Fast QLB algorithm and hypothesis tests in logistic model for ophthalmologic bilateral correlated data

Yi-Qi LIN[a,†], **Yu-Shun ZHANG**[b,†], **Guo-Liang TIAN**[b,*] and **Chang-Xing MA**[c]

[a]*Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, P. R. China*

[b]*Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen 518055, Guangdong Province, P. R. China*

[c]*Department of Biostatistics, The State University of New York at Buffalo, Buffalo, NY 14214, USA*

[†]Equally contributed
[*]Corresponding author's email: tiangl@sustech.edu.cn

Abstract. In ophthalmologic or otolaryngologic studies, bilateral correlated data often arise when observations involving paired organs (e.g., eyes, ears) are measured from each subject. Based on the model of Donner (1989), in this paper, we focus on investigating the relationship between the disease probability and covariates (such as ages, weights, gender and so on) via the logistic regression for the analysis of bilateral correlated data. We first propose a new *minorization–maximization* (MM) algorithm and a fast *quadratic lower bound* (QLB) algorithm to calculate the maximum likelihood estimates of the vector of regression coefficients, and then develop three large-sample tests (i.e., the likelihood ratio test, Wald test and score test) to test if covariates have a significant impact on the disease probability. Simulation studies are conducted to evaluate the performance of the proposed fast QLB algorithm and three testing methods. A real ophthalmologic data set in Iran is used to illustrate the proposed methods.

Keywords: Assembly and decomposition technique; Bilateral correlated data; Fast QLB algorithm; Logistic regression model; MM algorithm; Ophthalmologic study.

1. Introduction

In ophthalmologic (or otolaryngologic) studies, data collection not only includes the information from each subject but also contains the observations from each of two eyes (or two ears). Although the data from different subjects are independent, the observations from the two eyes (or ears) are generally positively correlated. Therefore, any statistical models and methods that ignore this feature of dependency could lead to incorrect conclusions (Rosner 1982, Dallal 1988, Donner & Banting 1988, Rosner & Milton 1988, Donner 1989, Bodian 1994, Tang et al 2008).

Rosner (1982) proposed a "constant R" model, which considers the intra-class correlation between the two eyes of the same subject.A basic assumption in Rosner's model is that the probability of disease in one eye given disease in the other eye is proportional to the prevalence of disease in the population. Donner (1989) further proposed an alternative model based on the assumption of constant intra-cluster correlation for all individuals in the sample, and he used a simple adjustment of the standard Pearson chi-square test for testing whether the proportion of affected eyes is the same among all groups of patients. Comparing Rosner's with Donner's model, they are based on different assumptions, however, in most cases, Donner's model is more practical, so we choose to use it in our article. More details can be seen in the paper of Donner(1989).

Ma & Liu (2017) further proposed three testing procedures for testing the equality of proportions for correlated binary data under Donner's model, however, they tested the difference among different groups by simply assuming different π's for different groups, which was not a flexible framework to link the difference to general covariates.

To our best knowledge, there are few papers exploring the relationship between the disease probability/rate and some covariates (e.g., age, weight and so on) under Donner's model. For example, from the last column of Table 7 in Ma & Liu (2017), we found a possible trend that the disease probabilities were increasing with the ages of the patient groups. Although Ma & Liu (2017) performed several hypothesis tests for the ophthalmologic data in Iran (Rajavi et al 2011) and found out that the disease rates in different age groups are different,

they did not reveal the exact relationship between ages and disease rates. Therefore, in this article, based on Donner (1989)'s model, we focus on investigating the relationship between the disease probability and covariates via the logistic regression for the analysis of bilateral correlated data in ophthalmologic studies. The method of Ma & Liu (2017) can be used to test the difference among different groups, while our methods can be used to test if covariates have a significant impact on the disease probability.

To calculate the *maximum likelihood estimates* (MLEs) of the vector of regression coefficients $\boldsymbol{\beta}$, we have noted that the well-known Newton–Raphson (NR) algorithm and the Fisher scoring algorithms are not available for the following reasons: (i) the log-likelihood function $\ell(\boldsymbol{\beta}, \rho)$ in (2.6) is not concave with respect to $\boldsymbol{\beta}$, and it is too complicated; (ii) it requires tedious calculations of the inverse Hessian matrix at each iteration, especially when sample size n is quite large; (iii) it is sensitive to the initial values, it does not converge if a poor initial value is chosen; (iv) the likelihood function does not necessarily increase at each iteration for the NR algorithm; (v) in our simulation experiments, instead of converging to the optimal point, it converges to completely different points given many different random initial values. To overcome the above difficulties, in Section 2 we shall utilize the *assembly and decomposition* (AD) approach of Tian, Huang & Xu (2018) to propose a new *minorization–maximization* (MM) algorithm and a fast *quadratic lower bound* (QLB) algorithm.

The rest of the paper is organized as follows. In section 3, we develop three large-sample tests (i.e., the likelihood ratio test, Wald test and score test) for the logistic regression model to test if covariates have a significant impact on the disease probability. Simulation studies are conducted to evaluate the performance of the proposed fast QLB algorithm and three testing methods in Section 4. In addition, we compare the proposed three tests to the test based on *generalized estimating equations* (GEE) method with the logit link function. The simulation results show that the test based on the GEE method has similar type I error rates, but lower powers compared to the proposed three tests. In Section 5, a real ophthalmologic data set in Iran is used to illustrate the proposed methods. Some discussions are given in Section 6. Two technical details are put in the Appendix.

2. Modeling bilateral correlated data by a logistic regression model

2.1 Model formulation

Suppose that we want to compare n different individuals on some ocular disease. Let $Z_{ik} = 1$ if the k-th eye of the i-th individual is diseased, and $Z_{ik} = 0$ otherwise for $i = 1, \ldots, n$ and $k = 1, 2$. The parametric model proposed by Donner (1989) is

$$\begin{cases} \Pr(Z_{ik} = 1) &= \pi_{ik}, \quad i = 1, \ldots, n, \quad k = 1, 2, \\ & \pi_{i1} = \pi_{i2} = \pi_i, \quad 0 \leqslant \pi_{ik} \leqslant 1, \\ \mathrm{Corr}(Z_{i1}, Z_{i2}) &= \rho, \quad 0 \leqslant \rho \leqslant 1, \end{cases} \tag{2.1}$$

where π_i denotes the disease probability for the i-th individual and ρ is assumed to be the common correlation coefficient between two eyes of the same individual across all patients. Note that the disease rates are assumed to be the same for both eyes within the same individual. The rationale of this assumption is that the disease rate π_i will be determined only by the number of eyes affected within the same individual, without differentiating whether it is the left or right eye. For more remarks on this model assumption, please see Ma & Liu (2017).

It is easy to derive that the disease probabilities for none, one or both eyes are

$$p_{i0} = (1 - \pi_i)(\rho\pi_i - \pi_i + 1), \; p_{i1} = 2\pi_i(1 - \rho)(1 - \pi_i) \;\text{ and }\; p_{i2} = \pi_i^2 + \rho\pi_i(1 - \pi_i), \tag{2.2}$$

respectively. We define $Y_i = 0, 1, 2$ if the i-th individual has none, one or both defective eyes. The log-likelihood function for the unknown parameters $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)^\top$ and ρ is

$$\begin{aligned} & \ell(\boldsymbol{\pi}, \rho) \\ = \; & \sum_{i=1}^{n} \Big\{ I(Y_i = 0) \log\big[(1 - \pi_i)(\rho\pi_i - \pi_i + 1)\big] + I(Y_i = 1) \log\big[2\pi_i(1 - \rho)(1 - \pi_i)\big] \\ & + I(Y_i = 2) \log\big[\pi_i^2 + \rho\pi_i(1 - \pi_i)\big] \Big\}, \end{aligned} \tag{2.3}$$

where $I(\cdot)$ denotes the indicator function.

Note that $\{\pi_i\}_{i=1}^n$ are assumed to be different since each patient has different covariates such as age, weight and so on. Let $\boldsymbol{x}_{(i)} = (1, x_{i1}, \ldots, x_{i,q-1})^\top$ be a vector of covariates of interest for $i = 1, \ldots, n$. To link the disease probability π_i with these covariates, we consider the following logistic model:

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}, \tag{2.4}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{q-1})^\top$ is the vector of unknown regression coefficients. From (2.4), it is easy to obtain

$$\pi_i = \frac{\mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}}{1+\mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}} = \frac{1}{1+\mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}} \quad \text{and} \quad 1-\pi_i = \frac{1}{1+\mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}}, \tag{2.5}$$

so that the log-likelihood function (2.3) can be rewritten as

$$\begin{aligned}
\ell(\boldsymbol{\beta}, \rho) \;=\; & \sum_{i=1}^n \left\{ I(Y_i = 0) \log\left[\frac{1}{1+\mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}}\left(\frac{\rho-1}{1+\mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}}+1\right)\right] \right. \\
& + I(Y_i = 1) \log\left[\frac{2(1-\rho)}{1+\mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}} \cdot \frac{1}{1+\mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}}\right] \\
& \left. + I(Y_i = 2) \log\left[\left(\frac{1}{1+\mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}}\right)^2 + \frac{\rho}{1+\mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}} \cdot \frac{1}{1+\mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}}\right] \right\}.
\end{aligned} \tag{2.6}$$

2.2 MLEs of parameters via a new MM algorithm

Given $\boldsymbol{\pi}$, the MLE of ρ can be obtained by solving the following equation

$$0 = \frac{\partial \ell(\boldsymbol{\pi}, \rho)}{\partial \rho} = \sum_{i=1}^n \left[\frac{\pi_i I(Y_i = 0)}{\pi_i(\rho-1)+1} + \frac{I(Y_i = 1)}{\rho - 1} + \frac{(1-\pi_i) I(Y_i = 2)}{\pi_i + \rho(1-\pi_i)}\right].$$

That is, given $\boldsymbol{\pi} = \boldsymbol{\pi}^{(t)}$ and the t-th approximation $\rho^{(t)}$, the $(t+1)$-th approximation of the MLE $\hat{\rho}$ can be calculated via the Newton–Raphson iteration:

$$\rho^{(t+1)} = \rho^{(t)} - \left[\frac{\partial^2 \ell(\boldsymbol{\pi}^{(t)}, \rho^{(t)})}{\partial \rho^2}\right]^{-1} \frac{\partial \ell(\boldsymbol{\pi}^{(t)}, \rho^{(t)})}{\partial \rho}. \tag{2.7}$$

Given ρ, to calculate the MLEs of $\boldsymbol{\beta}$ in (2.6), we have noted that the well-known NR algorithm and the Fisher scoring algorithms are not available for possible four reasons stated in Section 1. To overcome those difficulties, we utilize the AD approach of Tian, Huang & Xu

(2018) to propose a new MM algorithm by constructing a surrogate function $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho)$, which minorizes $\ell(\boldsymbol{\beta}, \rho)$ in (2.6). To this end, we first introduce the following important inequality

$$h(\boldsymbol{\alpha}^\top \boldsymbol{z}) \geqslant \sum_{i=1}^{n} \frac{\alpha_i z_i^{(t)}}{\boldsymbol{\alpha}^\top \boldsymbol{z}^{(t)}} h\left(\frac{\boldsymbol{\alpha}^\top \boldsymbol{z}^{(t)}}{z_i^{(t)}} z_i\right), \tag{2.8}$$

where $h(\cdot)$ is an arbitrary concave function, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^\top$, $\boldsymbol{z} = (z_1, \ldots, z_n)^\top$ and $\boldsymbol{z}^{(t)} = (z_1^{(t)}, \ldots, z_n^{(t)})^\top$ are three positive vectors. The inequality (2.8) is just a variant of the discrete version of Jensen's inequality: $h\left(\sum_{i=1}^n p_i z_i\right) \geqslant \sum_{i=1}^n p_i h(z_i)$, where $\{p_i\}_{i=1}^n$ are probability weights.

Let $\boldsymbol{\alpha}_1 = (1, \rho)^\top$ and $\boldsymbol{z} = (\pi_i, 1-\pi_i)^\top$, $h(\cdot) = \log(\cdot)$, we obtain

$$\begin{aligned}
&\log[\pi_i + \rho(1-\pi_i)] = \log(\boldsymbol{\alpha}_1^\top \boldsymbol{z}) \\
\overset{(2.8)}{\geqslant}\ & \frac{\pi_i^{(t)}}{\pi_i^{(t)} + \rho(1-\pi_i^{(t)})} \log\left[\frac{\pi_i^{(t)} + \rho(1-\pi_i^{(t)})}{\pi_i^{(t)}} \pi_i\right] \\
&+ \frac{\rho(1-\pi_i^{(t)})}{\pi_i^{(t)} + \rho(1-\pi_i^{(t)})} \log\left[\frac{\pi_i^{(t)} + \rho(1-\pi_i^{(t)})}{1-\pi_i^{(t)}} (1-\pi_i)\right].
\end{aligned} \tag{2.9}$$

Similarly, let $\boldsymbol{\alpha}_2 = (\rho, 1)^\top$, we have

$$\begin{aligned}
&\log[\rho\pi_i + (1-\pi_i)] = \log(\boldsymbol{\alpha}_2^\top \boldsymbol{z}) \\
\overset{(2.8)}{\geqslant}\ & \frac{\rho\pi_i^{(t)}}{\rho\pi_i^{(t)} + (1-\pi_i^{(t)})} \log\left[\frac{\rho\pi_i^{(t)} + (1-\pi_i^{(t)})}{\pi_i^{(t)}} \pi_i\right] \\
&+ \frac{(1-\pi_i^{(t)})}{\rho\pi_i^{(t)} + (1-\pi_i^{(t)})} \log\left[\frac{\rho\pi_i^{(t)} + (1-\pi_i^{(t)})}{1-\pi_i^{(t)}} (1-\pi_i)\right].
\end{aligned} \tag{2.10}$$

For the purpose of brevity, we define

$$A_i = I(Y_i = 1 \text{ or } 2), \quad B_i = I(Y_i = 0 \text{ or } 1), \quad C_i = I(Y_i = 2) \quad \text{and} \quad D_i = I(Y_i = 0),$$

so that we can re-express (2.3) as

$$
\begin{aligned}
\ell(\boldsymbol{\pi},\rho) \quad &= \quad \sum_{i=1}^{n}\Big\{A_i\log(\pi_i)+B_i\log(1-\pi_i)+C_i\log[\pi_i+\rho(1-\pi_i)] \\
&\quad + D_i\log[\rho\pi_i+(1-\pi_i)]+c_{i1}(\rho)\Big\} \\
&\overset{(2.9)\,\&\,(2.10)}{\geqslant} \quad \sum_{i=1}^{n}\Bigg\{A_i\log(\pi_i)+B_i\log(1-\pi_i) \\
&\quad + \frac{C_i\pi_i^{(t)}}{\pi_i^{(t)}+\rho(1-\pi_i^{(t)})}\log\left[\frac{\pi_i^{(t)}+\rho(1-\pi_i^{(t)})}{\pi_i^{(t)}}\pi_i\right] \\
&\quad + \frac{C_i\rho(1-\pi_i^{(t)})}{\pi_i^{(t)}+\rho(1-\pi_i^{(t)})}\log\left[\frac{\pi_i^{(t)}+\rho(1-\pi_i^{(t)})}{1-\pi_i^{(t)}}(1-\pi_i)\right] \\
&\quad + \frac{D_i\rho\pi_i^{(t)}}{\rho\pi_i^{(t)}+(1-\pi_i^{(t)})}\log\left[\frac{\rho\pi_i^{(t)}+(1-\pi_i^{(t)})}{\pi_i^{(t)}}\pi_i\right] \\
&\quad + \frac{D_i(1-\pi_i^{(t)})}{\rho\pi_i^{(t)}+(1-\pi_i^{(t)})}\log\left[\frac{\rho\pi_i^{(t)}+(1-\pi_i^{(t)})}{1-\pi_i^{(t)}}(1-\pi_i)\right]+c_{i1}(\rho)\Bigg\} \\
&= \quad \sum_{i=1}^{n}\Bigg\{\left[A_i+\frac{C_i\pi_i^{(t)}}{\pi_i^{(t)}+\rho(1-\pi_i^{(t)})}+\frac{D_i\rho\pi_i^{(t)}}{\rho\pi_i^{(t)}+(1-\pi_i^{(t)})}\right]\log(\pi_i) \\
&\quad + \left[B_i+\frac{C_i\rho(1-\pi_i^{(t)})}{\pi_i^{(t)}+\rho(1-\pi_i^{(t)})}+\frac{D_i(1-\pi_i^{(t)})}{\rho\pi_i^{(t)}+(1-\pi_i^{(t)})}\right]\log(1-\pi_i)+c_{i2}^{(t)}(\rho)\Bigg\} \\
&\hat{=} \quad Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)},\rho),
\end{aligned}
$$

where ρ is treated as a fixing constant, $\pi_i^{(t)}$ denotes the t-th approximation of π_i and $\{c_{i1}(\rho), c_{i2}^{(t)}(\rho)\}$ are functions of ρ but not depending on π_i for $i=1,\ldots,n$. Let

$$
\begin{aligned}
a_i(\pi_i^{(t)},\rho) &= A_i+\frac{C_i\pi_i^{(t)}}{\pi_i^{(t)}+\rho(1-\pi_i^{(t)})}+\frac{D_i\rho\pi_i^{(t)}}{\rho\pi_i^{(t)}+(1-\pi_i^{(t)})} \quad \text{and} \\
b_i(\pi_i^{(t)},\rho) &= B_i+\frac{C_i\rho(1-\pi_i^{(t)})}{\pi_i^{(t)}+\rho(1-\pi_i^{(t)})}+\frac{D_i(1-\pi_i^{(t)})}{\rho\pi_i^{(t)}+(1-\pi_i^{(t)})}, \quad i=1,\ldots,n,
\end{aligned}
$$

we have

$$
Q(\boldsymbol{\pi}|\boldsymbol{\pi}^{(t)},\rho)=\sum_{i=1}^{n}\Big\{a_i(\pi_i^{(t)},\rho)\log(\pi_i)+b_i(\pi_i^{(t)},\rho)\log(1-\pi_i)+c_{i2}^{(t)}(\rho)\Big\}. \tag{2.11}
$$

Furthermore, we have noted that

$$\begin{aligned} & a_i(\pi_i^{(t)}, \rho) + b_i(\pi_i^{(t)}, \rho) \\ = \;& A_i + B_i + C_i + D_i = 2[I(Y_i = 0) + I(Y_i = 1) + I(Y_i = 2)] = 2. \end{aligned} \tag{2.12}$$

From (2.5), by substituting

$$\pi_i = \frac{\mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}}{1 + \mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}} \quad \text{and} \quad \pi_i^{(t)} = \frac{\mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}^{(t)}}}{1 + \mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}^{(t)}}}$$

into (2.11), we obtain

$$\begin{aligned} Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho) &= \sum_{i=1}^{n} \left\{ a_i(\pi_i^{(t)}, \rho) \cdot \boldsymbol{x}_{(i)}^\top \boldsymbol{\beta} - \left[a_i(\pi_i^{(t)}, \rho) + b_i(\pi_i^{(t)}, \rho) \right] \log(1 + \mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}) + c_{i2}^{(t)}(\rho) \right\} \\ &\overset{(2.12)}{=} \sum_{i=1}^{n} \left\{ a_i(\pi_i^{(t)}, \rho) \cdot \boldsymbol{x}_{(i)}^\top \boldsymbol{\beta} - 2 \log(1 + \mathrm{e}^{\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}) + c_{i2}^{(t)}(\rho) \right\}. \end{aligned} \tag{2.13}$$

Obviously, $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho)$ satisfies the following conditions:

$$\begin{aligned} Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho) &\leqslant \ell(\boldsymbol{\beta}, \rho), \qquad \forall\, \boldsymbol{\beta}, \boldsymbol{\beta}^{(t)} \in \mathbb{R}^q \quad \text{and} \\ Q(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}, \rho) &= \ell(\boldsymbol{\beta}^{(t)}, \rho). \end{aligned}$$

That is, $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho)$ minorizes $\ell(\boldsymbol{\beta}, \rho)$ at $\boldsymbol{\beta} = \boldsymbol{\beta}^{(t)}$. By the MM principle (Becker *et al.*, 1997; Lange *et al.*, 2000; Hunter & Lange, 2004), given $(\boldsymbol{\beta}^{(t)}, \rho^{(t)})$, the $(t+1)$-th approximation of the MLE $\hat{\boldsymbol{\beta}}$ can be calculated by

$$\boldsymbol{\beta}^{(t+1)} = \arg \max_{\boldsymbol{\beta} \in \mathbb{R}^q} Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho^{(t)}). \tag{2.14}$$

2.3 MLEs of parameters via a fast QLB algorithm

We noted that $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho)$ defined by (2.13) is a special log-likelihood function up to a constant within a traditional logistic regression model. Although the calculation of $\boldsymbol{\beta}^{(t+1)}$ in (2.14) can be implemented by the NR algorithm (equivalently the Fisher scoring algorithm for the logistic regression setting), we want to employ the QLB algorithm (Böhning & Lindsay,

1988) with monotonic convergence since the NR algorithm cannot guarantee the increase of $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho)$ at each iteration. Specifically, the QLB iteration is defined by

$$\begin{aligned} \boldsymbol{\beta}^{(t+1)} &= \boldsymbol{\beta}^{(t)} + \boldsymbol{B}^{-1}\nabla Q(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}, \rho^{(t)}) \\ &= \boldsymbol{\beta}^{(t)} + 2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top[\boldsymbol{y}^{(t)} - 2\boldsymbol{\pi}^{(t)}], \end{aligned} \tag{2.15}$$

where

$$\nabla Q(\boldsymbol{\beta}^{(t)}|\boldsymbol{\beta}^{(t)}, \rho^{(t)}) \;\hat{=}\; \left.\frac{\partial Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho^{(t)})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(t)}},$$

$\boldsymbol{B}_{q\times q} = (1/2)\boldsymbol{X}^\top\boldsymbol{X}$, $\boldsymbol{X}_{n\times q} = (\boldsymbol{x}_{(1)}, \ldots, \boldsymbol{x}_{(n)})^\top$ and $\boldsymbol{y}^{(t)} = (a_1(\pi_1^{(t)}, \rho^{(t)}), \ldots, a_n(\pi_n^{(t)}, \rho^{(t)}))^\top$.

In addition, Tian, Tang & Liu (2012) proposes a novel 'shrinkage parameter' approach to accelerate the QLB algorithm while maintaining its simplicity and stability (i.e., monotonic increase in the target function). The fast QLB iteration is defined by

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + 4(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top[\boldsymbol{y}^{(t)} - 2\boldsymbol{\pi}^{(t)}], \tag{2.16}$$

which is double faster than (2.15), as shown by Tian, Tang & Liu (2012). By combining (2.16) with (2.7), we summarize the proposed algorithm as follows:

$$\begin{cases} \boldsymbol{\beta}^{(t+1)} &= \boldsymbol{\beta}^{(t)} + 4(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top[\boldsymbol{y}^{(t)} - 2\boldsymbol{\pi}^{(t)}], \\ \rho^{(t+1)} &= \rho^{(t)} - \left[\dfrac{\partial^2\ell(\boldsymbol{\pi}^{(t)}, \rho^{(t)})}{\partial\rho^2}\right]^{-1} \dfrac{\partial\ell(\boldsymbol{\pi}^{(t)}, \rho^{(t)})}{\partial\rho}. \end{cases} \tag{2.17}$$

The *standard errors* (se) of each component of the unconstrained MLEs $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\rho})^\top$ can be estimated by the square roots of the diagonal elements of the inverse Fisher information matrix $\boldsymbol{I}^{-1}(\hat{\boldsymbol{\theta}})$, where $\boldsymbol{I}(\boldsymbol{\theta})$ denotes the Fisher information matrix given by (A.2). Therefore, the asymptotic 95% confidence interval of θ_i can be obtained as $\hat{\theta}_i \pm 1.96 \times \widehat{\text{se}}$.

3. Hypothesis testing

In this section, we are interested in testing the following null hypothesis

$$H_0\text{: } \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{0}_m, \tag{3.1}$$

where $\boldsymbol{C}$ is an $m \times q$ matrix with $\text{rank}\,(\boldsymbol{C}) = r$ and $r < q$. In Ma & Liu (2017), they tested the equality of disease rates in several groups, which is just a special case of (3.1) with $\boldsymbol{C} = (0, 1)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$.

3.1 The likelihood ratio test (LRT)

The likelihood ratio statistic is given by

$$T_L = 2[\ell(\hat{\boldsymbol{\beta}}, \hat{\rho}) - \ell(\hat{\boldsymbol{\beta}}_{H_0}, \hat{\rho}_{H_0})],$$

where $(\hat{\boldsymbol{\beta}}, \hat{\rho})$ are calculated via the fast QLB algorithm (2.17), denoting the unconstrained MLEs of $(\boldsymbol{\beta}, \rho)$; while $(\hat{\boldsymbol{\beta}}_{H_0}, \hat{\rho}_{H_0})$ are the constrained MLEs of $(\boldsymbol{\beta}, \rho)$ under H_0, which can be obtained by

$$\begin{aligned} \boldsymbol{\beta}_{H_0}^{(t+1)} &= \arg\max_{\boldsymbol{\beta}} Q(\boldsymbol{\beta}|\boldsymbol{\beta}_{H_0}^{(t)}, \rho_{H_0}^{(t)}) \quad \text{subject to} \quad \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{0}, \quad \text{and} \qquad (3.2) \\ \rho_{H_0}^{(t+1)} &= \rho_{H_0}^{(t)} - \left[\frac{\partial^2 \ell(\boldsymbol{\pi}_{H_0}^{(t)}, \rho_{H_0}^{(t)})}{\partial \rho^2}\right]^{-1} \frac{\partial \ell(\boldsymbol{\pi}_{H_0}^{(t)}, \rho_{H_0}^{(t)})}{\partial \rho}. \end{aligned}$$

Since $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho)$ defined in (2.13) is a nonlinear concave function, the constrained optimization (3.2) can be solved by many effective methods such as the active-set or interior-point methods. In this paper, we employed the built-in fmincon function in MATLAB. Under the null hypothesis H_0, T_L is asymptotically distributed as a chi-squared distribution with r degrees of freedom.

3.2 The Wald test

Let $\boldsymbol{A}_{m\times(q+1)} = [\boldsymbol{C}, \boldsymbol{0}]$ and $\boldsymbol{\theta}_{(q+1)\times 1} = (\boldsymbol{\beta}^\top, \rho)^\top$, then the null hypothesis H_0 becomes $\boldsymbol{A}\boldsymbol{\theta} = [\boldsymbol{C}, \boldsymbol{0}](\boldsymbol{\beta}^\top, \rho)^\top = \boldsymbol{0}$. The Wald statistic is given by

$$T_W = (\boldsymbol{A}\hat{\boldsymbol{\theta}})^\top [\boldsymbol{A}\boldsymbol{I}^{-1}(\hat{\boldsymbol{\theta}})\boldsymbol{A}^\top]^{-1}(\boldsymbol{A}\hat{\boldsymbol{\theta}}),$$

where $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\rho})^\top$ denote the unconstrained MLEs of $\boldsymbol{\theta}$ and $\boldsymbol{I}(\boldsymbol{\theta})$ denotes the Fisher information matrix given by (A.2). T_W is asymptotically distributed as a chi-squared distribution with r degrees of freedom.

3.3 The score test

The score test statistic T_S is given by

$$T_S = [\boldsymbol{s}(\hat{\boldsymbol{\theta}}_{H_0})]^\top \boldsymbol{I}^{-1}(\hat{\boldsymbol{\theta}}_{H_0}) \boldsymbol{s}(\hat{\boldsymbol{\theta}}_{H_0}),$$

where $\hat{\boldsymbol{\theta}}_{H_0} = (\hat{\boldsymbol{\beta}}_{H_0}^\top, \hat{\rho}_{H_0})^\top$ denote the constrained MLEs of $\boldsymbol{\theta}$ under H_0, and

$$\boldsymbol{s}(\boldsymbol{\theta}) \hat{=} \left(\frac{\partial \ell(\boldsymbol{\beta}, \rho)}{\partial \beta_0}, \ldots, \frac{\partial \ell(\boldsymbol{\beta}, \rho)}{\partial \beta_{q-1}}, 0 \right)^\top,$$

which is given by (A.1). T_S is asymptotically distributed as a chi-squared distribution with r degrees of freedom.

4. Simulation studies

To evaluate the performance of the proposed fast QLB algorithm (2.17) and three testing methods in Section 3, we first investigate the accuracy of point estimates and confidence interval estimates for different parameter settings via simulation studies. Next, the performance of the LRT, the Wald test and the score test is assessed by comparing their type I error rates and powers.

4.1 Accuracy of point estimates and interval estimates

To evaluate the accuracy of point estimates and confidence intervals, we consider two cases for the dimension: $q = 2$ and $q = 4$. Parameter configurations are set as follows: The sample size $n = 100$, 200, 400; $\rho = 0.3$, 0.5, 0.7;

(a) When $q = 2$, $\boldsymbol{\beta} = (-1, 2)^\top$, $\boldsymbol{x}_{(i)} = (1, x_{i1})^\top$, $\{x_{i1}\}_{i=1}^n \overset{\text{iid}}{\sim} N(0.4, \sigma_0^2)$ with $\sigma_0^2 = 1.5 \times 10^{-3}$;

(b) When $q = 4$, $\boldsymbol{\beta} = (-1, 2, -1, 2)^\top$, $\boldsymbol{x}_{(i)} = (1, x_{i1}, x_{i2}, x_{i3})^\top$, $\{x_{i1}\}_{i=1}^n \overset{\text{iid}}{\sim} N(0.4, 1.5 \times 10^{-3})$, $\{x_{i2}\}_{i=1}^n \overset{\text{iid}}{\sim} N(0.45, 10^{-3})$, $\{x_{i3}\}_{i=1}^n \overset{\text{iid}}{\sim} 0.3 + 0.06 \times t(5)$.

First, we define a finite discrete distribution, denoted by $X \sim \text{FDiscrete}_m(\boldsymbol{x}, \boldsymbol{p})$ if $\Pr(X = x_j) = p_j$ for $j = 1, \ldots, m$, where $\boldsymbol{x} = (x_1, \ldots, x_m)^\top$, $\boldsymbol{p} = (p_1, \ldots, p_m)^\top$, $0 \leqslant p_j \leqslant 1$ and $\sum_{j=1}^m p_j = 1$. Second, for $i = 1, \ldots, n$, we generate Y_i from the following distribution:

Y_i	0	1	2
$\Pr(Y_i = y_i)$	p_{i0}	p_{i1}	p_{i2}

that is, $Y_i \sim \text{FDiscrete}_3((0,1,2)^\top, \boldsymbol{p}_i)$, where $\boldsymbol{p}_i = (p_{i0}, p_{i1}, p_{i2})^\top$ is defined by (2.2). Third, we calculate the MLEs of $(\boldsymbol{\beta}^\top, \rho)^\top$ via the fast QLB algorithm (2.17) and the corresponding standard errors. Finally, we independently repeat this process 10,000 times. The resultant average bias (denoted by Bias; i.e., average MLE minus true value of the parameter), the mean square error (denoted by MSE; i.e., Bias^2 + (standard deviation)^2, the standard deviation is estimated by the sample standard deviation of 10,000 MLEs), the coverage probability (denoted by CP) and the average width of confidence interval (denoted by CIW) are reported in Tables 1 and 2. The smaller the CIWs, the more precise the confidence interval estimation. As we can see, all Bias, MSEs and coverage probabilities are satisfactory.

Table 1 Parameter estimates based on 10,000 replications for Case (a)

	ρ	0.3			0.5			0.7		
	n	100	200	400	100	200	400	100	200	400
$\hat{\beta}_0$	Bias	−.0189	−.0120	−.0050	−.0235	−.0146	−.0094	−.0344	−.0113	−.0032
	MSE	.1994	.1032	.0536	.2600	.1399	.0683	.3414	.1624	.0682
	CP	.9552	.9482	.9494	.9535	.9535	.9463	.9551	.9479	.9522
	CIW	1.7443	1.2559	.9089	1.9659	1.4539	1.0087	2.2741	1.5544	1.0233
$\hat{\beta}_1$	Bias	.0441	.0291	.0110	.0555	.0333	.0192	.0760	.0204	.0047
	MSE	.9900	.5387	.2933	1.4741	.7716	.3623	1.9044	.8607	.4020
	CP	.9521	.9511	.9510	.9515	.9523	.9496	.9552	.9498	.9504
	CIW	3.8462	2.8736	2.1278	4.6610	3.4020	2.3229	5.3592	3.5943	2.4844
$\hat{\rho}$	Bias	−.0108	−.0044	−.0023	−.0077	−.0045	−.0019	−.0057	−.0023	−.0011
	MSE	.0098	.0048	.0024	.0081	.0040	.0019	.0055	.0027	.0013
	CP	.9409	.9446	.9462	.9402	.9473	.9520	.9409	.9438	.9484
	CIW	.3799	.2683	.1896	.3470	.2445	.1726	.2862	.2016	.1428

Table 2 Parameter estimates based on 10,000 replications for Case (b)

	ρ	0.3			0.5			0.7		
	n	100	200	400	100	200	400	100	200	400
$\hat{\beta}_0$	Bias	−.0648	−.0682	−.0132	−.0518	−.0255	−.0078	−.0436	−.0167	−.0136
	MSE	1.4208	.5655	.3024	1.8714	.7086	.3662	2.3333	.9540	.4613
	CP	.9483	.9485	.9506	.9494	.9486	.9482	.9510	.9544	.9507
	CIW	4.5808	2.8697	2.1330	5.1962	3.2369	2.3615	5.8134	3.8079	2.6294
$\hat{\beta}_1$	Bias	.0930	.0324	.0345	.0757	.0346	.0212	.1181	.0393	.0251
	MSE	1.0767	.5590	.3054	1.6137	.9241	.4025	1.9436	.8725	.4512
	CP	.9493	.9494	.9479	.9497	.9437	.9504	.9500	.9479	.9504
	CIW	3.9332	2.9019	2.1368	4.8028	3.6298	2.4716	5.2545	3.5782	2.6173
$\hat{\beta}_2$	Bias	−.0061	−.0053	−.0204	−.0253	−.0173	−.0225	−.0741	−.0387	−.0096
	MSE	3.0323	1.4664	.7467	3.5625	1.7114	.8559	5.4095	2.2990	.9219
	CP	.9512	.9476	.9471	.9478	.9483	.9527	.9514	.9503	.9498
	CIW	6.6641	4.6945	3.3705	7.1563	5.0258	3.6211	8.8206	5.8564	3.7432
$\hat{\beta}_3$	Bias	.1016	.1934	.0281	.1063	.0599	.0298	.0908	.0570	.0236
	MSE	6.3117	2.0239	1.2900	4.8338	2.8539	1.4032	9.5762	3.7430	1.7111
	CP	.9502	.9537	.9468	.9491	.9514	.9515	.9489	.9497	.9491
	CIW	9.6467	5.2832	4.3639	8.3203	6.5245	4.6293	11.6542	7.4092	5.0665
$\hat{\rho}$	Bias	−.0199	−.0094	−.0040	−.0147	−.0075	−.0037	−.0107	−.0049	−.0025
	MSE	.0104	.0051	.0024	.0085	.0041	.0019	.0057	.0027	.0014
	CP	.9370	.9396	.9494	.9389	.9465	.9494	.9441	.9499	.9462
	CIW	.3834	.2697	0.1896	.3508	.2450	.1728	.2911	.2034	.1430

4.2 Comparison of three tests

In Section 3, the LRT, the Wald test and the score test are developed for testing

$$H_0\colon \boldsymbol{C}\boldsymbol{\beta} = \mathbf{0} \quad \text{against} \quad H_1\colon \boldsymbol{C}\boldsymbol{\beta} \neq \mathbf{0}. \tag{4.1}$$

To compare the type I error rates for the three tests, we consider two cases for the dimension: $q = 2$ and $q = 4$. Parameter configurations are set as follows: The sample size $n = 100$, 200, 400; $\rho = 0.3$, 0.5, 0.7;

(A1) When $q = 2$, $\boldsymbol{C} = (1, -1)$, $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ so that (4.1) becomes H_0: $\beta_0 = \beta_1$. We set $\beta_0 = \beta_1 = 0$, $\boldsymbol{x}_{(i)} = (1, x_{i1})^\top$, $\{x_{i1}\}_{i=1}^n \overset{\text{iid}}{\sim} N(0.4, \sigma_0^2)$ with $\sigma_0^2 = 1.5 \times 10^{-3}$;

(B1) When $q = 4$,

$$\boldsymbol{C} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}, \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top,$$

so that (4.1) becomes $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3$. We set $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$, $\boldsymbol{x}_{(i)} = (1, x_{i1}, x_{i2}, x_{i3})^\top$, $\{x_{i1}\}_{i=1}^n \overset{\text{iid}}{\sim} N(0.4, 1.5 \times 10^{-3})$, $\{x_{i2}\}_{i=1}^n \overset{\text{iid}}{\sim} N(0.45, 10^{-3})$, $\{x_{i3}\}_{i=1}^n \overset{\text{iid}}{\sim} 0.3 + 0.06 \times t(5)$.

For $i = 1, \ldots, n$, we generate $Y_i \sim \text{FDiscrete}_3((0, 1, 2)^\top, \boldsymbol{p}_i)$ for 10,000 times, where $\boldsymbol{p}_i = (p_{i0}, p_{i1}, p_{i2})^\top$ is defined by (2.2). The type I error rate for a given test can be simply estimated by the number of rejections at the 0.05 significance level divided by 10,000. In addition, we compare the proposed three tests with the test based on GEE method (Ratcliffe & Shults 2008). Under the framework of GEE, we treat the left and right eye measures from the same object as repeated measure with exchangeable correlation structure. We also used Bernoulli distribution to model the response, namely the monocular data. In addition, model-based standard error estimates are used.

The test statistic based on the GEE method is denoted by T_{GEE}. The empirical type I error rates for the three test statistics and T_{GEE} are displayed in Table 3. From Table 3, we can see that all four tests perform similarly in controlling the type I error rates around the pre-chosen nominal level.

Table 3 The empirical type I error rates of statistics (T_L, T_W, T_S, T_{GEE}) for testing (4.1)

		Case (A1)				Case (B1)			
ρ	n	T_L	T_W	T_S	T_{GEE}	T_L	T_W	T_S	T_{GEE}
0.3	100	.0528	.0492	.0513	.0490	.0512	.0424	.0449	.0417
	200	.0463	.0449	.0451	.0445	.0496	.0370	.0421	.0360
	400	.0533	.0524	.0527	.0529	.0512	.0488	.0493	.0481
0.5	100	.0531	.0496	.0505	.0490	.0545	.0416	.0459	.0404
	200	.0516	.0493	.0513	.0499	.0568	.0490	.0524	.0476
	400	.0509	.0503	.0502	.0499	.0517	.0479	.0487	.0464
0.7	100	.0524	.0486	.0514	.0488	.0586	.0411	.0503	.0388
	200	.0476	.0447	.0468	.0450	.0526	.0427	.0475	.0408
	400	.0520	.0507	.0514	.0508	.0495	.0442	.0466	.0445

Next, to evaluate the power performance of the three tests, we consider two cases for the regression coefficients, where Case (A2): $\boldsymbol{\beta} = (-2, 2, -1, 1)^\top$ and Case (B2): $\boldsymbol{\beta} = (-1, 2, -1, 2)^\top$. Here, we omit the simpler case where q = 2, this is because their patterns of power are quite similar with that of q=4. Instead, we try to focus on the more complicated case. We tried 2 different parameter settings with q= 4, and it turns out that all the four test statistics have the similar pattern under different parameter settings.

Other parameter configurations are set as follows: The sample size $n = 50$, 80, 100, 200, 400; $\rho = 0.3$, 0.5, 0.7; $q = 4$,

$$\boldsymbol{C} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}, \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top,$$

so that (4.1) becomes $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3$ against H_1: At least two of $(\beta_0, \beta_1, \beta_2, \beta_3)$ are not equal; $\boldsymbol{x}_{(i)} = (1, x_{i1}, x_{i2}, x_{i3})^\top$, $\{x_{i1}\}_{i=1}^n \overset{\text{iid}}{\sim} N(0.4, 1.5 \times 10^{-3})$, $\{x_{i2}\}_{i=1}^n \overset{\text{iid}}{\sim} N(0.45, 10^{-3})$, $\{x_{i3}\}_{i=1}^n \overset{\text{iid}}{\sim} 0.3 + 0.06 \times t(5)$.

For $i = 1, \ldots, n$, we generate $Y_i \sim \text{FDiscrete}_3((0, 1, 2)^\top, \boldsymbol{p}_i)$ for 10,000 times, where $\boldsymbol{p}_i = (p_{i0}, p_{i1}, p_{i2})^\top$ is defined by (2.2). The power for a given test can be simply estimated by the number of rejections at the 0.05 significance level divided by 10,000. The simulated power for the three tests are displayed in Table 4. From Table 4, we can see that the performance of the LRT is the best among the three tests for all parameter configurations; and the score test performs better than the Wald test. Furthermore, we compare the powers of T_{GEE} with those of the proposed three statistics. From Table 4, we can see that powers of T_{GEE} are constantly lower than those of the proposed three test statistics. Particularly, when $n = 80$, most powers of T_L are 30% higher than those of T_{GEE}; when $n = 50$, most powers of T_L are 50% - 90% higher. In this sense, our method is more powerful than GEE in a limited study with small sample size.

In addition to the test performance, we also compare the estimates of regression coefficients from the GEE with our proposed method, it turns out that these two are nearly the same. In conclusion, our proposed method and GEE method have similar performance in parameter estimates and type I error rate, while the powers of our method are constantly higher than those of GEE, especially when sample sizes are small.

Table 4 The empirical powers of test statistics (T_L, T_W, T_S, T_{GEE}) for testing (4.1)

		Case (A2)				Case (B2)			
ρ	n	T_L	T_W	T_S	T_{GEE}	T_L	T_W	T_S	T_{GEE}
	50	.2163	.1343	.1569	.1394	.1689	.1286	.1417	.1303
	80	.4040	.3641	.3818	.3540	.3654	.3235	.3395	.3200
	100	.3844	.3479	.3630	.3413	.3963	.3581	.3721	.3583
0.3	200	.6538	.6214	.6333	.6071	.7126	.6938	.6992	.6884
	400	.9672	.9646	.9652	.9591	.9430	.9410	.9409	.9400
	50	.1595	.1006	.1385	.1056	.1530	.1072	.1251	.1046
	80	.2674	.2145	.2426	.2083	.2742	.2344	.2481	.2283
	100	.3279	.2877	.3114	.2789	.3149	.2613	.2876	.2556
0.5	200	.6792	.6509	.6624	.6363	.6836	.6632	.6715	.6591
	400	.9320	.9272	.9294	.9169	.9483	.9445	.9461	.9440
	50	.1600	.0800	.1130	.0820	.1924	.1140	.1594	.1096
	80	.2596	.1968	.2443	.1909	.3264	.2516	.2894	.2459
	100	.4215	.3670	.4039	.3576	.4255	.3679	.3976	.3588
0.7	200	.6916	.6567	.6734	.6444	.7245	.6750	.6987	.6671
	400	.8934	.8833	.8878	.8705	.8885	.8795	.8848	.8783

5. Analysis of an ophthalmologic data set

In this section, we analyze the ophthalmologic data set presented in Ma & Liu (2017) by employing the proposed methods. The study was from a cross-sectional, population-based sample in Iran to assess the prevalence of avoidable blindness (Rajavi *et al.*, 2011). Nearly 3000 patients were examined and blindness was assessed for seven age groups as shown in the first five columns of Table 5. Ma & Liu (2017) performed several hypothesis tests and found out that the disease rates in different age groups were different, however, they were unable to reveal the exact relationship between ages and disease rates.

Table 5 The ophthalmologic data in Iran (Rajavi *et al.*, 2011)

Age in year	Number of patients	Blindness			MLE of π_i
		None	Unilateral	Bilaterial	
50–54	989	964	23	2	0.015
55–59	566	541	17	8	0.026
60–64	491	469	18	4	0.026
65–69	278	257	16	5	0.045
70–74	277	242	32	3	0.074
75–79	166	127	30	9	0.145
80+	143	104	29	10	0.171

From (2.11), the MLEs of $\{\pi_i\}_{i=1}^n$ can be iteratively calculated via the proposed MM algorithm:

$$\pi_i^{(t+1)} = \frac{a_i(\pi_i^{(t)}, \rho^{(t)})}{2}, \quad i = 1, \ldots, n.$$

These results are displayed in the last column of Table 5, exhibiting a potential trend that $\{\hat{\pi}_i\}_{i=1}^n$ are increasing with the ages of the patient groups. To further explore the relationship between ages and disease rates, we consider

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + x_{i1}\beta_1,$$

where x_{i1} denotes the age covariate for patient i. In this example, since the exact ages are not available in the data set, we set the ages in different groups as their middle points and treat them as continuous variables. By the fast QLB algorithm (2.17), we can calculate the MLEs and confidence intervals of parameters. These results are reported in Table 6. From Table 6, $\hat{\rho} = 0.2733$ indicates a positive correlation between eyes from the same patient, while $\hat{\beta}_1 = 0.0688$ reveals a positive relationship between disease rate and age. The estimation results coincide with the common sense that older people have higher chance to suffer from eye disease. Besides, we compare the estimates of regression coefficients from the GEE with our proposed method: similarly as in section 4, these two are nearly the same, so we omit the GEE results in Table 6.

Table 6 MLEs and confidence intervals of parameters for the ophthalmologic data in Iran

Parameter	MLE	Standard error	95% CI
β_0	-6.9885	0.3525	$[-7.6795, -6.2975]$
β_1	0.0688	0.0056	[0.0577, 0.0798]
ρ	0.2733	0.0406	[0.1937, 0.3530]

Finally, we are interested in testing H_0: $\beta_1 = 0$ against H_1: $\beta_1 \neq 0$, which is corresponding to $\boldsymbol{C} = (0, 1)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ in (3.1). The values of three statistics T_L, T_W, T_S, T_{GEE} and their p-values are listed in Table 7. Since three p-values are almost equal to zero, the null hypothesis H_0 must be rejected with a strong evidence. This is consistent to the conclusion from the CI of β_1, which excludes zero. In addition, by comparing the p-values of four Tests, we can see that our three tests outperform GEE because GEE has a larger p-value. This result coincides with the previous observation in the simulation.

Table 7 Values of four statistics T_L, T_W, T_S, T_{GEE} and their p-values

Method	T_L	T_W	T_S	T_{GEE}
Statistic	151.7455	149.0259	169.1516	137.9613
p-value	7.2018e-35	2.1336e-34	1.8305e-38	7.4312e-32

In conclusion, our data analysis shows that age has a significant positive effect on disease rate. That is, older people are more likely to suffer from eye diseases. This result coincides with that obtained by Ma & Liu (2017), presenting a trend that the MLEs of disease probabilities are increasing with the ages of the patient groups as shown in the last column of Table 5.

6. Discussions

Based on the model of Donner (1989), in this paper, we introduced the logistic regression model to explore the relationship between ophthalmologic disease rate and covariates such as patients' ages, weights, gender and so on. To avoid direct optimization on the complicated, non-concave log-likelihood $\ell(\boldsymbol{\beta}, \rho)$, we successfully constructed a minimizing function

$Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}, \rho)$ by utilizing Jensen's inequality, where the Q function enjoyed a simple and concave form of the log-likelihood function in the original logistic regression. In Section 2, we combined the fast QLB algorithm with the coordinate ascend method to calculate the unconstrained MLEs of the parameters of interest. The confidence intervals of parameters could be derived by estimating standard errors via the inverse Fisher information matrix.

We have developed three large-sample tests (i.e., the LRT, Wald test, and score test) for testing if covariates have a significant impact on the disease probability. Simulations were implemented to assess the performance of the proposed fast QLB algorithm and three testing approaches. Simulation results showed that all three tests could satisfactorily control the type I error rates regardless of sample size or parameter configurations. Moreover, the LRT had the highest power among the three tests and its power increased steadily with the sample sizes. Therefore, the LRT is highly recommended. Besides, we compare the proposed three tests to the test based on *generalized estimating equations* (GEE) method with the logit link function. The simulation results show that the test based on the GEE method has similar type I error rates, but lower powers compared to the proposed three tests. This is not surprising because GEE method is not a likelihood-based method and requires fewer assumptions. Despite the lower power, GEE approach is more robust to model misspecification. On the other hand, this difference became smaller when the sample size increased to $n = 400$, where the power of GEE test seemed to be very close to the proposed Wald-test according to our simulation results. However, in all cases where $n = 50$, 80, 100 and 200, our proposed method presented a relatively better power, Especially when $n = 50$ and 80, most powers of our proposed LRT are 30% - 90% higher than those of GEE. In this sense, our method is more powerful when sample size is small.

In addition to logit transformation, there are many other link functions (e.g., probit, log-log, complementary log-log) connecting covariates to disease rate. One of our future works is to introduce other link functions in generalized linear Donner's model. Moreover, when the prior information about the ophthalmologic disease rates are available, we could employ the Bayesian method to generate posterior samples via MCMC. In addition, we can consider the mixed model, which provides the ability to model the group-specific random

effect. However, the introduction of random effect will finally lead to another optimization algorithm, which is similar to the idea of this work. The mixed model may be our future work as an extension to our proposed method.

Acknowledgements

G.L. Tian's research was supported by a grant from the National Natural Science Foundation of China (No. 11771199).

References

Becker, M.P., Yang, I. and Lange, K. (1997). EM algorithms without missing data. *Statistical Methods in Medical Research* **6**: 38–54.

Bodian, C.A. (1994). Intraclass correlation for two-by-two tables under three sampling designs. *Biometrics* **50**: 183–193.

Böhning, D. and Lindsay, B.G. (1988). Monotonicity of quadratic approximation algorithms. *Annals of the Institute of Statistical Mathematics* **40**: 641–663.

Dallal, G.E. (1988). Paired Bernoulli trials. *Biometrics* **44**: 253–257.

Donner, A. (1989). Statistical methods in ophthalmology: An adjusted chi-square approach. *Biometrics* **45**(2): 605–661.

Donner, A. and Banting, D. (1988). Analysis of site-specific data in dental studies. *Journal of Dental Research* **67**(11): 1392–1395.

Hunter, D.R. and Lange, K. (2004). A tutorial on MM algorithms. *The American Statistician* **58**(1): 30–37.

Lange, K., Hunter, D.R. and Yang, I. (2000). Optimization transfer using surrogate objective functions (with discussions). *Journal of Computational and Graphical Statistics* **9**: 1–20.

Ma, C.X. and Liu, S. (2017). Testing equality of proportions for correlated binary data in ophthalmologic studies. *Journal of Biopharmaceutical Statistics* **27**(4): 611–619.

Ratcliffe, S.J. and Shults, J. (2008). GEEQBOX: A MATLAB toolbox for generalized

estimating equations and quasi-least squares. *Journal of Statistical Software* **25**(14): 1–14.

Rajavi, Z., Katibeh, M., Ziaei, H., Fardesmaeilpour, N., Sehat, M., Ahmadieh, H. and Javadi, M.A. (2011). Rapid assessment of avoidable blindness in Iran. *Ophthalmology* **118**(9): 1812–1818.

Rosner, B. (1982). Statistical method in ophthalmology: An adjustment for the intraclass correlation between eyes. *Biometrics* **38**(1): 105–114.

Rosner, B. and Milton, R.C. (1988). Significance testing for correlated binary outcome data. *Biometrics* **44**: 505–512.

Tang, N.S., Tang, M.L. and Qiu, S.F. (2008). Testing the equality of proportions for correlated otolaryngologic data. *Computational Statistics and Data Analysis* **52**: 3719–3729.

Tian, G.L., Huang, X.F. and Xu, J.F. (2018). An assembly and decomposition approach for constructing separable minorizing functions in a class of MM algorithms. *Statistica Sinica*, in press.

Tian, G.L., Tang, M.L. and Liu, C.L. (2012). Accelerating the quadratic lower-bound algorithm via optimizing the shrinkage parameter. *Computational Statistics and Data Analysis* **56**(2): 255–265.

Appendix A: Derivation of $\boldsymbol{s}(\boldsymbol{\theta})$ and Fisher information matrix

A.1 The expression of $\boldsymbol{s}(\boldsymbol{\theta})$

The elements of the $(q+1)$-dimensional vector

$$\boldsymbol{s}(\boldsymbol{\theta}) = \left(\frac{\partial \ell(\boldsymbol{\beta}, \rho)}{\partial \beta_0}, \ldots, \frac{\partial \ell(\boldsymbol{\beta}, \rho)}{\partial \beta_{q-1}}, 0 \right)^{\top} \tag{A.1}$$

are given by

$$\frac{\partial \ell(\boldsymbol{\beta}, \rho)}{\partial \beta_s} = \sum_{i=1}^{n} \frac{\partial \ell(\boldsymbol{\pi}, \rho)}{\partial \pi_i} \cdot \frac{\partial \pi_i}{\partial \beta_s}, \quad s = 0, 1, \ldots, q-1,$$

where

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\pi},\rho)}{\partial \pi_i} &= \left\{\frac{(2\pi_i-1)I(Y_i=1)}{\pi_i(\pi_i-1)}+\frac{(\rho+2\pi_i-2\rho\pi_i)I(Y_i=2)}{\pi_i(\rho+\pi_i-\rho\pi_i)}\right.\\ &\quad \left.-\frac{(\rho+2\pi_i-2\rho\pi_i-2)I(Y_i=0)}{(\pi_i-1)(\rho\pi_i-\pi_i+1)}\right\},\\ \frac{\partial \pi_i}{\partial \beta_s} &= \frac{x_{is}\mathrm{e}^{-\boldsymbol{x}_{(i)}\boldsymbol{\beta}}}{(1+\mathrm{e}^{-\boldsymbol{x}_{(i)}\boldsymbol{\beta}})^2}, \quad i=1,\ldots,n.\end{aligned}$$

A.2 The Fisher information matrix

The Fisher information matrix is given by

$$\boldsymbol{I}(\boldsymbol{\theta}) = -E\begin{pmatrix}\boldsymbol{H}(\boldsymbol{\beta}) & \boldsymbol{h}(\boldsymbol{\theta})\\ \boldsymbol{h}^\top(\boldsymbol{\theta}) & \dfrac{\partial^2\ell}{\partial\rho^2}\end{pmatrix}, \tag{A.2}$$

where

$$\begin{aligned}\frac{\partial^2\ell}{\partial\rho^2} &= -\sum_{i=1}^n\left\{\frac{I(Y_i=1)}{(\rho-1)^2}+\frac{\pi_i^2 I(Y_i=0)}{(\rho\pi_i-\pi_i+1)^2}+\frac{(\pi_i-1)^2I(Y_i=2)}{(\rho+\pi_i-\rho\pi_i)^2}\right\},\\ \boldsymbol{h}(\boldsymbol{\theta}) &= \left(\frac{\partial^2\ell}{\partial\beta_0\partial\rho},\ldots,\frac{\partial^2\ell}{\partial\beta_{q-1}\partial\rho}\right)^\top, \quad \text{and}\\ \boldsymbol{H}(\boldsymbol{\beta}) &= \left(\frac{\partial^2\ell}{\partial\beta_s\partial\beta_t}\right), \quad s,t=0,1,\ldots,q-1.\end{aligned}$$

The elements of $\boldsymbol{h}(\boldsymbol{\theta})$ and $\boldsymbol{H}(\boldsymbol{\beta})$ are given by

$$\begin{aligned}\frac{\partial^2\ell}{\partial\beta_s\partial\rho} &= \frac{\partial}{\partial\rho}\left(\sum_{i=1}^n\frac{\partial\ell}{\partial\pi_i}\frac{\partial\pi_i}{\partial\beta_s}\right)=\sum_{i=1}^n\frac{\partial^2\ell}{\partial\rho\partial\pi_i}\frac{\partial\pi_i}{\partial\beta_s}+\sum_{i=1}^n\frac{\partial^2\pi_i}{\partial\beta_s\partial\rho}\frac{\partial\ell}{\partial\pi_i}\\ &= \sum_{i=1}^n x_{is}\mathrm{e}^{\boldsymbol{x}_{(i)}^\top\boldsymbol{\beta}}\Big[n_{i0}(\rho+\mathrm{e}^{-\boldsymbol{x}_{(i)}^\top\boldsymbol{\beta}})^{-2}-n_{i2}(2\rho+\rho\mathrm{e}^{-\boldsymbol{x}_{(i)}^\top\boldsymbol{\beta}}-1)^{-2}\Big],\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \beta_s^2} &= \sum_{i=1}^{n} \Bigg\{ \Big[\frac{x_{is} \mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}}{(1+\mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}})^2} \Big]^2 \Big\{ \frac{(-2\pi_i^2 + 2\pi_i - 1) I(Y_i = 1)}{\pi_i^2 (\pi_i - 1)^2} \\
&\quad + \frac{[-2(\rho+1)\pi_i^2 + (\rho^2+2\rho)\pi_i - \rho^2] I(Y_i = 2)}{[\pi_i(\rho + \pi_i - \rho\pi_i)]^2} \\
&\quad - \frac{[2(\rho-1)^2\pi_i^2 + (2-\rho)^2 \rho \pi_i + \rho^2 - 2\rho + 2] I(Y_i = 0)}{(\pi_i - 1)^2 (\rho\pi_i - \pi_i + 1)^2} \Big\} \\
&\quad + \Bigg[\frac{(2\pi_i - 1) I(Y_i = 1)}{\pi_i(\pi_i - 1)} + \frac{(\rho + 2\pi_i - 2\rho\pi_i) I(Y_i = 2)}{\pi_i(\rho + \pi_i - \rho\pi_i)} \\
&\quad - \frac{(\rho + 2\pi_i - 2\rho\pi_i - 2) I(Y_i = 0)}{(\pi_i - 1)(\rho\pi_i - \pi_i + 1)} \Bigg] \\
&\quad \times \frac{x_{is}(1+\mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}) \mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}} [-x_{is}(1+\mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}) + 2x_{is}\mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}]}{(1+\mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}})^4} \Bigg\}, \\
\frac{\partial^2 \ell}{\partial \beta_s \partial \beta_t} &= \frac{\partial}{\partial \beta_t} \left(\sum_{i=1}^{n} \frac{\partial \ell}{\partial \pi_i} \frac{\partial \pi_i}{\partial \beta_s} \right) = \sum_{i=1}^{n} \frac{\partial^2 \ell}{\partial \beta_s \partial \pi_i} \frac{\partial \pi_i}{\partial \beta_s} + \sum_{i=1}^{n} \frac{\partial^2 \pi_i}{\partial \beta_t \partial \beta_s} \frac{\partial \ell}{\partial \pi_i} \\
&= \sum_{i=1}^{n} \frac{\partial \pi_i}{\partial \beta_s} \left(\sum_{j=1}^{n} \frac{\partial^2 \ell}{\partial \pi_j \partial \pi_i} \frac{\partial \pi_j}{\partial \beta_s} \right) + \sum_{i=1}^{n} \frac{\partial^2 \pi_i}{\partial \beta_t \partial \beta_s} \frac{\partial \ell}{\partial \pi_i} \\
&= \sum_{i=1}^{n} \left(\frac{\partial \pi_i}{\partial \beta_s} \right)^2 \left(\frac{\partial^2 \ell}{\partial \pi_i^2} \right) + \sum_{i=1}^{n} \frac{\partial^2 \pi_i}{\partial \beta_t \partial \beta_s} \frac{\partial \ell}{\partial \pi_i} \\
&= \sum_{i=1}^{n} \Bigg\{ \left[\frac{x_{is}^2 \mathrm{e}^{-2\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}}}{(1+\mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}})^4} \right] \Bigg[- \frac{(2\pi_i^2 - 2\pi_i + 1) I(Y_i = 1)}{\pi_i^2 (\pi - i - 1)^2} \\
&\quad - \frac{(2\rho^2\pi_i^2 - 2\rho^2\pi_i + \rho^2 - 4\rho\pi_i^2 + 2\rho\pi_i + 2\pi_i^2) I(Y_i = 2)}{\pi_i^2 (\rho + \pi_i - \rho\pi_i)^2} \\
&\quad - \frac{(2\rho^2\pi_i^2 - 2\rho^2\pi_i + \rho^2 - 4\rho\pi_i^2 + 6\rho\pi_i - 2\rho + 2\pi_i^2 - 4\pi_i + 2)}{(\pi_i - 1)^2 (\rho\pi_i - \pi_i + 1)^2} \Bigg] \Bigg\} \\
&\quad - \sum_{i=1}^{n} \Bigg\{ \left[\frac{-x_{is} x_{it} \mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}} (3\mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}} + 1)}{(1+\mathrm{e}^{-\boldsymbol{x}_{(i)}^\top \boldsymbol{\beta}})^3} \right] \Bigg[\frac{(2\pi_i - 1) I(Y_i = 1)}{\pi_i(\pi_i - 1)} \\
&\quad + \frac{(\rho + 2\pi_i - 2\rho\pi_i) I(Y_i = 2)}{\pi_i(\rho + \pi_i - \rho\pi_i)} - \frac{(\rho + 2\pi_i - 2\rho\pi_i - 2) I(Y_i = 0)}{(\pi_i - 1)(\rho\pi_i - \pi_i + 1)} \Bigg] \Bigg\},
\end{aligned}
$$

where $s \neq t$ and $\frac{\partial^2 \ell}{\partial \pi_i \partial \pi_j} = 0$. Since

$$\begin{aligned} E[I(Y_i = 0)] &= p_{i0} = (1 - \pi_i)(\rho\pi_i - \pi_i + 1), \\ E[I(Y_i = 1)] &= p_{i1} = 2\pi_i(1 - \rho)(1 - \pi_i), \\ E[I(Y_i = 2)] &= p_{i2} = \pi_i^2 + \rho\pi_i(1 - \pi_i), \quad i = 1, \ldots, n, \end{aligned}$$

substituting $E[I(\cdot)]$ for $I(\cdot)$ in the above formulae, we can obtain the Fisher information matrix (A.2).

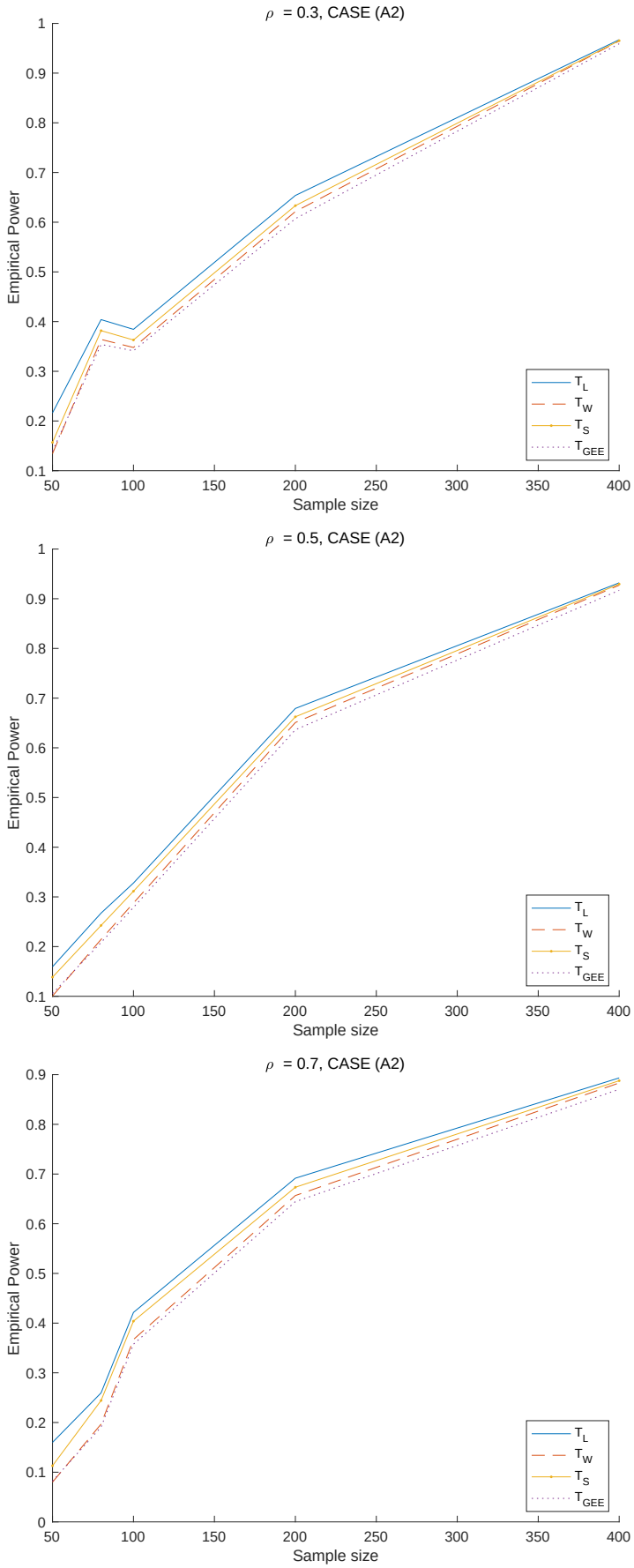

Figure 1. The empirical powers of three test statistics (T_L, T_W, T_S) for testing (4.1).

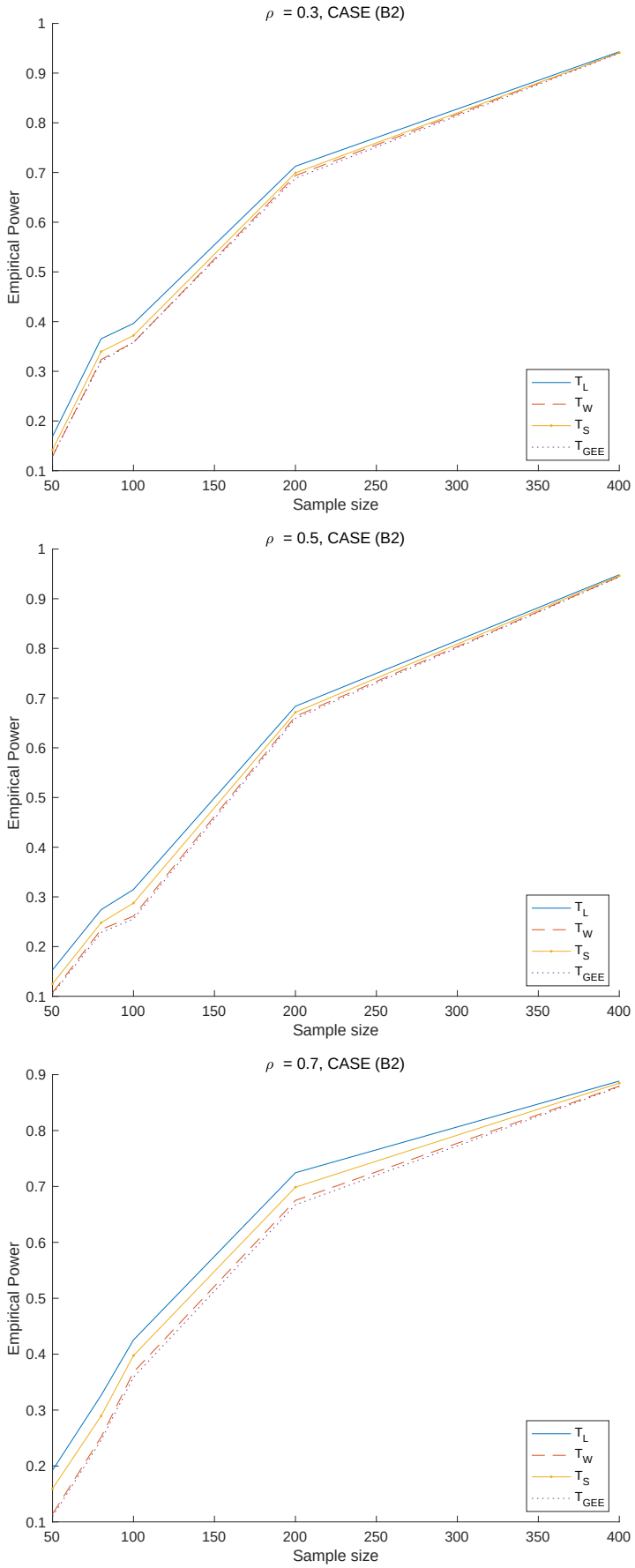

Figure 2. The empirical powers of three test statistics (T_L, T_W, T_S) for testing (4.1).