.4586	0.4187	0.4083	0.4579	0.4173	0.4069	0.4580	0.4171	0.4065
$\widehat{\boldsymbol{\beta}}^{S+}_{\lambda_1,\lambda_2}$	0.4586	0.4187	0.4083	0.4579	0.4173	0.4069	0.4580	0.4171	0.4065
$\widehat{\boldsymbol{\beta}}^{PTE+}_{\lambda_1,\lambda_2}$	0.4591	0.4192	0.4088	0.4584	0.4179	0.4074	0.4585	0.4177	0.4071

Table 4
SMSE of the estimates for $\boldsymbol{\delta} = (0, 0, 0)$, $\mathbf{n} = n^2$

λ_1	0			2/3			1		
λ_2	0	2/3	1	0	2/3	1	0	2/3	1
$\widehat{\boldsymbol{\beta}}_{\phi_{\lambda_2}}$	1.1212	1.0198	0.9921	1.1212	1.0198	0.9921	1.1212	1.0198	0.9921
$\widehat{\boldsymbol{\beta}}^{H_0}_{\phi_{\lambda_2}}$	0.1222	0.1063	0.1051	0.1222	0.1063	0.1051	0.1222	0.1063	0.1051
$\widehat{\boldsymbol{\beta}}^{SRE}_{\lambda_1,\lambda_2}$	0.3800	0.3398	0.3306	0.3800	0.3398	0.3306	0.3800	0.3398	0.3306
$\widehat{\boldsymbol{\beta}}^{PTE}_{\lambda_1,\lambda_2}$	0.3571	0.3091	0.3198	0.3187	0.2665	0.2563	0.3182	0.2545	0.2447
$\widehat{\boldsymbol{\beta}}^{SPT}_{\lambda_1,\lambda_2}$	0.5554	0.4921	0.4917	0.5272	0.4604	0.4447	0.5269	0.4516	0.4360
$\widehat{\boldsymbol{\beta}}^{S}_{\lambda_1,\lambda_2}$	0.7928	0.7078	0.6984	0.7809	0.6918	0.6812	0.7784	0.6870	0.6759
$\widehat{\boldsymbol{\beta}}^{S+}_{\lambda_1,\lambda_2}$	0.7152	0.6383	0.6202	0.7031	0.6218	0.6025	0.7006	0.6168	0.5969
$\widehat{\boldsymbol{\beta}}^{PTE+}_{\lambda_1,\lambda_2}$	0.3135	0.2712	0.2789	0.2818	0.2357	0.2273	0.2812	0.2260	0.2178

alternative hypotheses far away from the null hypothesis $\widehat{\boldsymbol{\beta}}^{SRE}_{\lambda_1,\lambda_2}$, $\widehat{\boldsymbol{\beta}}^{S+}_{\lambda_1,\lambda_2}$ and $\widehat{\boldsymbol{\beta}}^{S}_{\lambda_1,\lambda_2}$ seem to be the preferable.

Another point is that in our case each estimator of the first column in fact is a family of estimators, so we can ask us about the best members of those families. For all tables, i.e., under the null hypothesis or alternative hypotheses the best choice is $\lambda_1 = \lambda_2 = 1$ for all the cases except when $\boldsymbol{\delta} = (-1, 10, -5)$ and $\mathbf{n} = n^1$ for $\widehat{\boldsymbol{\beta}}^{SPT}_{\lambda_1,\lambda_2}$ and $\widehat{\boldsymbol{\beta}}^{PTE}_{\lambda_1,\lambda_2}$.

6. Conclusions

In this paper, we have considered a new family of estimators for the parameters of the GLM with binary data. This new family of estimators depends on both minimum ϕ-divergence estimators and ϕ-divergence statistics. Minimum ϕ-divergence estimators appear as a natural generalization of the maximum likelihood estimator in the GLM. Based on their asymptotic

Table 5
SMSE of the estimates for $\boldsymbol{\delta} = (1, 1, 1)$, $\mathbf{n} = n^2$

λ_1	0			2/3			1		
λ_2	0	2/3	1	0	2/3	1	0	2/3	1
$\widehat{\boldsymbol{\beta}}_{\phi_{\lambda_2}}$	1.1412	1.0253	0.9941	1.1412	1.0253	0.9941	1.1412	1.0253	0.9941
$\widehat{\boldsymbol{\beta}}^{H_0}_{\phi_{\lambda_2}}$	0.1518	0.1300	0.1279	0.1518	0.1300	0.1279	0.1518	0.1300	0.1279
$\widehat{\boldsymbol{\beta}}^{SRE}_{\lambda_1,\lambda_2}$	0.3946	0.3474	0.3367	0.3946	0.3474	0.3367	0.3946	0.3474	0.3367
$\widehat{\boldsymbol{\beta}}^{PTE}_{\lambda_1,\lambda_2}$	0.4643	0.4081	0.4019	0.4421	0.3764	0.3657	0.4319	0.3669	0.3559
$\widehat{\boldsymbol{\beta}}^{SPT}_{\lambda_1,\lambda_2}$	0.6297	0.5575	0.5441	0.6130	0.5337	0.5169	0.6056	0.5266	0.5096
$\widehat{\boldsymbol{\beta}}^{S}_{\lambda_1,\lambda_2}$	0.8283	0.7227	0.7014	0.8173	0.7082	0.6861	0.8151	0.7040	0.6813
$\widehat{\boldsymbol{\beta}}^{S+}_{\lambda_1,\lambda_2}$	0.7601	0.6723	0.6516	0.7489	0.6575	0.6360	0.7467	0.6531	0.6310
$\widehat{\boldsymbol{\beta}}^{PTE+}_{\lambda_1,\lambda_2}$	0.4083	0.3583	0.3529	0.3882	0.3298	0.3207	0.3801	0.3216	0.3121

Table 6
SMSE of the estimates for $\boldsymbol{\delta} = (-1, 10, -5)$, $\mathbf{n} = n^2$

λ_1	0			2/3			1		
λ_2	0	2/3	1	0	2/3	1	0	2/3	1
$\widehat{\boldsymbol{\beta}}_{\phi_{\lambda_2}}$	1.8120	1.6531	1.6142	1.8120	1.6531	1.6142	1.8120	1.6531	1.6142
$\widehat{\boldsymbol{\beta}}^{H_0}_{\phi_{\lambda_2}}$	1.1552	1.2624	1.3097	1.1552	1.2624	1.3097	1.1552	1.2624	1.3097
$\widehat{\boldsymbol{\beta}}^{SRE}_{\lambda_1,\lambda_2}$	0.7092	0.7135	0.7232	0.7092	0.7135	0.7232	0.7092	0.7135	0.7232
$\widehat{\boldsymbol{\beta}}^{PTE}_{\lambda_1,\lambda_2}$	1.8091	1.6507	1.6119	1.8091	1.6507	1.6119	1.8091	1.6507	1.6119
$\widehat{\boldsymbol{\beta}}^{SPT}_{\lambda_1,\lambda_2}$	1.8090	1.6505	1.6117	1.8090	1.650	1.6117	1.8090	1.6505	1.6117
$\widehat{\boldsymbol{\beta}}^{S}_{\lambda_1,\lambda_2}$	1.7171	1.5684	1.5327	1.7090	1.5592	1.5233	1.7077	1.5571	1.5210
$\widehat{\boldsymbol{\beta}}^{S+}_{\lambda_1,\lambda_2}$	1.7171	1.5684	1.5327	1.7090	1.5592	1.5233	1.7077	1.5571	1.5210
$\widehat{\boldsymbol{\beta}}^{PTE+}_{\lambda_1,\lambda_2}$	1.7172	1.5686	1.5329	1.7091	1.5594	1.5235	1.7078	1.5573	1.5212

quadratic risk, some new estimators emerge for the GLM with binary data. These results are asymptotic (large sample sizes), but for small and moderate sample sizes we get results in accordance with the asymptotic results. These results are established when the postulated model is true as well as when is not true, i.e., when we consider contiguous alternative hypotheses.

Acknowledgments

This work was supported by Grants MTM2006-06872 and UCM2007-910707. The authors would like to thanks the referees and the Associate Editor for critically reading this paper and making suggestions.

References

[1] S.M. Ali, S.D. Silvey, A general class of coefficients of divergence of one distribution from another, J. Roy. Statist. Soc. B 26 (1966) 131–142.

[2] T.A. Bancroft, On biases in estimation due to use of preliminary tests of significance, Ann. Math. Statist. 15 (1944) 190–204.
[3] N. Cressie, T.R.C. Read, Multinomial goodness-of-fit tests, J. Roy. Statist. Soc. B 46 (1984) 440–464.
[4] I. Csiszár, Eine Informationtheorestiche Ungleichung und ihre Anwendung anf den Beweis der Ergodizität Markoffshen Ketten. Publications of the Mathematical Institute of Hungarian Academy of Sciences, Series A, 8 (1963) 84–108.
[5] T.S. Ferguson, A course in large sample theory, in: Texts in Statistical Science, Chapman & Hall, New York, 1996.
[6] W. James, C. Stein, Estimation with quadratic loss, in: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, CA, 1961, pp. 361–379.
[7] S. Kullback, Kullback information, in: S. Kotz, N.L. Johnson (Eds.), Encyclopedia of Statistical Sciences, vol. 4, John Wiley, New York, 1985, pp. 421–425.
[8] M.A. Matin, A.K. Md, E. Saleh, Some improved estimators in logistic regression model, J. Stat. Res. 39 (2) (2005) 37–58.
[9] M.A Matin, A.K. Md, E. Saleh, Small-sample properties of some improved estimators in logistic regression model with skew-normally distributed explanatory variables, J. Stat. Res. 40 (1) (2006) 1–21.
[10] P. McCullagh, J.A. Nelder, Generalized Linear Models, second ed., Chapman and Hall, London, 1989.
[11] L. Pardo, Statistical Inference Based on Divergence Measures, Chapman & Hall/CRC, New York, 2006.
[12] J.A. Pardo, M.C. Pardo, Minimum phi-divergence estimators and phi-divergence statistics in GLM with binary data, Methodol. Comput. Appl. (2008), doi:10.1007/s11009-007-9054-2.
[13] A.K. Md, E. Saleh, P.K. Sen, Nonparametric estimation of location parameter after a preliminary test on regression, Ann. Statist. 6 (1978) 154–168.
[14] A.K. Md, E. Saleh, P.K. Sen, On shrinkage least-squares estimation in a parallelism problem, Comm. Statist. Theory Methods 15 (1986) 1451–1466.
[15] A.K. Md, E. Saleh, Theory of Preliminary Test and Stein-Type Estimation with Applications, Wiley, 2006.
[16] P.K. Sen, A.K. Md, E. Saleh, On preliminary test and shrinkage M-estimation in linear models, Ann. Statist. 15 (1987) 1580–1592.
[17] C. Stein, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, in: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, University of California Press, Berkeley, CA, 1956, pp. 197–206.
[18] I. Vajda, Theory of Statistical Inference and Information, Kluwer, Boston, 1989.