Confidence sets and coverage probabilities based on preliminary estimators in logistic regression models

M.L. Menéndez\textsuperscript{a}, L. Pardo\textsuperscript{b}, M.C. Pardo\textsuperscript{b,∗}

\textsuperscript{a} Department of Applied Mathematics, E.T.S.A.M., Technical University of Madrid, Spain
\textsuperscript{b} Department of Statistics and O.R. I, Complutense University of Madrid, Spain

\textbf{A R T I C L E I N F O}

Article history:
Received 28 October 2007
Received in revised form 11 April 2008

MSC:
62F25
62F30
62F12
62J99

Keywords:
Preliminary test estimator
Asymptotic coverage probability
Minimum \phi-divergence estimator
Recentered confidence set

\textbf{A B S T R A C T}

In this paper we present recentered confidence sets for the parameters of a logistic regression model based on preliminary minimum \phi-divergence estimators. Asymptotic coverage probabilities are given as well as a simulation study in order to analyze the coverage probabilities for small and moderate sample sizes.

© 2008 Elsevier B.V. All rights reserved.

\section{1. Introduction}

Let \(Y_i, i = 1, \ldots, n\), be independent binomial random variables with parameters \(\pi_i\) and \(n_i, i = 1, \ldots, n\). We shall assume that the parameters \(\pi_i = \Pr(Y_i = 1), i = 1, \ldots, n\), depend on the unknown parameters \(\beta = (\beta_0, \ldots, \beta_k)^T, \beta_i \in (-\infty, \infty)\) and explanatory variables \(x_i = (x_{i0}, \ldots, x_{ik}), x_{i0} = 1, i = 1, \ldots, n\) through the linear predictor

\[
\text{logit}(\pi_i) = \sum_{j=0}^{k} x_{ij}\beta_j, \quad i = 1, \ldots, n
\]  

where \(\text{logit}(p) = \log(p/(1-p))\). In the following we shall denote the binomial parameter \(\pi_i\) by \(\pi_i = \pi(x_i^T\beta)\) and by \(\mathbf{X}\) the \(n \times (k + 1)\) matrix with rows \(x_i, i = 1, \ldots, n\). We also assume that \(\text{rank}(\mathbf{X}) = k + 1\).

In [4] a preliminary test estimator for \(\hat{\beta}_{\phi_1,\phi_2}\) (see (8) in Section 2) was considered. This estimator is based on the restricted \(\hat{\beta}_{\phi_0}\) (see (7) in Section 2) and the unrestricted \(\hat{\beta}_{\phi_2}\) (see (2) in Section 2) minimum \phi-divergence estimators of \(\beta\). An important problem for the point estimation of \(\beta\) is to provide associated confidence sets. In this paper we consider asymptotic recentered confidence sets for \(\beta\) based on \(\hat{\beta}_{\phi_1,\phi_2}, \hat{\beta}_{\phi_0}\) and \(\hat{\beta}_{\phi_2}\) and we study their coverage probabilities.

In Section 2 we present some notation as well as some preliminary results that will be necessary in the paper. Section 3 is devoted to the definition of recentered confidence sets as well as an analytical study of their asymptotic coverage probabilities. Finally, in Section 4 a simulation study is carried out in order to analyze the coverage probabilities for small and moderate sample sizes and different choices on the functions \(\phi_1\) and \(\phi_2\).
2. Background and notation

We denote by \( y_i \), the number of “successes” associated with the binomial random variable \( Y_i, i = 1, \ldots, n \). Minimum \( \phi_2 \)-divergence estimator (\( M\phi_2E \)) of \( \beta, \hat{\beta}_{\phi_2} = \hat{\beta}_{\phi_2}(Y_1, \ldots, Y_n) \) is defined as

\[
\hat{\beta}_{\phi_2} = \arg \min_{\beta \in \Theta} \sum_{i=1}^{n} n_i D_{\phi_2}(\tilde{p}_i, \pi_i(\beta))
\]

(2)

where

\[
\tilde{p}_i = \left( \frac{y_i}{n_i}, \frac{n_i - y_i}{n_i} \right)^T \quad \text{and} \quad \pi_i(\beta) = \left( \pi(\chi_i^T \beta), 1 - \pi(\chi_i^T \beta) \right)^T. \quad i = 1, \ldots, n,
\]

\( \Theta = \{ \beta = (\beta_0, \beta_1, \ldots, \beta_k) : \beta_j \in (-\infty, +\infty), j = 0, \ldots, k \} \) and \( D_{\phi_2}(\tilde{p}, \pi(\beta)) \) is the \( \phi_2 \)-divergence measure between the probability vectors \( \tilde{p} \) and \( \pi(\beta) \), given by

\[
D_{\phi_2}(\tilde{p}_i, \pi_i(\beta)) = \pi(\chi_i^T \beta) \phi_2 \left( \frac{y_i}{\pi(\chi_i^T \beta) n_i} \right) + \left( 1 - \pi(\chi_i^T \beta) \right) \phi_2 \left( \frac{n_i - y_i}{1 - \pi(\chi_i^T \beta) n_i} \right),
\]

(4)

\( \phi_2 \in \Phi, \phi \) is the class of all convex functions \( \phi_2(x), x > 0 \), such that at \( x = 1, \phi_2(1) = \phi_2'(1) = 0, \phi_2''(1) > 0 \). In (4) we shall assume the conventions \( 0\phi_2(0/0) = 0 \) and \( 0\phi_2(p/0) = p \lim_{u \to \infty} \phi_2(u) / u \). For a systematic study of \( \phi_2 \)-divergences see Pardo et al. [6].

For \( \phi_2(x) = x \log x - x + 1 \) we obtain in (4) the Kullback–Leibler divergence,

\[
D_{\text{Kull}}(\tilde{p}_i, \pi_i(\beta)) = y_i \log \frac{y_i}{\pi(\chi_i^T \beta) n_i} + (n_i - y_i) \log \frac{(n_i - y_i)}{(1 - \pi(\chi_i^T \beta) n_i)}
\]

and it is immediately seen that

\[
\sum_{i=1}^{n} n_i D_{\text{Kull}}(\tilde{p}_i, \pi_i(\beta)) = -l(\beta) + k,
\]

where \( l(\beta) \) is the loglikelihood function defined by

\[
l(\beta) = \sum_{i=1}^{n} \log \left( \pi(\chi_i^T \beta)^{y_i} \left( 1 - \pi(\chi_i^T \beta) \right)^{n_i - y_i} \right).
\]

Therefore, the maximum likelihood estimator defined by \( \hat{\beta} = \arg \max_{\beta \in \Theta} l(\beta) \) can also be defined by

\[
\hat{\beta} = \arg \min_{\beta \in \Theta} \sum_{i=1}^{n} n_i D_{\text{Kull}}(\tilde{p}_i, \pi_i(\beta))
\]

and the minimum \( \phi_2 \)-divergence estimator defined in (2) is a natural extension of the maximum likelihood estimator.

We denote \( N = \sum_{i=1}^{n} n_i \),

\[
W_N(\beta) = \text{diag} \left( (C_i(\beta))^T \right) \text{diag} \left( (C_i(\beta)) \right), \quad i = 1, \ldots, n
\]

with

\[
C_i(\beta) = \left( \frac{n_i}{N} \pi(\chi_i^T \beta) \left( 1 - \pi(\chi_i^T \beta) \right) \right)^{1/2} \left( \frac{1 - \pi(\chi_i^T \beta)}{\pi(\chi_i^T \beta)} \right)^{1/2}.
\]

(5)

In the following we shall assume \( \lambda_i = \lim_{n \to \infty} n_i / N, i = 1, \ldots, n \). Under the assumption that \( \pi \) has continuous second partial derivatives in a neighborhood of the true value of the parameter \( \beta_0 \), and \( \phi_2(\beta) \) is twice differentiable at \( x > 0 \), \( \hat{\beta}_{\phi_2} \) verifies

\[
\sqrt{N} (\hat{\beta}_{\phi_2} - \beta_0) \xrightarrow{d} N \left( 0, (X'W(\beta_0)X)^{-1} \right),
\]

(6)

where \( W(\beta_0) = \lim_{n \to \infty} W_N(\beta_0) \). For more properties about \( \hat{\beta}_{\phi_2} \) see Pardo et al. [5].

Now we assume that we have the additional information that \( \beta \in \Theta_0 = \{ \beta \in \Theta : \beta^T K = m \} \), where \( K^T \) is any matrix of \( r \) rows and \( k + 1 \) columns and \( m \) is a vector of order \( r \) of specified constants. The minimum \( \phi_2 \)-divergence estimator restricted to \( \Theta_0 \) is given by

\[
\hat{\beta}_{\phi_2}^{k_0} = \arg \min_{\beta \in \Theta_0} \sum_{i=1}^{n} n_i D_{\phi_2}(\tilde{p}_i, \pi_i(\beta))
\]

(7)

We refer to it as the restricted minimum \( \phi_2 \)-divergence estimator (RM\( \phi_2E \)) of \( \beta \in \Theta_0 \). The RM\( \phi_2E \) verifies

\[
\sqrt{N} (\hat{\beta}_{\phi_2}^{k_0} - \beta_0) \xrightarrow{d} N \left( 0, (X'W(\beta_0)X)^{-1} \right),
\]

where \( H(\beta_0) = I - (X'W(\beta_0)X)^{-1} K (K^T (X'W(\beta_0)X)^{-1} K)^{-1} K^T \).
If we consider $\phi_2(x) = x \log x - x + 1$ in (7) we obtain the classical restricted maximum likelihood estimator. 

In [3] in order to test the compatibility of the restricted and the unrestricted minimum $\phi_2$-divergence estimators $\hat{\beta}_{\phi_2}$ and $\hat{\beta}_{\phi_2}^{H_0}$, i.e., for testing 

$$H_0 : K^T \beta = m \quad \text{versus} \quad H_1 : K^T \beta \neq m$$

the following family of $\phi$-divergence statistics was considered

$$T_{N, \phi_2}^{\phi_1, \phi_2} = \frac{2}{\phi_2^N(1)} \sum_{i=1}^{n} D_{\phi_2}(\pi_i(\hat{\beta}_{\phi_2}), \pi_i(\hat{\beta}_{\phi_2}^{H_0})), $$

where $\pi_i(\hat{\beta}_{\phi_2})$ and $\pi_i(\hat{\beta}_{\phi_2}^{H_0})$ are obtained from (3) replacing $\beta$ by $\hat{\beta}_{\phi_2}$ and $\hat{\beta}_{\phi_2}^{H_0}$ respectively. We can observe that the statistic $T_{N, \phi_2}^{\phi_1, \phi_2}$ involves two functions $\phi_1$ and $\phi_2$. The function $\phi_2$ is used to compute the minimum $\phi_2$-divergence estimators $\hat{\beta}_{\phi_2}$ and $\hat{\beta}_{\phi_2}^{H_0}$, while $\phi_1$ is used to calculate the “distance” between the two probability vectors.

It is interesting to observe that for $\phi_2(x) = \phi_1(x) = x \log x - x + 1$ we obtain $T_{N, \phi_2}^{\phi_1, \phi_2} = LR + o_p(1)$, where $LR$ is the likelihood-ratio test.

If we accept $H_0$ we choose the RM$\phi_2$E and if we reject $H_0$ we choose the M$\phi_2$E, i.e., the preliminary minimum $(\phi_1, \phi_2)$-divergence estimator,

$$\hat{\beta}_{\phi_1, \phi_2}^{Pre} = \hat{\beta}_{\phi_2}^{H_0} \{0, \chi_2, \infty\} (T_{N, \phi_2}^{\phi_1, \phi_2}) + \hat{\beta}_{\phi_2} \{1, \chi_{2, \infty}, \infty\} (T_{N, \phi_2}^{\phi_1, \phi_2})$$

or equivalently

$$\hat{\beta}_{\phi_1, \phi_2}^{Pre} = \hat{\beta}_{\phi_2}^{H_0} + (\hat{\beta}_{\phi_2} - \hat{\beta}_{\phi_2}^{H_0}) I_{\{\chi_{2, \infty}, \infty\}} (T_{N, \phi_2}^{\phi_1, \phi_2}),$$

(8)

where $I_A(y)$ denotes an indicator function taking the value 1 if $y \in A$ and 0 if $y \notin A$. Hence, the preliminary estimator depends on $\phi_1$ and $\phi_2$.

In [4] the asymptotic bias and the asymptotic distributional quadratic risk for $\hat{\beta}_{\phi_1, \phi_2}^{Pre}$, $\hat{\beta}_{\phi_2}$ and $\hat{\beta}_{\phi_2}^{H_0}$ were studied. A closely related problem is the confidence sets based on the preliminary test estimators. Our interest in this paper is to provide asymptotic confidence sets based on $\hat{\beta}_{\phi_1, \phi_2}^{Pre}$, $\hat{\beta}_{\phi_2}$ and $\hat{\beta}_{\phi_2}^{H_0}$, for contiguous alternative hypotheses and to obtain the asymptotic expressions for their coverage probabilities. Whereas exact expressions have been studied in the multinomial distributional problem, [1] among others, in logistic regression models it is not possible to obtain exact results. Recentered confidence sets are well documented in [7] for different statistical problems.

3. Coverage probabilities: An analytical study

We define the recentered confidence set based on the estimator $\hat{\beta}_{\phi_2}^*$, where $\hat{\beta}_{\phi_2}^*$ is equal to $\hat{\beta}_{\phi_1, \phi_2}^{Pre}$, $\hat{\beta}_{\phi_2}$ or $\hat{\beta}_{\phi_2}^{H_0}$, as

$$C_\beta (\hat{\beta}_{\phi_2}^*) = \left\{ \beta : N \| \beta - \hat{\beta}_{\phi_2}^* \|_{X^T W_\phi(\hat{\beta}_{\phi_2}) X} \leq \chi_{2+1, \alpha}^2 \right\},$$

where $\| Y \|_C^2 = Y^T C Y$.

We are going to see the asymptotic behavior of $\hat{\beta}_{\phi_1, \phi_2}^{Pre}$, $\hat{\beta}_{\phi_2}$ and $\hat{\beta}_{\phi_2}^{H_0}$ under fixed alternative hypotheses defined by

$$H_1 : K^T \beta = m + s$$

with $s \in \mathbb{R}^r$ and fixed. The main results are presented in the following theorem:

**Theorem 1.** Under fixed alternative hypotheses $H_1 : K^T \beta = m + s$ with $s \in \mathbb{R}^r$, we have:

(a) $\sqrt{N} (\hat{\beta}_{\phi_1, \phi_2}^{Pre} - \beta) = \sqrt{N} (\hat{\beta}_{\phi_2} - \beta) + o_p(1)$ .

(b) $\sqrt{N} (\hat{\beta}_{\phi_2}^{H_0} - \beta)$ has a degenerate asymptotic distribution.

**Proof.** (a) First we are going to establish that $T_{N, \phi_2}^{\phi_1, \phi_2} \to \infty$ as $N \to \infty$. On the one hand

$$\sqrt{N} (K^T \hat{\beta}_{\phi_2} - m) = \sqrt{N} K^T \hat{\beta}_{\phi_2} - \sqrt{N} m - \sqrt{N} K^T \beta + \sqrt{N} K^T \beta$$

$$= \sqrt{N} K^T (\hat{\beta}_{\phi_2} - \beta) + \sqrt{N} (K^T \beta - m)$$

$$= \sqrt{N} K^T (\hat{\beta}_{\phi_2} - \beta) + \sqrt{N} s$$
and in [3] we obtain,
\[ T_{N}^{\beta_{1}, \beta_{2}} = \sqrt{N} \left( \mathbf{K}^{T} \tilde{\beta}_{\phi_{2}} - \mathbf{m} \right)^{T} \left( \mathbf{K}^{T} \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X}^{-1} \mathbf{K} \right)^{-1} \sqrt{N} \left( \mathbf{K}^{T} \tilde{\beta}_{\phi_{2}} - \mathbf{m} \right) + o_{p} \ (1) \]
\[ = \sqrt{N} \left( \tilde{\beta}_{\phi_{2}} - \beta \right)^{T} \mathbf{K} \left( \mathbf{K}^{T} \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X}^{-1} \mathbf{K} \right)^{-1} \sqrt{N} \mathbf{K}^{T} \left( \tilde{\beta}_{\phi_{2}} - \beta \right) \]
\[ + \sqrt{Ns} \left( \mathbf{K}^{T} \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X}^{-1} \mathbf{K} \right)^{-1} \sqrt{Ns} + 2N \left( \tilde{\beta}_{\phi_{2}} - \beta \right)^{T} \mathbf{K} \left( \mathbf{K}^{T} \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X}^{-1} \mathbf{K} \right)^{-1} \mathbf{s}. \]

It is not difficult to see that
\[ \sqrt{N} \mathbf{K} \left( \mathbf{K}^{T} \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X}^{-1} \mathbf{K} \right)^{-1/2} \left( \tilde{\beta}_{\phi_{2}} - \beta \right) \xrightarrow[N \to \infty]{l} \mathcal{N} \left( \mathbf{0}, \mathbf{I} \right) \]
and
\[ Ns^{2} \left( \mathbf{K}^{T} \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X}^{-1} \mathbf{K} \right)^{-1} \mathbf{s} \xrightarrow[N \to \infty]{l} \infty \]
\[ 2N \left( \tilde{\beta}_{\phi_{2}} - \beta \right)^{T} \mathbf{K} \left( \mathbf{K}^{T} \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X}^{-1} \mathbf{K} \right)^{-1} \mathbf{s} \xrightarrow[N \to \infty]{l} \infty. \]

Therefore \( T_{N}^{\beta_{1}, \beta_{2}} \to \infty. \)

In order to establish (a) we consider, based on (8), the quadratic difference
\[ N \left\| \tilde{\beta}_{\phi_{1}, \phi_{2}}^{\text{Pre}} - \tilde{\beta}_{\phi_{2}} \right\|_{X^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X}}^{2} = N \left( \tilde{\beta}_{\phi_{2}} - \tilde{\beta}_{\phi_{2}}^{\text{H0}} \right)^{T} \left( \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X} \right) \left( \tilde{\beta}_{\phi_{2}} - \tilde{\beta}_{\phi_{2}}^{\text{H0}} \right) \mathbf{I}_{0, \mathcal{X}_{\alpha}^{2}} \left( T_{N}^{\beta_{1}, \beta_{2}} \right) \]
\[ = \left[ T_{N}^{\beta_{1}, \beta_{2}} + o_{p} \left( 1 \right) \right] \mathbf{I}_{0, \mathcal{X}_{\alpha}^{2}} \left( T_{N}^{\beta_{1}, \beta_{2}} \right). \]

Therefore
\[ \lim_{N \to \infty} E \left[ N \left\| \tilde{\beta}_{\phi_{1}, \phi_{2}}^{\text{Pre}} - \tilde{\beta}_{\phi_{2}} \right\|_{X^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X}}^{2} \right] \leq \lim_{N \to \infty} E \left[ \left[ X_{1, \alpha}^{2} + o_{p} \left( 1 \right) \right] \mathbf{I}_{0, \mathcal{X}_{\alpha}^{2}} \left( T_{N}^{\beta_{1}, \beta_{2}} \right) \right] = 0 \]
which means
\[ \sqrt{N} \tilde{\beta}_{\phi_{1}, \phi_{2}}^{\text{Pre}} - \sqrt{N} \tilde{\beta}_{\phi_{2}} \xrightarrow[N \to \infty]{q.m.} 0 \]
and \( \sqrt{N} \tilde{\beta}_{\phi_{1}, \phi_{2}}^{\text{Pre}} - \sqrt{N} \tilde{\beta}_{\phi_{2}} \xrightarrow[N \to \infty]{p.} 0. \) Then,
\[ \sqrt{N} \left( \tilde{\beta}_{\phi_{1}, \phi_{2}}^{\text{Pre}} - \beta \right) = \sqrt{N} \left( \tilde{\beta}_{\phi_{2}} - \beta \right) + o_{p} \left( 1 \right). \]

(b) Based on Pardo et al. [5]
\[ \tilde{\beta}_{\phi_{2}} = \beta_{0} + \left( \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X} \right)^{-1} \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X} \text{diag} \left( \mathbf{C}_{\alpha} \left( \beta_{0} \right) \right) \text{diag} \left( \mathbf{p} \left( \beta_{0} \right)^{-1/2} \right) \left( \mathbf{p} - \mathbf{p} \left( \beta_{0} \right) \right) + o_{p} \left( N^{-1/2} \right) \]
and based on Menéndez et al. [3],
\[ \tilde{\beta}_{\phi_{2}}^{\text{H0}} = \beta_{0} + \mathbf{H}_{N} \left( \beta_{0} \right) \left( \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta_{0} \right) \mathbf{X} \right)^{-1} \mathbf{X}^{T} \text{diag} \left( \mathbf{C}_{\alpha} \left( \beta_{0} \right) \right) \text{diag} \left( \mathbf{p} \left( \beta_{0} \right)^{-1/2} \right) \left( \mathbf{p} - \mathbf{p} \left( \beta_{0} \right) \right) + o_{p} \left( N^{-1/2} \right). \]

Therefore,
\[ \tilde{\beta}_{\phi_{2}}^{\text{H0}} - \beta = \tilde{\beta}_{\phi_{2}} - \beta - \left( \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X} \right)^{-1} \mathbf{K} \left( \mathbf{K}^{T} \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X} \right)^{-1} \left( \mathbf{K}^{T} \tilde{\beta}_{\phi_{2}} - \mathbf{m} \right) + o_{p} \left( N^{-1/2} \right). \]

Now taking into account that \( \mathbf{m} = \mathbf{K}^{T} \beta - \mathbf{s} \) we have
\[ \sqrt{N} \left( \tilde{\beta}_{\phi_{2}}^{\text{H0}} - \beta \right) = \sqrt{N} \left( \tilde{\beta}_{\phi_{2}} - \beta \right) - \left( \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X} \right)^{-1} \mathbf{K} \left( \mathbf{K}^{T} \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X} \right)^{-1} \sqrt{N} \left( \mathbf{K}^{T} \tilde{\beta}_{\phi_{2}} - \mathbf{K}^{T} \beta + \mathbf{s} \right) \]
\[ = \sqrt{N} \left( \tilde{\beta}_{\phi_{2}} - \beta \right) - \left( \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X} \right)^{-1} \mathbf{K} \left( \mathbf{K}^{T} \mathbf{X}^{T} \mathbf{W}_{N} \left( \beta \right) \mathbf{X} \right)^{-1} \sqrt{N} \mathbf{K}^{T} \left( \tilde{\beta}_{\phi_{2}} - \beta \right) + \sqrt{N} \mathbf{s}. \]

and the asymptotic distribution of \( \sqrt{N} \left( \tilde{\beta}_{\phi_{2}}^{\text{H0}} - \beta \right) \) is degenerated under the fixed alternative hypotheses \( H_{1} : \mathbf{K}^{T} \beta = \mathbf{m} + \mathbf{s} \).

The result in the previous theorem is important because it reveals that in order to obtain meaningful asymptotic coverage probabilities of the confidence set \( C_{\beta} \left( \tilde{\beta}_{\phi} \right) \) we must consider contiguous alternative hypotheses to \( H_{0} \), i.e., we shall consider hypotheses of the type,
\[ H_{1, N} : \beta_{N} = \beta_{0} + N^{-1/2} \Delta, \]
with \( \beta_{0} \in \Theta_{0} \) and \( \Delta \in \mathbb{R}^{k+1} \).
If we consider the function $g(\beta) = K^T \beta - m$ it is clear that $\theta_0 = \{ \beta \in \Theta : g(\beta) = 0 \}$ and the hypothesis $H_{1,N}$ is equivalent to the hypothesis

$$H_{1,N}^* : \beta_{1,N} = N^{-1/2} \delta(H_{1,N}^* : K^T \beta_{1,N} = m + N^{-1/2} \delta).$$

A Taylor expansion of $g(\beta_{1,N})$ around $\beta_0 \in \Theta_0$ yields

$$g(\beta_{1,N}) = g(\beta_0) + K^T (\beta_{1,N} - \beta_0) + o(1),$$

but $g(\beta_0) = 0$ and $\beta_{1,N} - \beta_0 = N^{-1/2} \Delta$, hence

$$g(\beta_{1,N}) = N^{-1/2} K^T \Delta + o(1).$$

Now if we consider $\delta = K^T \Delta$ we have the equivalence in the limit.

On the other hand, we know that

$$N \| \beta_{1,N} - \hat{\beta}_{\phi}^* \|_{X^T W_{0}(\beta_0) X}^2 - N \| \beta_{1,N} - \hat{\beta}_{\phi}^* \|_{X^T W_{0}(\beta_0) X}^2 \xrightarrow{p} 0.$$

Therefore in order to study the asymptotic behavior of $C_\beta (\hat{\beta}_{\phi}^*)$ we shall consider that our centered confidence sets are given by

$$C_{\beta_{1,N}} (\hat{\beta}_{\phi}^*) = \left\{ \beta_{1,N} : N \| \beta_{1,N} - \hat{\beta}_{\phi}^* \|_{X^T W_{0}(\beta_0) X}^2 \leq X^2_{\alpha + 1} \right\}.$$

We need an auxiliary lemma to obtain the asymptotic coverage probabilities of $C_{\beta_{1,N}} (\hat{\beta}_{\phi_{21}})$ and $C_{\beta_{1,N}} (\hat{\beta}_{\phi_{22}})$.

**Lemma 2.** We denote by $\Gamma = \left( \Gamma_1^T, \Gamma_2^T \right)^T$, $\Gamma_1$ is an $r \times (k + 1)$ matrix and $\Gamma_2$ a $(k + 1 - r) \times (k + 1)$ matrix, the orthogonal matrix that diagonalizes the idempotent matrix

$$(X^T W_{0} (\beta_0) X)^{-1/2} K^T (X^T W_{0} (\beta_0) X)^{-1} K (X^T W_{0} (\beta_0) X)^{-1/2},$$

$\eta_{1,N} = (\eta_{11}, \eta_{12})$ $(\eta_{11})$ is an $r \times 1$ random vector and $\eta_{12}$ a $(k + 1 - r) \times 1$ random vector) the random vector defined as

$$\eta_{1,N} = \sqrt{N} \left( \Gamma (X^T W_{0} (\beta_0) X)^{-1/2} \hat{\beta}_{\phi_{21}} - \Gamma (X^T W_{0} (\beta_0) X)^{-1/2} K^T (X^T W_{0} (\beta_0) X)^{-1} m \right) + o_p(1).$$

Then, we have:

(a) $\eta_{1,N} - E[\eta_{1,N}] \xrightarrow{L \to \infty} N(0, \Gamma)$ where $\Gamma = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$.

(b) $N \| \beta_{1,N} - \hat{\beta}_{\phi_{21}} \|_{X^T W_{0}(\beta_0) X}^2 = \left\| E[\eta_{1,N}] - \eta_{1,N} \right\|_{X^T W_{0}(\beta_0) X}^2 + \left\| E[\eta_{1,N}] - \eta_{1,N} \right\|_{X^T W_{0}(\beta_0) X}^2 + o_p(1)$.

**Proof.** Part (a). Based on the definition of $\eta_{1,N}$, given in (10), we have

$$\eta_{1,N} - E[\eta_{1,N}] = \Gamma (X^T W_{0} (\beta_0) X)^{-1/2} \sqrt{N} (\hat{\beta}_{\phi_{21}} - \beta_{1,N}).$$

Now by (6) we obtain

$$\eta_{1,N} - E[\eta_{1,N}] \xrightarrow{L \to \infty} N(0, \Gamma).$$

Now we consider part (b). The matrix $\Gamma$ verifies

$$\Gamma (X^T W_{0} (\beta_0) X)^{-1/2} K^T (X^T W_{0} (\beta_0) X)^{-1} K (X^T W_{0} (\beta_0) X)^{-1/2} \Gamma = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

In (3) it was obtained,

$$\Gamma_{\phi_{21}, \phi_{21}} = \sqrt{N} (K^T \hat{\beta}_{\phi_{21}} - m)^T (K^T (X^T W_{0} (\beta_0) X)^{-1} K)^{-1} \sqrt{N} (K^T \hat{\beta}_{\phi_{21}} - m) + o_p(1).$$

Now we have,

$$\Gamma_{\phi_{21}, \phi_{21}} = \sqrt{N} (K^T \hat{\beta}_{\phi_{21}} - m)^T (K^T (X^T W_{0} (\beta_0) X)^{-1} K)^{-1} K^T (X^T W_{0} (\beta_0) X)^{-1/2}

\times (X^T W_{0} (\beta_0) X)^{-1/2} K^T (X^T W_{0} (\beta_0) X)^{-1} K^{-1} \sqrt{N} (K^T \hat{\beta}_{\phi_{21}} - m) + o_p(1)$$

$$= \sqrt{N} \left( \Gamma (X^T W_{0} (\beta_0) X)^{-1/2} \hat{\beta}_{\phi_{21}} - \Gamma (X^T W_{0} (\beta_0) X)^{-1/2} K^T (X^T W_{0} (\beta_0) X)^{-1} m \right)^T

\times \Gamma (X^T W_{0} (\beta_0) X)^{-1/2} K^T (X^T W_{0} (\beta_0) X)^{-1} K^{-1} \sqrt{N} (K^T \hat{\beta}_{\phi_{21}} - m) + o_p(1)$$

$$= \Gamma (X^T W_{0} (\beta_0) X)^{-1/2} K^T (X^T W_{0} (\beta_0) X)^{-1} K^{-1} \sqrt{N} (K^T \hat{\beta}_{\phi_{21}} - m) + o_p(1).$$

- \sqrt{N} \Gamma (X^T W_{0} (\beta_0) X)^{-1/2} K^T (X^T W_{0} (\beta_0) X)^{-1} m + o_p(1).$$
Therefore, 
\[ T_N^{1, \phi_2} = \eta_N^T \left( \begin{array}{cc} l_1 & 0 \\ 0 & 0 \end{array} \right) \left( \begin{array}{cc} l_2 & 0 \\ 0 & 0 \end{array} \right) \eta_N + o_p(1) = \eta_1^T \eta_1 + o_p(1) \]
and the asymptotic distribution of \( T_N^{1, \phi_2} \) is a noncentral chi-square with \( r \) degrees of freedom and noncentrality parameter 
\[ \lambda = E[\eta_1^T \eta_1] = \Delta^T K^T (K^T X^T W (\beta_0) X)^{-1} K^{-1} \Delta. \]  
(11)

This result follows since \( \sqrt{N}(\eta_1 - E[\eta_1]) \) converges in law to an \( r \)-normal random vector with mean vector zero and variance covariance matrix \( I \).

Using the definition of \( \hat{\beta}_{\phi_1, \phi_2} \) given in (8), we obtain
\[
N \left| \beta_N - \hat{\beta}_{\phi_1, \phi_2} \right|^2 |X^T W_n(\beta_0) X| = \frac{1}{\sqrt{N}} \left( \begin{array}{c} \hat{\beta}_{\phi_1} - \hat{\beta}_{\phi_2} \\ \hat{\beta}_{\phi_2} - \hat{\beta}_{\phi_2} \end{array} \right) \eta_N^T \left( \begin{array}{c} \hat{\beta}_{\phi_1} - \hat{\beta}_{\phi_2} \\ \hat{\beta}_{\phi_2} - \hat{\beta}_{\phi_2} \end{array} \right) \left( \begin{array}{c} \hat{\beta}_{\phi_1} - \hat{\beta}_{\phi_2} \\ \hat{\beta}_{\phi_2} - \hat{\beta}_{\phi_2} \end{array} \right) \eta_N + o_p(1) 
\]
(12)
In the following theorem we are going to obtain the coverage probabilities of the sets \( C_{\beta_0} (\hat{\beta}_0) \) with \( \hat{\beta}_0 \) equal to \( \hat{\beta}_{H0} \) or \( \hat{\beta}_{\phi_1, \phi_2} \).

**Theorem 3.** We have, under \( H_{1,N} \):

(a) \( \lim_{N \to \infty} \Pr \left( C_{\beta_0} (\hat{\beta}_{\phi_1, \phi_2}) \right) = G_{r} (\chi^2_{k+1, \alpha}; \lambda) G_{k+1-r} (\chi^2_{k+1, \alpha} - \lambda; 0) + \lim_{N \to \infty} \int_0^\infty \Pr (\|E \eta_1\| - \eta_1^2 \leq \chi^2_{k+1, \alpha} - \tau; \|E \eta_1\|^2 > \chi^2_{r, \alpha}) dG_{k+1-r} (\tau; 0) \),

(b) \( \lim_{N \to \infty} \Pr \left( C_{\beta_0} (\hat{\beta}_{H0}) \right) = G_{k+1-r} (\chi^2_{k+1, \alpha} - \lambda; 0) \).

By \( G_{r} (b; \mu) \) we are denoting the distribution function of a noncentral chi-square random variable with noncentrality parameter \( \mu \) and "a" degrees of freedom evaluated at "b".

**Proof.** (a) We denote \( I = \lim_{N \to \infty} \Pr \left( C_{\beta_0} (\hat{\beta}_{\phi_1, \phi_2}) \right) \). We have,

\[
I = \lim_{N \to \infty} \Pr \left( \left( \|E \eta_1\| - \eta_1^2 \leq \chi^2_{k+1, \alpha} \right) \right.
+ \lim_{N \to \infty} \Pr \left( \left( \|E \eta_2\| - \eta_2^2 \leq \chi^2_{k+1, \alpha} \right) \right.
+ \lim_{N \to \infty} \Pr \left( \left( \|E \eta_1\| - \eta_1^2 \leq \chi^2_{k+1, \alpha} \right) \right.
\]

Therefore

\[
\lim_{N \to \infty} \Pr \left( C_{\beta_0} (\hat{\beta}_{\phi_1, \phi_2}) \right) = \lim_{N \to \infty} \Pr \left( \left( \|E \eta_1\| - \eta_1^2 \leq \chi^2_{k+1, \alpha} \right) \right.
+ \lim_{N \to \infty} \Pr \left( \left( \|E \eta_2\| - \eta_2^2 \leq \chi^2_{k+1, \alpha} \right) \right.
\]

Remark 4. We know that under \( H_{1,N} \)

\[
\lim_{N \to \infty} \Pr \left( N \|E \eta_1\|^2 \leq \chi^2_{k+1, \alpha} \right) = 1 - \alpha
\]

and this probability does not depend on \( \lambda \), i.e.,

\[
\lim_{N \to \infty} \Pr \left( C_{\beta_0} (\hat{\beta}_{\phi_1, \phi_2}) \right) = 1 - \alpha.
\]

If we consider

\[
C_{\beta_0} (\hat{\beta}_{\phi_2}) \quad \text{defined by (b) in Theorem 3 that}
\]

we have by (b) in **Theorem 3** that

\[
\lim_{N \to \infty} \Pr \left( C_{\beta_0} (\hat{\beta}_{H0}) \right) = G_{k+1-r} (\chi^2_{k+1, \alpha} - \lambda; 0).
\]

We can observe that \( G_{k+1-r} (\chi^2_{k+1, \alpha} - \lambda; 0) \) is a decreasing function on \( \lambda \). At \( \lambda = 0 \), it attains the maximum value \( G_{k+1-r} (\chi^2_{k+1, \alpha}; 0) \) and it tends to zero as \( \lambda \to \chi^2_{k+1, \alpha} \). The coverage probabilities of \( C_{\beta} (\hat{\beta}_{\phi_2}) \) and \( C_{\beta} (\hat{\beta}_{H0}) \) are equal if \( \lambda = \chi^2_{k+1, \alpha} - C_{k+1-r} (1 - \alpha; 0) \).

The asymptotic coverage probability of \( C_{\beta_0} (\hat{\beta}_{\phi_1, \phi_2}) \) depends on the noncentrality parameter \( \lambda \) in the following way:

**Theorem 5.** The following results hold:

(i) If \( 0 \leq \lambda < \chi^2_{k+1, \alpha} \), then

\[
\lim_{N \to \infty} \Pr \left( C_{\beta_0} (\hat{\beta}_{\phi_1, \phi_2}) \right) \geq 1 - \alpha.
\]
(ii) If \( \chi_{k+1, \alpha}^2 \leq \lambda \leq \left( (\chi_{k+1, \alpha}^2 + (\chi_{t, \alpha}^2)^{1/2}) \right)^2 \), then

\[
\lim_{N \to \infty} \Pr \left( C_{\beta_{\text{Pre}}} (\hat{\beta}_{\psi_1, \phi_2}) \right) \leq 1 - \alpha.
\]

(iii) If \( \lambda > \left( (\chi_{k+1, \alpha}^2 + (\chi_{t, \alpha}^2)^{1/2}) \right)^2 \), then

\[
\lim_{N \to \infty} \Pr \left( C_{\beta_{\text{Pre}}} (\hat{\beta}_{\psi_1, \phi_2}) \right) = 1 - \alpha.
\]

**Proof.** (i) We assume \( \lambda < \chi_{k+1, \alpha}^2 \). We denote

\[
l = \lim_{N \to \infty} \Pr \left( C_{\beta_{\text{Pre}}} (\hat{\beta}_{\psi_1, \phi_2}) \right),
\]

we have

\[
l = \lim_{N \to \infty} \Pr \left( \left\| E[\eta_2] - \eta_2 \right\|^2 + \lambda \leq \chi_{k+1, \alpha}^2; \left\| \eta_1 \right\|^2 \leq \chi_{t, \alpha}^2 \right)
\]

\[
+ \lim_{N \to \infty} \int_{0}^{\chi_{k+1, \alpha}^2} \Pr \left( \left\| E[\eta_1] - \eta_1 \right\|^2 \leq \chi_{k+1, \alpha}^2 - t; \left\| \eta_1 \right\|^2 > \chi_{t, \alpha}^2 \right) dG_{k+1, \alpha}(t; 0)
\]

\[
= \lim_{N \to \infty} \Pr \left( \left\| E[\eta_1] - \eta_1 \right\|^2 \leq \chi_{k+1, \alpha}^2 - t; \left\| \eta_1 \right\|^2 > \chi_{t, \alpha}^2 \right) dG_{k+1, \alpha}(t; 0)
\]

\[
+ \lim_{N \to \infty} \Pr \left( \left\| E[\eta_1] - \eta_1 \right\|^2 \leq \chi_{k+1, \alpha}^2 - t; \left\| \eta_1 \right\|^2 > \chi_{t, \alpha}^2 \right) dG_{k+1, \alpha}(t; 0)
\]

\[
= \lim_{N \to \infty} \Pr \left( \left\| E[\eta_1] - \eta_1 \right\|^2 + \left\| E[\eta_2] - \eta_2 \right\|^2 \leq \chi_{k+1, \alpha}^2 \right)
\]

\[
= \lim_{N \to \infty} \Pr \left( \eta_1 \leq \chi_{k+1, \alpha}^2 \right) = \Pr \left( \chi_{k+1} \leq \chi_{k+1, \alpha}^2 \right) = 1 - \alpha.
\]

(ii) We assume \( \chi_{k+1, \alpha}^2 \leq \lambda \leq \left( (\chi_{k+1, \alpha}^2 + (\chi_{t, \alpha}^2)^{1/2}) \right)^2 \). On the other hand we have established before that

\[
l = G_r(\chi_{k+1, \alpha}^2; \lambda)G_{k+1, \alpha} - \lambda; 0) + \lim_{N \to \infty} \int_{0}^{\chi_{k+1, \alpha}^2} \Pr \left( \left\| E[\eta_1] - \eta_1 \right\|^2 \leq \chi_{k+1, \alpha}^2 - t; \left\| \eta_1 \right\|^2 > \chi_{t, \alpha}^2 \right) dG_{k+1, \alpha}(t; 0).
\]

where \( l \) was defined in (13). But if \( \lambda \geq \chi_{k+1, \alpha}^2 \) then \( \lambda \geq \chi_{k+1, \alpha}^2 \geq \chi_{t, \alpha}^2 \), hence \( \chi_{t, \alpha}^2 - \lambda \leq 0 \) and \( G_{k+1, \alpha}(\chi_{k+1, \alpha}^2 - \lambda; 0) = 0 \). Therefore

\[
l = \lim_{N \to \infty} \Pr \left( \left\| E[\eta_1] - \eta_1 \right\|^2 \leq \chi_{k+1, \alpha}^2 - t; \left\| \eta_1 \right\|^2 > \chi_{t, \alpha}^2 \right) dG_{k+1, \alpha}(t; 0)
\]

\[
= \lim_{N \to \infty} \Pr \left( \left\| E[\eta_1] - \eta_1 \right\|^2 \leq \chi_{k+1, \alpha}^2 - t \right) dG_{k+1, \alpha}(t; 0)
\]

\[
= \Pr \left( \chi_{k+1}^2 \leq \chi_{k+1, \alpha}^2 \right) = 1 - \alpha.
\]

(iii) If \( \lambda > \left( (\chi_{k+1, \alpha}^2 + (\chi_{t, \alpha}^2)^{1/2}) \right)^2 \), then \( G_{k+1, \alpha}(\chi_{k+1, \alpha}^2 - \lambda; 0) = 0 \), hence

\[
l = \lim_{N \to \infty} \int_{0}^{\chi_{k+1, \alpha}^2} \Pr \left( \left\| E[\eta_1] - \eta_1 \right\|^2 \leq \chi_{k+1, \alpha}^2 - t; \left\| \eta_1 \right\|^2 > \chi_{t, \alpha}^2 \right) dG_{k+1, \alpha}(t; 0).
\]

On the other hand if \( \left( (\chi_{k+1, \alpha}^2 + (\chi_{t, \alpha}^2)^{1/2}) \right)^2 < \lambda \) then \( (\chi_{k+1, \alpha}^2)^{1/2} + (\chi_{t, \alpha}^2)^{1/2} < \lambda^{1/2} \) and further

\[
\left\| E[\eta_1] - \eta_1 \right\|^2 \leq \chi_{k+1, \alpha}^2 - t \implies \left\| E[\eta_1] - \eta_1 \right\| \leq \sqrt{\chi_{k+1, \alpha}^2 - t},
\]

then

\[
\left\| E[\eta_1] - \eta_1 \right\| = \lambda^{1/2} - \left\| \eta_1 \right\| \leq \left\| \eta_1 - E[\eta_1] \right\| \implies \lambda^{1/2} \leq \left\| \eta_1 \right\| + \sqrt{\chi_{k+1, \alpha}^2 - t}.
\]

Since

\[
\sqrt{\chi_{k+1, \alpha}^2 - t} + (\chi_{t, \alpha}^2)^{1/2} \leq (\chi_{k+1, \alpha}^2)^{1/2} + (\chi_{t, \alpha}^2)^{1/2} < \lambda^{1/2} \leq \left\| \eta_1 \right\| + \sqrt{\chi_{k+1, \alpha}^2 - t}
\]
hence \( \| \eta_1 \| > (\frac{\chi^2_{\alpha}}{n})^{1/2} \rightarrow \| \eta_1 \|^2 > \chi^2_{\alpha} \).

Therefore

\[
I = \lim_{N \to \infty} \frac{1}{2} \int_0^{\chi^2_{1,\alpha}} \Pr \left( \| E \eta_1 - \eta_1 \|^2 \leq \chi^2_{k+1,\alpha} - t; \| \eta_1 \|^2 > \chi^2_{\alpha} \right) dG_{k+1-t}(t; 0)
\]

\[
= \lim_{N \to \infty} \frac{1}{2} \int_0^{\chi^2_{1,\alpha}} \Pr \left( \| E \eta_1 - \eta_1 \|^2 \leq \chi^2_{k+1,\alpha} - t \right) dG_{k+1-t}(t; 0)
\]

\[
= \Pr (\chi^2_{k+1} \leq \chi^2_{1,\alpha}) = 1 - \alpha. \]

4. Simulation results

We study the coverage probability (CP) of the confidence sets based on preliminary minimum \((\phi_1, \phi_2)\)-divergence test estimators, \(\hat{\beta}^{\text{pre}}_{\phi_1, \phi_2}\), under the null hypothesis as well as under contiguous alternative hypotheses using Monte Carlo experiments. Our idea is to check the advantage of using the minimum \(\phi\)-divergence estimators instead of the classical likelihood-ratio test or Pearson test statistic. In our study we shall consider the power divergence measures introduced and studied in [2], the expression of the function associated with this family of divergence measures is

\[
\phi_2(x) = \begin{cases} 
\frac{x^\lambda + 1 - x}{\lambda + 1}, & \lambda \neq 0, -1 \\
\log x - x + 1, & \lambda = 0 \\
\log x + x - 1, & \lambda = -1.
\end{cases}
\]

This family will be used for testing and estimating. That it is to say, we consider for our study the family of preliminary test estimators

\[
\hat{\beta}^{\text{pre}}_{\phi_1, \phi_2} = \hat{\beta}^{i_0}_{\phi_2} l(x, \chi^2_{\alpha}) (T_{n, \phi_1, \phi_2}) + \hat{\beta}^{i_0}_{\phi_2} l(x, \chi^2_{\alpha}) (T_{n, \phi_1, \phi_2}),
\]

for some choices of the parameters \(\lambda_1\) and \(\lambda_2\). More concretely we shall use \(\lambda_1 = -1/2, 0, 2/3, 1\) and \(\lambda_2 = 0, 2/3\) and 1. It is interesting to note that for \(\lambda_2 = 0\), \(\hat{\beta}^{\text{pre}}_{\phi_1, \phi_2}\) and \(\hat{\beta}^{i_0}_{\phi_2}\) are the unrestricted and restricted MLE of \(\beta\) respectively. Note that

\[
T_{n, \phi_1, \phi_2} = LR + \phi_0(1), \text{ where LR is the likelihood-ratio test.}
\]

The logistic regression model considered in the simulation study consists of a dichotomous dependent variable and three normally distributed with zero mean and unit variance explanatory variables. We generated 10 000 samples of different sample sizes \(n = (n_1, \ldots, n_3)^T \in N = \{n^1, n^2, n^3, n^4, n^5\} \) with \(n^1 = 15, n^2 = 20, n^3 = 80, i = 1, \ldots, 8, n^4 = (25, 25, 25, 25, 10, 10, 10, 10)\) and \(n^5 = (40, 40, 15, 15, 5, 5, 25, 25)\). The regression coefficients \(\beta^i = (\beta_0, \beta_1, \beta_2, \beta_3)\) were generated from a uniform over \((0, 2)\).

We analyze the CP under the null hypothesis \(H_1: \beta_N = \beta + N^{-1/2} \Delta\), with \(\beta \in \Theta_0\) and different values of \(\Delta_1 = (0, 0, 0, 30), \Delta_2 = (0, 0, 0, 20), \Delta_3 = (0, 0, 0, -20)\) and \(\Delta_4 = (0, 0, 0, -30)\). We present the results obtained in Tables 1–5.

| Table 1 |
|------------------|------------------|------------------|------------------|------------------|------------------|
| \(\lambda_1\)   | \(\lambda_2\)   | \(n^1\)          | \(n^2\)          | \(n^3\)          | \(n^4\)          | \(n^5\)          |
| 0                | -1/2             | 0.9734           | 0.9670           | 0.9606           | 0.9757           | 0.9720           |
| 0                | 0                | 0.9658           | 0.9670           | 0.9655           | 0.9692           | 0.9651           |
| 0                | 2/3              | 0.9484           | 0.9551           | 0.9578           | 0.9437           | 0.9421           |
| 0                | 1                | 0.9344           | 0.9440           | 0.9512           | 0.9259           | 0.9299           |
| 0                | 2                | 0.9022           | 0.9015           | 0.9238           | 0.8790           | 0.8874           |
| 2/3              | -1/2             | 0.9738           | 0.9673           | 0.9607           | 0.9760           | 0.9714           |
| 2/3              | 0                | 0.9647           | 0.9661           | 0.9658           | 0.9684           | 0.9628           |
| 2/3              | 1                | 0.9461           | 0.9538           | 0.9571           | 0.9406           | 0.9396           |
| 2/3              | 2                | 0.9312           | 0.9426           | 0.9506           | 0.9226           | 0.9260           |
| 2                | 0.8890           | 0.8980           | 0.9227           | 0.8754           | 0.8817           |
| 2                | -1/2             | 0.9739           | 0.9672           | 0.9612           | 0.9759           | 0.9711           |
| 2                | 0                | 0.9647           | 0.9663           | 0.9659           | 0.9677           | 0.9627           |
| 2                | 2/3              | 0.9446           | 0.9530           | 0.9568           | 0.9391           | 0.9382           |
| 2                | 1                | 0.9304           | 0.9417           | 0.9502           | 0.9219           | 0.9249           |
| 2                | 2                | 0.8975           | 0.8972           | 0.9218           | 0.8738           | 0.8802           |
From Tables 2 and 3 that correspond with $\Delta_1$, $\Delta_2$ it is clear that $\hat{\beta}_{0,-1/2}$ is preferred to the rest. For $\Delta = 0$, this estimator is the first or second best. However, for $\Delta_3$, $\Delta_4$ it can be seen from Tables 4 and 5 that $\hat{\beta}_{1,2}$ is preferred to the rest. Therefore, $\hat{\beta}_{2/3,2/3}$ can be considered as a good compromise for all the cases. Note that if we want to use the LRT ($\lambda_1 = 0$) statistic for the preliminary estimator, the largest CP corresponds to $\lambda_2 = -1/2$ for $\Delta = 0$, $\Delta_1$, $\Delta_2$ and $\lambda_2 = 2$ for $\Delta_3$ and $\Delta_4$. So, $\hat{\beta}_{0,2/3}$ is a good compromise between these two. On the other hand, we can fix the MLE ($\lambda_2 = 0$) for obtaining the preliminary...
### Table 5
CP of the estimates for $\Delta = \Delta_4$

<table>
<thead>
<tr>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$n^1$</th>
<th>$n^2$</th>
<th>$n^3$</th>
<th>$n^4$</th>
<th>$n^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$-1/2$</td>
<td>0.9613</td>
<td>0.9473</td>
<td>0.9353</td>
<td>0.9552</td>
<td>0.9456</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.9637</td>
<td>0.9581</td>
<td>0.9472</td>
<td>0.9549</td>
<td>0.9605</td>
</tr>
<tr>
<td>2/3</td>
<td>0</td>
<td>0.9670</td>
<td>0.9606</td>
<td>0.9486</td>
<td>0.9608</td>
<td>0.9660</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.9688</td>
<td>0.9605</td>
<td>0.9469</td>
<td>0.9616</td>
<td>0.9669</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.9708</td>
<td>0.9585</td>
<td>0.9402</td>
<td>0.9625</td>
<td>0.9665</td>
</tr>
<tr>
<td>2/3</td>
<td>$-1/2$</td>
<td>0.9618</td>
<td>0.9483</td>
<td>0.9385</td>
<td>0.9575</td>
<td>0.9482</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.9642</td>
<td>0.9588</td>
<td>0.9488</td>
<td>0.9579</td>
<td>0.9616</td>
</tr>
<tr>
<td>2/3</td>
<td>2/3</td>
<td>0.9674</td>
<td>0.9610</td>
<td>0.9493</td>
<td>0.9621</td>
<td>0.9664</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.9692</td>
<td>0.9609</td>
<td>0.9479</td>
<td>0.9631</td>
<td>0.9671</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.9711</td>
<td>0.9589</td>
<td>0.9407</td>
<td>0.9635</td>
<td>0.9669</td>
</tr>
</tbody>
</table>

estimator and to look for the best statistic. In this case, for $\Delta = 0$, $\Delta_1$, $\Delta_3$ LRT is the best but for $\Delta_2$, $\Delta_4$ the minimum chi-square statistic is the best, so a good compromise for all $\Delta$ seems to be the statistic corresponding with $\lambda_1 = 2/3$.

### Acknowledgements

The research in this paper was supported in part by Grants MTM2006-06872 and UCM2007-910707.

### References