

Quantile estimation under a two-sample semi-parametric model

BIAO ZHANG

*Department of Mathematics, The University of Toledo, Toledo OH 43606, USA.
E-mail: bzhang@math.utoledo.edu*

We consider quantile estimation under a two-sample semi-parametric model in which the log ratio of two unknown density functions has a known parametric form. This two-sample semi-parametric model, arising naturally from case-control studies and logistic discriminant analysis, can be regarded as a biased sampling model. A new quantile estimator is constructed on the basis of the maximum semi-parametric likelihood estimator of the underlying distribution function. It is shown that the proposed quantile estimator is asymptotically normally distributed with smaller asymptotic variance than that of the standard quantile estimator. Also presented are some results on simulation and from the analysis of a real data set.

Keywords: biased sampling problem; case-control data; Gaussian process; logistic discriminant analysis; logistic regression; relative efficiency; weak convergence

1. Introduction

Let X_1, \dots, X_{n_0} be a random sample from a population with distribution function G and density or frequency function g . Independent of the X_i s, let Z_1, \dots, Z_{n_1} be another random sample from a population with distribution function H and density or frequency function h . The quantile function associated with G is the function defined by $G^{-1}(s) = \inf\{t; G(t) \geq s\}$ for $s \in (0, 1)$. On the basis of the sample X_1, \dots, X_{n_0} , the standard estimator of the quantile function $G^{-1}(s)$ is the empirical quantile function $\hat{G}^{-1}(s)$, where \hat{G} is the empirical distribution function of X_1, \dots, X_{n_0} given by $\hat{G}(x) = (1/n_0) \sum_{i=1}^{n_0} I_{[X_i \leq x]}$. The process $\sqrt{n}(\hat{G}^{-1} - G^{-1})$ is referred to as the quantile process. In this paper, we consider estimating $G^{-1}(s)$ under the following two-sample semi-parametric model in which the two unknown density functions g and h are linked by an ‘exponential tilt’ $\exp[\alpha + r(x)\beta]$:

$$X_1, \dots, X_{n_0} \stackrel{\text{i.i.d.}}{\sim} g(x), \quad Z_1, \dots, Z_{n_1} \stackrel{\text{i.i.d.}}{\sim} h(x) = \exp[\alpha + r(x)\beta]g(x), \quad (1.1)$$

where $r(x) = (r_1(x), \dots, r_p(x))$ is a $1 \times p$ vector of functions of x , $\beta = (\beta_1, \dots, \beta_p)^T$ is a $p \times 1$ parameter vector, and α is a normalizing parameter that makes $h(x)$ integrate to 1. In most applications, $r(x) = x$ or $r(x) = (x, x^2)$.

For $r(x) = x$, model (1.1) encompasses many common distributions, including two exponential distributions with different means and two normal distributions with common variance but different means. Furthermore, model (1.1) with $r(x) = x$ or $r(x) = (x, x^2)$ has

wide applications in the logistic discriminant analysis (Anderson 1972; 1979). Moreover, model (1.1) with $r(x) = x$ arises naturally from case-control studies. Let Y be a binary response variable and X be the associated covariate; then the (prospective) logistic regression model is of the form

$$P(Y = 1|X = x) = \frac{\exp(\alpha^* + x\beta)}{1 + \exp(\alpha^* + x\beta)}, \quad (1.2)$$

where α^* and β are parameters and the marginal distribution of X is not specified. In case-control studies, data are collected retrospectively in the sense that for samples of subjects having $Y = 1$ ('cases') and having $Y = 0$ ('controls'), the value x of X is observed. Specifically, let X_1, \dots, X_{n_0} be a random sample from $F(x|y = 0)$ and, independent of the X_i , let Z_1, \dots, Z_{n_1} be a random sample from $F(x|y = 1)$. If $\pi = P(Y = 1) = 1 - P(Y = 0)$ and $f(x|Y = i)$ is the conditional density or frequency function of X given $Y = i$ for $i = 0, 1$, then it can be shown from (1.2) and Bayes's rule that model (1.1) holds with $g(x) = f(x|Y = 0)$, $h(x) = f(x|Y = 1)$, $\alpha = \alpha^* + \log\{(1 - \pi)/\pi\}$ and $r(x) = x$. For an exposition on the application of logistic regression to case-control studies, see for example Breslow and Day (1980), Prentice and Pyke (1979) and Farewell (1979). For a specially designed exponential family of densities obtained by putting an exponential family 'through' a kernel estimator, Efron and Tibshirani (1996) considered density estimation by employing Poisson regression techniques. It can be shown in the case of two-sample problems that Efron and Tibshirani's exponential family of densities coincides with model (1.1) with $r(x) = (x, x^2)$.

As an alternative to the two-sample location-scale model, model (1.1), which is reminiscent of the Cox proportional hazards model, is equivalent to a two-sample semi-parametric model in which the ratio of two unspecified density functions has a known parametric form. Furthermore, that we test the equality of G and H can be regarded as a special case of model (1.1) with $\alpha = \beta = 0$. Moreover, model (1.1) can also be viewed as a biased sampling model with weight function $\exp[\alpha + r(x)\beta]$ depending on the unknown parameters α and β . Vardi (1982; 1985), Gill *et al.* (1988) and Qin (1993) discussed estimating distribution functions in biased sampling models with known weight functions. Fokianos *et al.* (1998) have applied model (1.1) with $r(x) = (x, x^2)$ to a data set from spaceborne precipitation radar and a spaceborne radiometer. Gilbert *et al.* (1998) have employed model (1.1) with $r(x) = (x, x^2)$ to analyse HIV vaccine trial data for assessing differential vaccine protection against human immunodeficiency virus types. For a more general model $h(x) = w(x, \theta)g(x)$, with w non-negative and of a known parametric form, Gilbert *et al.* (1999) have discussed identifiability issues and maximum likelihood estimation of θ and G . One application of estimating G is to test the validity of model (1.1) or to test the consistency of the 'exponential tilt' in model (1.1). Qin and Zhang (1997) considered a goodness-of-fit test for logistic regression model (1.2) based on case-control data by employing the maximum semi-parametric likelihood estimator of G to test the validity of model (1.1) with $r(x) = x$. Another application of estimating G is to facilitate estimation of functionals of G such as the generalized impact fraction discussed by Drescher and Becher (1997). In the present paper, however, we are interested in the problem of estimating quantiles of G under model (1.1). Note that since the form of $g(x)$ is

not specified, statistical inferences based on model (1.1) would be more robust than those based on a full parametric model in which the form of $g(x)$ is known.

Under model (1.1), the standard empirical quantile function $\hat{G}^{-1}(s)$ based on the (control) sample X_1, \dots, X_{n_0} ignores the information contained in the (case) sample Z_1, \dots, Z_{n_1} . In this paper, we consider an alternative yet appealing estimator of $G^{-1}(s)$, which takes both samples X_1, \dots, X_{n_0} and Z_1, \dots, Z_{n_1} into account. Our approach is first to estimate the parameters α and β by maximizing a semi-parametric likelihood function and then to obtain the maximum semi-parametric likelihood estimator \tilde{G} of G by putting weights on all the observations X_i and Z_j . This approach leads us to estimate $G^{-1}(s)$ by $\tilde{G}^{-1}(s)$. It is shown that $\tilde{G}^{-1}(s)$ is asymptotically normal with smaller asymptotic variance than that of $\hat{G}^{-1}(s)$ for each $s \in (0, 1)$. The maximum semi-parametric likelihood estimator \tilde{G} of G is derived in Section 2 by employing the empirical likelihood method developed by Owen (1988; 1990). For a more complete survey of developments in empirical likelihood, see Hall and La Scala (1990) and Owen (1991).

This paper is organized as follows. In Section 2, we propose an alternative quantile estimator $\tilde{G}^{-1}(s)$ by deriving the maximum semi-parametric likelihood estimators of α , β and G . Section 3 pertains to the asymptotic relative efficiency of the maximum semi-parametric likelihood estimator of (α, β) relative to the maximum likelihood estimator (α, β) under a two-sample parametric model. Section 4 concerns the asymptotic behaviour of $\tilde{G}^{-1}(s)$. Some numerical results are presented in Section 5 to demonstrate the performance of the proposed quantile estimator. Finally, proofs of the main theoretical results are provided in Section 6.

2. The modified empirical quantile function

Let $\{T_1, \dots, T_n\}$ denote the pooled sample $\{X_1, \dots, X_{n_0}; Z_1, \dots, Z_{n_1}\}$ with $n = n_0 + n_1$. Based on the observed data in (1.1), we can write the likelihood function as

$$\mathcal{L}(\alpha, \beta, G) = \prod_{i=1}^{n_0} dG(X_i) \prod_{j=1}^{n_1} w(Z_j) dG(Z_j) = \left\{ \prod_{i=1}^n p_i \right\} \left\{ \prod_{j=1}^{n_1} w(Z_j) \right\},$$

where $w(x) = \exp[\alpha + r(x)\beta]$ and $p_i = dG(T_i)$, $i = 1, \dots, n$, are (non-negative) jumps with total mass unity. Similar to the approach of Owen (1988; 1990) and Qin and Lawless (1994), it can be shown by using the method of Lagrange multipliers that the maximum value of \mathcal{L} , subject to constraints $\sum_{i=1}^n p_i = 1$, $p_i \geq 0$, $\sum_{i=1}^n p_i \{w(T_i) - 1\} = 0$, is attained at

$$\tilde{p}_i = \frac{1}{n_0} \frac{1}{1 + \rho \exp[\tilde{\alpha} + r(T_i)\tilde{\beta}]}, \tag{2.1}$$

where $\rho = n_1/n_0$ and $(\tilde{\alpha}, \tilde{\beta})$ is the maximum semi-parametric likelihood estimator of (α, β) obtained as a solution to the following system of score equations:

$$\frac{\partial l(\alpha, \beta)}{\partial \alpha} = n_1 - \sum_{i=1}^n \frac{\rho \exp[\alpha + r(T_i)\beta]}{1 + \rho \exp[\alpha + r(T_i)\beta]} = 0,$$

$$\frac{\partial l(\alpha, \beta)}{\partial \beta} = \sum_{j=1}^{n_1} r(Z_j) - \sum_{i=1}^n \frac{\rho \exp[\alpha + r(T_i)\beta]}{1 + \rho \exp[\alpha + r(T_i)\beta]} r(T_i) = 0,$$

with $l(\alpha, \beta)$ being the profile log-likelihood function of (α, β) given by

$$l(\alpha, \beta) = \sum_{j=1}^{n_1} [\alpha + r(Z_j)\beta] - \sum_{i=1}^n \log\{1 + \rho \exp[\alpha + r(T_i)\beta]\} - n \log n_0. \tag{2.2}$$

As a result, the maximum semi-parametric likelihood estimator of $G(t)$ is identical to

$$\tilde{G}(t) = \sum_{i=1}^n \tilde{p}_i I_{[T_i \leq t]} = \frac{1}{n_0} \sum_{i=1}^n \frac{I_{[T_i \leq t]}}{1 + \rho \exp[\tilde{\alpha} + r(T_i)\tilde{\beta}]}. \tag{2.3}$$

On the basis of \tilde{G} in (2.3), we propose, under model (1.1), to estimate the quantile function $G^{-1}(s)$ by $\tilde{G}^{-1}(s)$ for $s \in (0, 1)$. Throughout this paper, we refer to the function $\tilde{G}^{-1}(s)$ as the modified empirical quantile function and the process $\sqrt{n}(\tilde{G}^{-1} - G^{-1})$ as the modified quantile process.

Remark 2.1. The profile log-likelihood function $l(\alpha, \beta)$ in (2.2) can alternatively be written as

$$l(\alpha, \beta) = \sum_{i=1}^{n_0} [(-\alpha) + r(X_i)(-\beta)] - \sum_{i=1}^n \log\{1 + \rho^{-1} \exp[(-\alpha) + r(T_i)(-\beta)]\} - n \log n_1.$$

This can also be derived from the following two-sample semi-parametric model:

$$X_1, \dots, X_{n_0} \stackrel{\text{i.i.d.}}{\sim} g(x) = \exp[-\alpha - r(x)\beta]h(x), \quad Z_1, \dots, Z_{n_1} \stackrel{\text{i.i.d.}}{\sim} h(x). \tag{2.4}$$

Thus, the maximum semi-parametric likelihood estimator $(\tilde{\alpha}, \tilde{\beta})$ of (α, β) is the same under model (1.1) as that under model (2.4).

Remark 2.2. Remark 2.1 indicates that $H(t)$ can be estimated by

$$\tilde{H}(t) = \frac{1}{n_1} \sum_{i=1}^n \frac{1}{1 + \rho^{-1} \exp[-\tilde{\alpha} - r(T_i)\tilde{\beta}]} I_{[T_i \leq t]} = \sum_{i=1}^n \tilde{p}_i \exp[\tilde{\alpha} + r(T_i)\tilde{\beta}] I_{[T_i \leq t]}, \tag{2.5}$$

where \tilde{p}_i is defined in (2.1). On the basis of \tilde{H} in (2.5) and under model (1.1), the quantile function $H^{-1}(s)$ associated with H can be estimated by $\tilde{H}^{-1}(s)$ for $s \in (0, 1)$.

Remark 2.3. Our proposed procedure can also be applied to mixture sampling data in which a sample of $n = n_0 + n_1$ members is randomly selected from the whole population with both n_0 and n_1 being random.

We close this section by presenting the following theorem whose proof is similar to that in Prentice and Pyke (1979) and to that in Qin and Zhang (1997).

Theorem 2.1. Let (α_0, β_0) be the true value of (α, β) under model (1.1). Let

$$\begin{aligned}
 A_{11}(t) &= \int_{-\infty}^t \frac{\exp[\alpha_0 + r(x)\beta_0]}{1 + \rho \exp[\alpha_0 + r(x)\beta_0]} dG(x), & A_{11} &= A_{11}(\infty), \\
 A_{12}(t) &= \int_{-\infty}^t \frac{\exp[\alpha_0 + r(x)\beta_0]}{1 + \rho \exp[\alpha_0 + r(x)\beta_0]} r(x) dG(x), & A_{12} &= A_{12}(\infty), \\
 A_{22} &= \int \frac{\exp[\alpha_0 + r(x)\beta_0]}{1 + \rho \exp[\alpha_0 + r(x)\beta_0]} r(x)r(x)^T dG(x), \\
 A &= \begin{pmatrix} A_{11} & A_{12}^T \\ A_{12} & A_{22} \end{pmatrix}, & \Sigma &= \frac{1 + \rho}{\rho} \left[A^{-1} - \begin{pmatrix} 1 + \rho & 0 \\ 0 & 0 \end{pmatrix} \right], & S &= \frac{\rho}{1 + \rho} A.
 \end{aligned}$$

(a) If model (1.1) holds and A^{-1} exists, we can write

$$\begin{pmatrix} \tilde{\alpha} - \alpha_0 \\ \tilde{\beta} - \beta_0 \end{pmatrix} = \frac{1}{n} S^{-1} \begin{pmatrix} \frac{\partial l(\alpha_0, \beta_0)}{\partial \alpha} \\ \frac{\partial l(\alpha_0, \beta_0)}{\partial \beta} \end{pmatrix} + o_p(n^{-1/2}),$$

where

$$\frac{\partial l(\alpha_0, \beta_0)}{\partial \alpha} = \frac{\partial l(\alpha, \beta)}{\partial \alpha} \Big|_{(\alpha, \beta) = (\alpha_0, \beta_0)} \quad \text{and} \quad \frac{\partial l(\alpha_0, \beta_0)}{\partial \beta} = \frac{\partial l(\alpha, \beta)}{\partial \beta} \Big|_{(\alpha, \beta) = (\alpha_0, \beta_0)}.$$

As a result,

$$\sqrt{n} \begin{pmatrix} \tilde{\alpha} - \alpha_0 \\ \tilde{\beta} - \beta_0 \end{pmatrix} \xrightarrow{d} N_{p+1}(0, \Sigma).$$

(b) If model (1.1) holds and A^{-1} exists, we can write

$$\tilde{G}(t) - \hat{G}(t) = H_1(t) - H_2(t) - \hat{G}(t) + R_n(t), \tag{2.6}$$

where

$$H_1(t) = \frac{1}{n_0} \sum_{i=1}^n \frac{I_{[T_i \leq t]}}{1 + \rho \exp[\alpha_0 + r(T_i)\beta_0]}, \quad H_2(t) = \frac{\rho}{n} (A_{11}(t), A_{12}^T(t)) S^{-1} \begin{pmatrix} \frac{\partial l(\alpha_0, \beta_0)}{\partial \alpha} \\ \frac{\partial l(\alpha_0, \beta_0)}{\partial \beta} \end{pmatrix}, \tag{2.7}$$

and the remainder term $R_n(t)$ satisfies

$$\sup_{-\infty \leq t \leq \infty} |R_n(t)| = o_p(n^{-1/2}). \tag{2.8}$$

As a result,

$$\sqrt{n}(\tilde{G} - \hat{G}) \xrightarrow{\mathcal{D}} W \quad \text{in } D[-\infty, \infty],$$

where W is a Gaussian process with continuous sample path and satisfies, for $-\infty \leq s \leq t \leq \infty$,

$$\begin{aligned} EW(t) &= 0, \\ EW(s)W(t) &= \rho(1 + \rho)(A_{11}(s), A_{12}^T(s)) \left[\begin{pmatrix} 1 \\ 0 \end{pmatrix} - A^{-1} \begin{pmatrix} A_{11}(t) \\ A_{12}(t) \end{pmatrix} \right]. \end{aligned} \tag{2.9}$$

3. Asymptotic relative efficiency of $(\tilde{\alpha}, \tilde{\beta})$

In this section, we consider the asymptotic relative efficiency of the maximum semi-parametric likelihood estimator $(\tilde{\alpha}, \tilde{\beta})$ relative to the maximum likelihood estimator of (α, β) under the following two-sample parametric model in which (X_1, \dots, X_{n_0}) and (Z_1, \dots, Z_{n_1}) are independent and

$$\begin{aligned} X_1, \dots, X_{n_0} &\stackrel{\text{i.i.d.}}{\sim} g(x) = \exp[\theta_1 T(x) + d(\theta_1) + S(x)], \\ Z_1, \dots, Z_{n_1} &\stackrel{\text{i.i.d.}}{\sim} h(x) = \exp[\theta_2 T(x) + d(\theta_2) + S(x)], \end{aligned} \tag{3.1}$$

where θ_1 and θ_2 are real parameters and $d(\cdot)$, $T(\cdot)$, and $S(\cdot)$ are real-valued functions. In other words, we assume that the distributions of both X_1 and Z_1 belong to the one-parameter exponential family of distributions (Bickel and Doksum 1977, p. 67). It is easy to see that model (3.1) is a special case of model (1.1) with $\alpha = d(\theta_2) - d(\theta_1)$, $\beta = \theta_2 - \theta_1$ and $r(x) = T(x)$. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ denote, respectively, the maximum likelihood estimators of θ_1 and θ_2 under model (3.1). Then according to the standard results on maximum likelihood estimation in exponential families, we can show that $\sqrt{n_0}(\hat{\theta}_1 - \theta_1) \xrightarrow{d} N(0, [-d''(\theta_1)]^{-1})$ and $\sqrt{n_1}(\hat{\theta}_2 - \theta_2) \xrightarrow{d} N(0, [-d''(\theta_2)]^{-1})$. Furthermore, according to the invariance property of maximum likelihood estimation, the maximum likelihood estimators of α and β under model (3.1) are, respectively, given by $\hat{\alpha} = d(\hat{\theta}_2) - d(\hat{\theta}_1)$ and $\hat{\beta} = \hat{\theta}_2 - \hat{\theta}_1$. Since $\hat{\theta}_1$ and $\hat{\theta}_2$ are independent, it can be shown after some algebra that $\sqrt{n}(\hat{\alpha} - \alpha_0) \xrightarrow{d} N(0, \sigma_{\hat{\alpha}}^2)$ and $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \sigma_{\hat{\beta}}^2)$, where (α_0, β_0) is the true value of (α, β) under model (3.1) and

$$\sigma_{\hat{\alpha}}^2 = -(1 + \rho) \left(\frac{[d'(\theta_1)]^2}{d''(\theta_1)} + \frac{1}{\rho} \frac{[d'(\theta_2)]^2}{d''(\theta_2)} \right), \quad \sigma_{\hat{\beta}}^2 = -(1 + \rho) \left(\frac{1}{d''(\theta_1)} + \frac{1}{\rho} \frac{1}{d''(\theta_2)} \right). \tag{3.2}$$

Equations (3.2), along with part (a) of Theorem 2.1, imply that the asymptotic relative efficiency of $\tilde{\alpha}$ relative to $\hat{\alpha}$ is given by

$$e(\tilde{\alpha}, \hat{\alpha}) = \left(\frac{[d'(\theta_1)]^2}{d''(\theta_1)} + \frac{1}{\rho} \frac{[d'(\theta_2)]^2}{d''(\theta_2)} \right) / \left(\frac{1 + \rho}{\rho} - \frac{1}{\rho} \frac{A_{22}}{A_{11}A_{22} - A_{12}^2} \right) \tag{3.3}$$

and the asymptotic relative efficiency of $\tilde{\beta}$ relative to $\hat{\beta}$ is given by

$$e(\tilde{\beta}, \hat{\beta}) = -\left(\frac{1}{d''(\theta_1)} + \frac{1}{\rho} \frac{1}{d''(\theta_2)}\right) \bigg/ \left(\frac{1}{\rho} \frac{A_{11}}{A_{11}A_{22} - A_{12}^2}\right), \tag{3.4}$$

where A_{11} , A_{12} , and A_{22} are defined in Theorem 2.1 with $r(x) = T(x)$.

Example 3.1. Let $X_1, \dots, X_{n_0} \stackrel{i.i.d.}{\sim} N(\theta_1, 1)$ and, independent of the X_i , let $Z_1, \dots, Z_{n_1} \stackrel{i.i.d.}{\sim} N(\theta_2, 1)$. Then model (3.1) holds with $d(\theta_1) = -\theta_1^2/2$, $d(\theta_2) = -\theta_2^2/2$, $T(x) = x$, $S(x) = -\{x^2 + \log(2\pi)\}/2$ and model (1.1) holds with $\alpha = (\theta_1^2 - \theta_2^2)/2$ and $\beta = \theta_2 - \theta_1$. According to (3.3) and (3.4), the asymptotic relative efficiency of $\tilde{\alpha}$ relative to $\hat{\alpha}$ is given by

$$e(\tilde{\alpha}, \hat{\alpha}) = (\rho\theta_1^2 + \theta_2^2) \bigg/ \left(\frac{A_{22}}{A_{11}A_{22} - A_{12}^2} - (1 + \rho)\right)$$

and the asymptotic relative efficiency of $\tilde{\beta}$ relative to $\hat{\beta}$ is given by

$$e(\tilde{\beta}, \hat{\beta}) = (1 + \rho) \bigg/ \frac{A_{11}}{A_{11}A_{22} - A_{12}^2},$$

where A_{11} , A_{12} , and A_{22} are defined in Theorem 2.1 with $r(x) = x$. It can be shown that if $\alpha = -\beta^2/2$, then $e(\tilde{\beta}, \hat{\beta})$ remains the same when ρ is replaced by ρ^{-1} . Table 1 presents some numerical values of $e(\tilde{\alpha}, \hat{\alpha})$ and $e(\tilde{\beta}, \hat{\beta})$ when $\theta_1 = 0$, $\theta_2 = 1$ and $n = 60$. As expected, the maximum likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ are more efficient than the maximum semi-parametric likelihood estimators $\tilde{\alpha}$ and $\tilde{\beta}$ and the values of $e(\tilde{\beta}, \hat{\beta})$ are the same for ρ and ρ^{-1} .

Example 3.2. Let $X_1, \dots, X_{n_0} \stackrel{i.i.d.}{\sim} g(x) = \theta_1 \exp(-\theta_1 x)$ and, independent of the X_i , let $Z_1, \dots, Z_{n_1} \stackrel{i.i.d.}{\sim} h(x) = \theta_2 \exp(-\theta_2 x)$. Then model (3.1) holds with $d(\theta_1) = \log(\theta_1)$, $d(\theta_2) = \log(\theta_2)$, $T(x) = -x$ and $S(x) = 0$, and model (1.1) holds with $\alpha = \log(\theta_2) - \log(\theta_1)$ and $\beta = \theta_2 - \theta_1$. According to (3.3) and (3.4), the asymptotic relative efficiency of $\tilde{\alpha}$ relative to $\hat{\alpha}$ is given by

$$e(\tilde{\alpha}, \hat{\alpha}) = (1 + \rho) \bigg/ \left(\frac{A_{22}}{A_{11}A_{22} - A_{12}^2} - (1 + \rho)\right)$$

Table 1. Asymptotic relative efficiencies $e(\tilde{\alpha}, \hat{\alpha})$ and $e(\tilde{\beta}, \hat{\beta})$ when $\theta_1 = 0$, $\theta_2 = 1$ and $n = 60$

(n_0, n_1)	ρ	$e(\tilde{\alpha}, \hat{\alpha})$	$e(\tilde{\beta}, \hat{\beta})$
(50, 10)	0.2	0.884 74	0.742 10
(40, 20)	0.5	0.829 44	0.674 49
(30, 30)	1.0	0.784 88	0.656 71
(20, 40)	2.0	0.741 21	0.674 49
(10, 50)	5.0	0.689 18	0.742 10

and the asymptotic relative efficiency of $\tilde{\beta}$ relative to $\hat{\beta}$ is given by

$$e(\tilde{\beta}, \hat{\beta}) = (\rho\theta_1^2 + \theta_2^2) / \frac{A_{11}}{A_{11}A_{22} - A_{12}^2},$$

where A_{11} , A_{12} , and A_{22} are defined in Theorem 2.1 with $r(x) = -x$. Table 2 presents some numerical values of $e(\tilde{\alpha}, \hat{\alpha})$ and $e(\tilde{\beta}, \hat{\beta})$ when $\theta_1 = 1$, $\theta_2 = 2$ and $n = 60$. As expected, the maximum likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ are more efficient than the maximum semi-parametric likelihood estimators $\tilde{\alpha}$ and $\tilde{\beta}$. It is seen that the values of $e(\tilde{\beta}, \hat{\beta})$, ranging from 0.67 to 0.92, are slightly more variable than those of $e(\tilde{\alpha}, \hat{\alpha})$, ranging from 0.71 to 0.87.

4. Asymptotic results

In this section, we study the asymptotic properties of the modified empirical quantile function $\tilde{G}^{-1}(s)$. To this end, let (α_0, β_0) be the true value of (α, β) under model (1.1). We assume that $\rho = n_1/n_0$ is positive and finite and remains fixed as $n = n_0 + n_1 \rightarrow \infty$.

We first establish the weak convergence of the modified quantile process $\sqrt{n}(\tilde{G}^{-1} - G^{-1})$ to a Gaussian process on a subinterval of $[0, 1]$ by expressing \tilde{G} as the mean of a sequence of independent and identically distributed stochastic processes with a remainder term of order $o_p(n^{-1/2})$.

Theorem 4.1. *Let $0 < a < b < 1$ be given. Suppose that the conditions of Theorem 2.1 hold. Suppose further that G has continuous positive density g on $[G^{-1}(a) - \varepsilon, G^{-1}(b) + \varepsilon]$ for some $\varepsilon > 0$. Then, one can write*

$$\tilde{G}^{-1}(s) - G^{-1}(s) = Q_n(s) + r_n(s), \tag{4.1}$$

where

$$Q_n(s) = - \frac{H_1(G^{-1}(s)) - H_2(G^{-1}(s)) - s}{g(G^{-1}(s))} \tag{4.2}$$

and the remainder term $r_n(s)$ satisfies

$$\sup_{a \leq s \leq b} |r_n(s)| = o_p(n^{-1/2}). \tag{4.3}$$

Table 2. Asymptotic relative efficiencies $e(\tilde{\alpha}, \hat{\alpha})$ and $e(\tilde{\beta}, \hat{\beta})$ when $\theta_1 = 1$, $\theta_2 = 2$ and $n = 60$

(n_0, n_1)	ρ	$e(\tilde{\alpha}, \hat{\alpha})$	$e(\tilde{\beta}, \hat{\beta})$
(50, 10)	0.2	0.870 92	0.921 36
(40, 20)	0.5	0.786 96	0.847 83
(30, 30)	1.0	0.733 69	0.778 69
(20, 40)	2.0	0.706 86	0.715 73
(10, 50)	5.0	0.715 65	0.671 43

As a result,

$$\sqrt{n}(\tilde{G}^{-1} - G^{-1}) \xrightarrow{\mathcal{D}} V(G^{-1})/g(G^{-1}) \quad \text{on } D[a, b], \tag{4.4}$$

where V is a Gaussian process with continuous sample path and satisfies, for $-\infty \leq s \leq t \leq \infty$,

$$EV(t) = 0,$$

$$EV(s)V(t) = (1 + \rho) \left\{ G(s)[1 - G(t)] - \rho(A_{11}(s), A_{12}^T(s)) \left[\begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} - A^{-1} \begin{pmatrix} A_{11}(t) \\ A_{12}(t) \end{pmatrix} \right] \right\}. \tag{4.5}$$

Theorem 4.1 enables us to establish the asymptotic normality of the sample quantiles associated with \tilde{G} . We present two results. The first concerns the asymptotic joint distribution of k sample quantiles, whereas the second pertains to the asymptotic distribution of a single sample quantile $\tilde{\xi}_q = \tilde{G}^{-1}(q)$.

Corollary 4.1. *Let $0 < q_1 < \dots < q_k < 1$. Suppose that the conditions of Theorem 2.1 hold. Suppose further that G has continuous positive density g in neighbourhoods of $G^{-1}(q_1), \dots, G^{-1}(q_k)$. Then as $n \rightarrow \infty$,*

$$\sqrt{n}(\tilde{G}^{-1}(q_1) - G^{-1}(q_1), \dots, \tilde{G}^{-1}(q_k) - G^{-1}(q_k)) \xrightarrow{d} N_k(\mathbf{0}, \mathbf{B}),$$

where $\mathbf{0} = (0, 0, \dots, 0)^T$ and $\mathbf{B} = (b_{ij})_{k \times k}$ with

$$b_{ij} = \frac{1 + \rho}{g(G^{-1}(q_i))g(G^{-1}(q_j))} \times \left\{ q_i(1 - q_j) - \rho[A_{11}(G^{-1}(q_i)), A_{12}^T(G^{-1}(q_j))] \left[\begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} - A^{-1} \begin{pmatrix} A_{11}(G^{-1}(q_j)) \\ A_{12}(G^{-1}(q_j)) \end{pmatrix} \right] \right\} \tag{4.6}$$

and $b_{ij} = b_{ji}$ for $i > j$.

Corollary 4.2. *Let $0 < q < 1$. Suppose that the conditions of Theorem 2.1 hold. Suppose further that G has a continuous positive density g in a neighbourhood of $G^{-1}(q)$. Then, as $n \rightarrow \infty$,*

$$\sqrt{n}(\tilde{G}^{-1}(q) - G^{-1}(q)) \xrightarrow{d} N(0, \sigma_{\tilde{\xi}}^2),$$

where

$$\sigma_{\tilde{\xi}}^2 = \frac{(1 + \rho)}{[g(G^{-1}(q))]^2} \left\{ q(1 - q) - \rho[A_{11}(G^{-1}(q)), A_{12}^T(G^{-1}(q))] \left[\begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} - A^{-1} \begin{pmatrix} A_{11}(G^{-1}(q)) \\ A_{12}(G^{-1}(q)) \end{pmatrix} \right] \right\}. \tag{4.7}$$

Remark 4.1. The results in Theorem 4.1 and its corollaries also hold for $\tilde{H}^{-1}(s)$ defined in Remark 2.2. We only need to replace ρ by ρ^{-1} and (α_0, β_0) by $(-\alpha_0, -\beta_0)$ in (4.5)–(4.7).

Next we consider the asymptotic relative efficiency of the standard quantile estimator $\hat{\xi}_q = \hat{G}^{-1}(q)$ relative to the proposed quantile estimator $\tilde{\xi}_q = \tilde{G}^{-1}(q)$. On the basis of the (control) sample X_1, \dots, X_{n_0} , let

$$\sigma_{\xi}^2 = \frac{q(1-q)}{[g(G^{-1}(q))]^2}$$

be the asymptotic variance of $\sqrt{n_0}\hat{\xi}_q$. If $\lim_{n_0 \rightarrow \infty} \text{var}[\sqrt{n_0}\hat{\xi}_q] = \sigma_{\xi}^2$ and $\lim_{n \rightarrow \infty} \text{var}[\sqrt{n}\tilde{\xi}_q] = \sigma_{\tilde{\xi}}^2$, then

$$\lim_{n \rightarrow \infty} \frac{\text{var}(\hat{\xi}_q)}{\text{var}(\tilde{\xi}_q)} = (1 + \rho) \lim_{n \rightarrow \infty} \frac{\text{var}[\sqrt{n_0}\hat{\xi}_q]}{\text{var}[\sqrt{n}\tilde{\xi}_q]} = (1 + \rho) \frac{\sigma_{\xi}^2}{\sigma_{\tilde{\xi}}^2}.$$

This motivates the following definitions regarding the asymptotic relative efficiency of $\hat{\xi}_q$ relative to $\tilde{\xi}_q$. Under the conditions of Corollary 4.2, we define the asymptotic relative efficiency of $\hat{\xi}_q$ relative to $\tilde{\xi}_q$ by

$$\begin{aligned} e(\hat{\xi}_q, \tilde{\xi}_q) &= \frac{\sigma_{\xi}^2}{(1 + \rho)\sigma_{\tilde{\xi}}^2} \\ &= 1 - \frac{1}{q(1-q)} \left\{ \rho [A_{11}(G^{-1}(q)), A_{12}^T(G^{-1}(q))] \left[\begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} - A^{-1} \begin{pmatrix} A_{11}(G^{-1}(q)) \\ A_{12}(G^{-1}(q)) \end{pmatrix} \right] \right\}. \end{aligned}$$

The following corollary, which is a straightforward consequence of (2.9), indicates that the proposed quantile estimator $\tilde{\xi}_q$ has smaller asymptotic variance than the standard quantile estimator $\hat{\xi}_q$. Thus, $\tilde{\xi}_q$ is more efficient than $\hat{\xi}_q$ under model (1.1).

Corollary 4.3. *Under the conditions of Corollary 4.2, we have $e(\hat{\xi}_q, \tilde{\xi}_q) \leq 1$.*

Finally, we consider estimating the asymptotic variance σ_{ξ}^2 given in Corollary 4.2. On the basis of \tilde{G} in (2.3), we propose to estimate σ_{ξ}^2 by

$$\begin{aligned} \tilde{\sigma}_{\xi}^2 &= (1 + \rho) \left[\frac{\tilde{G}(\tilde{\xi}_q + b_n) - \tilde{G}(\tilde{\xi}_q - b_n)}{2b_n} \right]^{-2} \\ &\quad \times \left\{ q(1-q) - \rho [\tilde{A}_{11}(\tilde{\xi}_q), \tilde{A}_{12}^T(\tilde{\xi}_q)] \left[\begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} - \tilde{A}^{-1} \begin{pmatrix} \tilde{A}_{11}(\tilde{\xi}_q) \\ \tilde{A}_{12}(\tilde{\xi}_q) \end{pmatrix} \right] \right\}, \end{aligned} \tag{4.8}$$

where

$$\begin{aligned} \tilde{A}_{11}(t) &= \sum_{i=1}^n \tilde{p}_i \frac{\exp[\tilde{\alpha} + r(T_i)\tilde{\beta}]}{1 + \rho \exp[\tilde{\alpha} + r(T_i)\tilde{\beta}]} I_{[T_i \leq t]} = \frac{1}{n_0} \sum_{i=1}^n \frac{\exp[\tilde{\alpha} + r(T_i)\tilde{\beta}]}{\{1 + \rho \exp[\tilde{\alpha} + r(T_i)\tilde{\beta}]\}^2} I_{[T_i \leq t]}, \\ \tilde{A}_{12}(t) &= \sum_{i=1}^n \tilde{p}_i \frac{\exp[\tilde{\alpha} + r(T_i)\tilde{\beta}]}{1 + \rho \exp[\tilde{\alpha} + r(T_i)\tilde{\beta}]} r(T_i) I_{[T_i \leq t]} = \frac{1}{n_0} \sum_{i=1}^n \frac{\exp[\tilde{\alpha} + r(T_i)\tilde{\beta}]}{\{1 + \rho \exp[\tilde{\alpha} + r(T_i)\tilde{\beta}]\}^2} r(T_i) I_{[T_i \leq t]}, \\ \tilde{A}_{11} &= \tilde{A}_{11}(\infty), \quad \tilde{A}_{12} = \tilde{A}_{12}(\infty), \\ \tilde{A}_{22} &= \sum_{i=1}^n \tilde{p}_i \frac{\exp[\tilde{\alpha} + r(T_i)\tilde{\beta}]}{1 + \rho \exp[\tilde{\alpha} + r(T_i)\tilde{\beta}]} r(T_i) r^T(T_i) = \frac{1}{n_0} \sum_{i=1}^n \frac{\exp[\tilde{\alpha} + r(T_i)\tilde{\beta}]}{\{1 + \rho \exp[\tilde{\alpha} + r(T_i)\tilde{\beta}]\}^2} r(T_i) r^T(T_i), \\ \tilde{A} &= \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12}^T \\ \tilde{A}_{12} & \tilde{A}_{22} \end{pmatrix}. \end{aligned}$$

Note that $\tilde{g}(t) \equiv (2b_n)^{-1}[\tilde{G}(t + b_n) - \tilde{G}(t - b_n)]$ is a special case of the following kernel estimator of the density function $g(t)$ of $G(t)$,

$$\tilde{g}(t) = \frac{1}{b_n} \int K\left(\frac{t - y}{b_n}\right) d\tilde{G}(y),$$

when K is the uniform kernel function, that is, $K(x) = \frac{1}{2}$ if $x \in [-1, 1)$ and $K(x) = 0$ if $x \notin [-1, 1)$.

Under the conditions given in Corollary 4.2, it can be shown that

$$\tilde{A}_{11}(\tilde{\xi}_q) = A_{11}(G^{-1}(q)) + o_p(1), \quad \tilde{A}_{12}(\tilde{\xi}_q) = A_{12}(G^{-1}(q)) + o_p(1), \quad \tilde{A} = A + o_p(1).$$

Moreover, if $nb_n^2 \rightarrow \infty$ as $n \rightarrow \infty$, then we can show under the conditions of Corollary 4.2 that

$$\frac{\tilde{G}(\tilde{\xi}_q + b_n) - \tilde{G}(\tilde{\xi}_q - b_n)}{2b_n} = g(G^{-1}(q)) + o_p(b_n) + O_p(n^{-1/2}b_n^{-1}).$$

Therefore, we have the following result.

Theorem 4.2. *Under the conditions of Corollary 4.2 and the condition that $nb_n^2 \rightarrow \infty$ as $n \rightarrow \infty$, we have*

$$\tilde{\sigma}_{\tilde{\xi}_q}^2 = \sigma_{\tilde{\xi}_q}^2 + o_p(1),$$

where $\tilde{\sigma}_{\tilde{\xi}_q}^2$ is given by (4.8).

5. Numerical results

In this section, we report the results of a real data problem and a simulation study.

5.1. An example

By employing the logistic regression model (1.2), Hosmer and Lemeshow (1989) analysed the relationship between age and coronary heart disease status based on a study of 100 subjects. The complete data set is listed on page 3 in their book. Let X denote age and $Y = 1$ or 0 represent the presence or absence of coronary heart disease; then the sample data (X_i, Y_i) , $i = 1, \dots, 100$, can be thought of as being drawn independently and identically from the joint distribution of (X, Y) . Remark 2.3 implies that our proposed procedure can be applied to this data set. By fitting model (1.1), we have $(\tilde{\alpha}, \tilde{\beta}) = (-5.02760, 0.11092)$. Since $n_0 = 57$ and $n_1 = 43$, we can estimate α^* in model (1.2) by $\tilde{\alpha}^* = -5.02760 - \log(57/43) = -5.309451$, which coincides with Hosmer and Lemeshow's result. Figure 1 shows the curves of four estimated quantile functions $\tilde{G}^{-1}(s)$, $\hat{G}^{-1}(s)$, $\tilde{H}^{-1}(s)$, and $\hat{H}^{-1}(s)$ for $s \in (0, 1)$. The curve of $\tilde{G}^{-1}(s)$ ($\tilde{H}^{-1}(s)$) bears a resemblance to that of $\hat{G}^{-1}(s)$ ($\hat{H}^{-1}(s)$). In addition, the left-hand panel of Figure 1 reveals that the median age in the disease-free population is estimated to be $\tilde{G}^{-1}(0.5) = \hat{G}^{-1}(0.5) = 38$, whereas the right-hand panel of Figure 1 indicates that the median age in the diseased population is estimated to be $\tilde{H}^{-1}(0.5) = 53$ or $\hat{H}^{-1}(0.5) = 54$. According to Corollary 4.3, Theorem 4.2 and equation (4.8), an approximate 95% confidence interval for $G^{-1}(0.5)$ is given by $\tilde{I} = (34.16, 41.84)$ on the basis of \tilde{G} . On the other hand, based on \hat{G} and $\hat{\sigma}_{\hat{\xi}}^2 = [\{\hat{G}(\hat{\xi}_{0.5} + b_{n_0}) - \hat{G}(\hat{\xi}_{0.5} - b_{n_0})\}/b_{n_0}]^{-2}$, an approximate 95% confidence interval for $G^{-1}(0.5)$ is given by $\hat{I} = (34.05, 41.95)$. It is seen that \tilde{I} is narrower than \hat{I} , indicating that statistical inference based on \tilde{G} is more accurate than that based on \hat{G} . In computing \tilde{I} and \hat{I} , we chose $b_n = b_{n_0} = 4.0$ so that the curve of \tilde{g} resembles the curve of $\hat{g}(t) = (1/b_{n_0}) \int K((t - y)/b_{n_0}) d\hat{G}(y)$.

5.2. A simulation study

In our simulation study, we consider estimating the first quartile $G^{-1}(0.25)$ of G . We assume that $g(x)$ is the standard normal density function and $h(x)$ is the density function of an $N(\mu, 1)$ distribution. It is easy to see that $g(x)$ and $h(x)$ are related by $h(x) = \exp(\alpha + x\beta)g(x)$, where $\alpha = \mu^2/2$ and $\beta = \mu$. Thus, model (1.1) holds with $r(x) = x$. Let $\mu = 1$ be fixed. Our aim is to compare the performance of $\hat{\xi}_{0.25} = \hat{G}^{-1}(0.25)$ with that of $\tilde{\xi}_{0.25} = \tilde{G}^{-1}(0.25)$ by examining their biases, variances and relative efficiencies. In our simulations, we generated, for each pair (n_0, n_1) , 1000 independent sets of combined random samples of size $n = 60$ from the $N(0, 1)$ and $N(1, 1)$ populations. The simulation results are summarized in Table 3.

In Table 3, $\text{bias}(\hat{\xi}_{0.25})$ and $\text{var}(\hat{\xi}_{0.25})$ stand for, respectively, the average of 1000 biases of $\hat{\xi}_{0.25}$ and the sample variance of 1000 estimates $\hat{\xi}_{0.25}$, whereas $\text{bias}(\tilde{\xi}_{0.25})$ and $\text{var}(\tilde{\xi}_{0.25})$ stand for, respectively, the average of 1000 biases of $\tilde{\xi}_{0.25}$ and the sample variance of 1000 estimates $\tilde{\xi}_{0.25}$. In addition, we use $\widehat{\text{var}}(\hat{\xi}_{0.25})$ and $\widehat{\text{var}}(\tilde{\xi}_{0.25})$ to represent, respectively, the averages of 1000 variance estimates of $\hat{\xi}_{0.25}$ and of $\tilde{\xi}_{0.25}$. Moreover, $e(\hat{\xi}_{0.25}, \tilde{\xi}_{0.25})$ stands for the relative efficiency of $\hat{\xi}_{0.25}$ relative to $\tilde{\xi}_{0.25}$, that is $e(\hat{\xi}_{0.25}, \tilde{\xi}_{0.25}) = \text{var}(\tilde{\xi}_{0.25})/\text{var}(\hat{\xi}_{0.25})$. Finally, in calculating variance estimates, we have employed (4.8) with $q = 0.25$ to estimate the asymptotic variance of $\tilde{\xi}_{0.25}$, and have employed $\hat{\sigma}_{\hat{\xi}}^2 = [\{\hat{G}(\hat{\xi}_{0.25} + b_{n_0}) - \hat{G}(\hat{\xi}_{0.25} -$

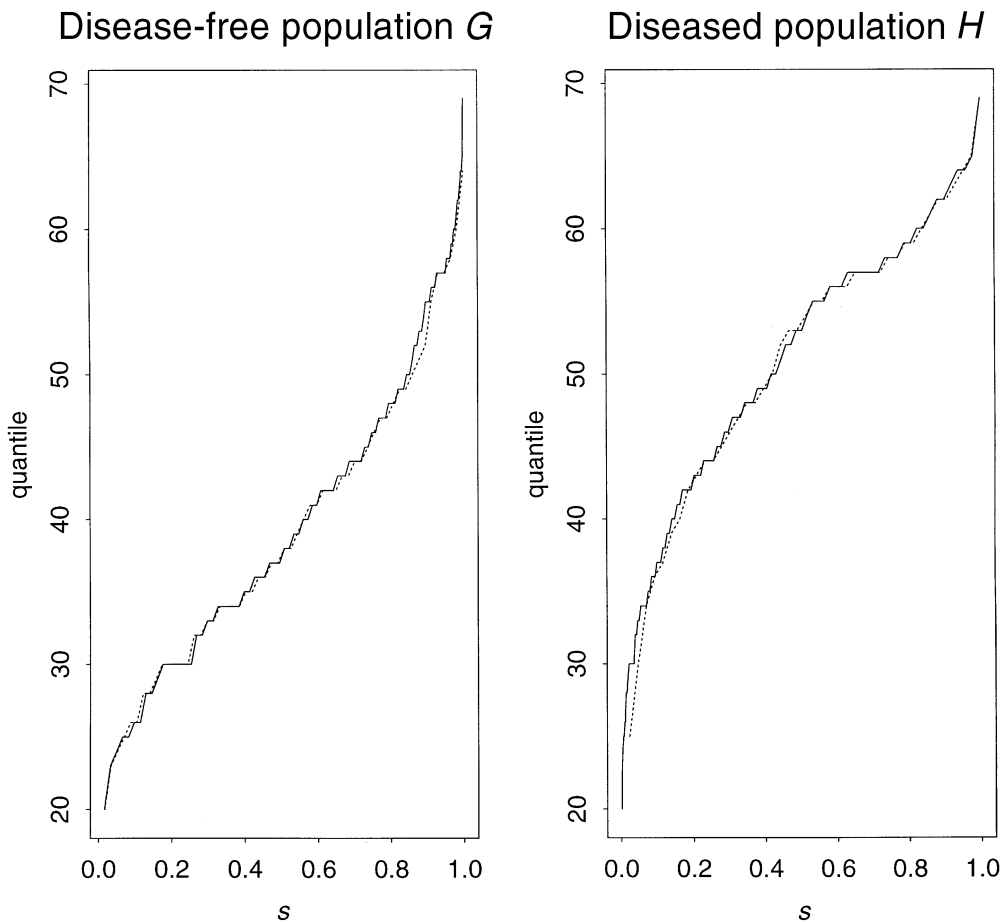


Figure 1. Estimated quantile functions $\tilde{G}^{-1}(s)$, $\hat{G}^{-1}(s)$, $\tilde{H}^{-1}(s)$, and $\hat{H}^{-1}(s)$ for $s \in (0, 1)$. The solid curve and dashed curve on the left panel represent $\tilde{G}^{-1}(s)$ and $\hat{G}^{-1}(s)$, respectively, while the solid curve and dashed curve on the right panel stand for $\tilde{H}^{-1}(s)$ and $\hat{H}^{-1}(s)$, respectively.

$b_{n_0}\}/b_{n_0}]^{-2}$ to estimate the asymptotic variance of $\hat{\xi}_{0.25}$. In this simulation study, the bandwidths b_n and b_{n_0} were selected according to the optimal bandwidth $h_{opt} = k_2^{-2/5}[\int K^2(x)]^{1/5}[\int\{f''(x)\}^2 dx]^{-1/5}n^{-1/5}$ in the usual kernel density estimation (Silverman 1986, p. 40), where $f(\cdot)$ is the density function of the population distribution and $k_2 = \int x^2 K(x) dx$ with K being the uniform kernel function given in Section 4. It can be shown that $h_{opt} = (12\sqrt{\pi}/n)^{1/5}$ when the population distributions is $N(0, 1)$. Consequently, we take $b_n = (12\sqrt{\pi}/n)^{1/5}$ and $b_{n_0} = (12\sqrt{\pi}/n_0)^{1/5}$ in our simulations.

It is seen from Table 3 that the biases and variances of $\hat{\xi}_{0.25}$ are all smaller than those of

Table 3. Biases, variances, and efficiencies of $\tilde{\xi}_{0.25}$ and $\hat{\xi}_{0.25}$

(n_0, n_1)	ρ	$\text{bias}(\hat{\xi}_{0.25})$	$\text{bias}(\tilde{\xi}_{0.25})$	$\text{var}(\hat{\xi}_{0.25})$	$\text{var}(\tilde{\xi}_{0.25})$	$\widehat{\text{var}}(\hat{\xi}_{0.25})$	$\widehat{\text{var}}(\tilde{\xi}_{0.25})$	$e(\hat{\xi}_{0.25}, \tilde{\xi}_{0.25})$
(50, 10)	0.2	0.014 08	0.013 02	0.038 13	0.037 20	0.043 10	0.041 20	0.975 61
(40, 20)	0.5	0.046 19	0.008 32	0.047 42	0.043 00	0.053 22	0.049 91	0.906 79
(30, 30)	1	0.007 99	0.005 40	0.065 14	0.056 80	0.077 38	0.063 77	0.871 97
(20, 40)	2	-0.084 44	0.001 02	0.097 13	0.077 21	0.135 85	0.090 11	0.794 91
(10, 50)	5	0.035 56	0.025 82	0.172 38	0.127 56	0.278 70	0.152 09	0.739 99

$\hat{\xi}_{0.25}$ for various values of (n_0, n_1) . By examining the magnitude of the relative efficiency $e(\hat{\xi}_{0.25}, \tilde{\xi}_{0.25})$, we see that the improvement of the proposed first quartile estimator $\tilde{\xi}_{0.25}$ over the standard first quartile estimator $\hat{\xi}_{0.25}$ in terms of variance reduction is more apparent as $\rho = n_1/n_0$ gets larger. This phenomenon is not unexpected because, on the basis of the (control) sample X_1, \dots, X_{n_0} , we would expect to estimate $G^{-1}(0.25)$ by $\hat{\xi}_{0.25}$ less accurately when n_0 gets smaller. With regard to the variance estimates, Table 3 indicates that the proposed estimator $\tilde{\sigma}_{\text{var}}^2$ in (4.8) works well in most cases. This simulation study indicates that when the second sample Z_1, \dots, Z_{n_1} is available, it is advantageous to employ the proposed quantile estimate $\tilde{G}^{-1}(s)$ to estimate $G^{-1}(s)$ because $\tilde{G}^{-1}(s)$ improves $\hat{G}^{-1}(s)$ in terms of variance reduction for various values of ρ and such an improvement is expected to be even more obvious for $\rho \geq 1$.

6. Proofs

First we present a lemma which will be used in the proof of Theorem 4.1.

Lemma 6.1. *Let $0 < a < b < 1$ be given. Suppose that model (1.1) holds and the conditions of Theorem 2.1 hold.*

- (a) $\sup_{-\infty \leq t \leq \infty} |\tilde{G}(t) - G(t)| = O_p(n^{-1/2})$.
- (b) *If G has continuous positive density g on $[G^{-1}(a), G^{-1}(b)]$, then $\sup_{a \leq s \leq b} |\tilde{G}^{-1}(s) - G^{-1}(s)| = O_p(n^{-1/2})$.*
- (c) $\sup_{a \leq s \leq b} |\tilde{G}(\tilde{G}^{-1}(s)) - s| = O_p(n^{-1})$.
- (d) *If G has continuous positive density g on $[G^{-1}(a) - \varepsilon, G^{-1}(b) + \varepsilon]$ for some $\varepsilon > 0$, then for any $C > 0$, $\sup_{G^{-1}(a) \leq x \leq G^{-1}(b)} \sup_{|y-x| \leq Cn^{-1/2}} |\tilde{G}(x) - \tilde{G}(y) - G(x) + G(y)| = o_p(n^{-1/2})$.*

Proof. Part (a) is a straightforward consequence of part (b) of Theorem 2.1 along with $\tilde{G}(t) - G(t) = \hat{G}(t) - \hat{G}(t) + [\hat{G}(t) - G(t)]$ and $\sup_{-\infty \leq t \leq \infty} |\hat{G}(t) - G(t)| = O_p(n^{-1/2})$.

For part (b), since $\max_{1 \leq i \leq n} \tilde{p}_i \leq 1/n_0 = (1 + \rho)/n$ by (2.1), it follows from part (a) that

$$\begin{aligned} \Delta_n &\equiv \max_{1 \leq i \leq n} \sup_{\tilde{G}(T_{(i-1)}) < s \leq \tilde{G}(T_{(i)})} |s - G(T_{(i)})| \leq \max_{1 \leq i \leq n} |\tilde{G}(T_{(i)}) - G(T_{(i)})| + \max_{1 \leq i \leq n} |\tilde{G}(T_{(i-1)}) - G(T_{(i)})| \\ &\leq 2 \sup_{-\infty < t < \infty} |\tilde{G}(t) - G(t)| + \max_{1 \leq i \leq n} \tilde{p}_i = O_p(n^{-1/2}), \end{aligned} \tag{6.1}$$

where $T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(n)}$ are the order statistics of T_1, \dots, T_n and $T_{(0)} = -\infty$. Since G has continuous positive density g on $[G^{-1}(a), G^{-1}(b)]$, $g(G^{-1}(s))$ is bounded away from zero on $[a, b]$, which, along with (6.1), implies that

$$\begin{aligned} \sup_{a \leq s \leq b} |\tilde{G}^{-1}(s) - G^{-1}(s)| &\leq \max_{1 \leq i \leq n} \sup_{\tilde{G}(T_{(i-1)}) < s \leq \tilde{G}(T_{(i)})} |T_{(i)} - G^{-1}(s)| \\ &= \max_{1 \leq i \leq n} \sup_{\tilde{G}(T_{(i-1)}) < s \leq \tilde{G}(T_{(i)})} |G^{-1}(G(T_{(i)})) - G^{-1}(s)| \\ &\leq \left[\sup_{a \leq s \leq b} \frac{1}{g(G^{-1}(s))} \right] \Delta_n + o_p(\Delta_n) = O_p(n^{-1/2}). \end{aligned}$$

For part (c), we have

$$\begin{aligned} \sup_{a \leq s \leq b} |\tilde{G}(\tilde{G}^{-1}(s)) - s| &\leq \max_{1 \leq i \leq n} \sup_{\tilde{G}(T_{(i-1)}) < s \leq \tilde{G}(T_{(i)})} |\tilde{G}(T_{(i)}) - s| \\ &\leq \max_{1 \leq i \leq n} |\tilde{G}(T_{(i)}) - \tilde{G}(T_{(i-1)})| = \max_{1 \leq i \leq n} \tilde{p}_i = O_p(n^{-1}). \end{aligned}$$

Turning to the proof of part (d), let $\delta_n = Cn^{-1/2}$ and $\{m_n\}$ be a sequence of positive integers such that $m_n = \lfloor n^\theta \rfloor$ with $\theta > \frac{1}{2}$, where $\lfloor n^\theta \rfloor$ denotes the largest integer less or equal to n^θ . We first divide the interval $[[m_n G^{-1}(a)]/m_n, [m_n G^{-1}(b) + 1]/m_n]$ into subintervals $I_i = [t_i, t_{i+1}]$ for $i = 0, 1, \dots, d_n - 1$, where $t_k = ([m_n G^{-1}(a)] + k)/m_n$ for $k = 0, 1, \dots, d_n$ with $d_n = [m_n G^{-1}(b) + 1] - [m_n G^{-1}(a)]$. We then subdivide each interval $[t_i - [m_n \delta_n + 1]/m_n, t_i + [m_n \delta_n + 1]/m_n]$ for $i = 0, 1, \dots, d_n$ into subintervals $I_{ij} = [t_{ij}, t_{i(j+1)}]$ for $j = -b_n, \dots, b_n - 1$, where $t_{ij} = t_i + j/m_n$ for $j = -b_n, b_n + 1, \dots, b_n - 1, b_n$ with $b_n = [m_n \delta_n + 1]$. Now let

$$\begin{aligned} a_i &= G(t_{i+1}) - G(t_i), \quad a_{ij} = G(t_{i(j+1)}) - G(t_{ij}), \quad i = 0, \dots, d_n, j = -b_n, \dots, b_n, \\ Z_{ij} &= [H_1(t_i) - H_2(t_i) - G(t_i)] - [H_1(t_{ij}) - H_2(t_{ij}) - G(t_{ij})], \quad i = 0, \dots, d_n, j = -b_n, \dots, b_n, \\ H_n(x, y) &= \tilde{G}(x) - \tilde{G}(y) - G(x) + G(y), \quad K_n = \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} |H_n(t_i, t_{ij})|, \end{aligned} \tag{6.2}$$

where H_1 and H_2 are defined in (2.7). Since G has continuous density g on $[G^{-1}(a) - \varepsilon, G^{-1}(b) + \varepsilon]$, g is bounded on $[G^{-1}(a) - \varepsilon, G^{-1}(b) + \varepsilon]$, that is $g(x) \leq M_f$ for all x in $[G^{-1}(a) - \varepsilon, G^{-1}(b) + \varepsilon]$, with M_f being some constant. Furthermore, applying the mean value theorem gives $\max_{0 \leq i \leq d_n} a_i \leq M_f/m_n$ and $\max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} a_{ij} \leq M_f/m_n$ for large n . These facts, along with the monotonicity of \tilde{G} and G , imply that, for large n ,

$$\sup_{t_i \leq x \leq t_{i+1}} \sup_{t_{ij} \leq y \leq t_{i(j+1)}} |H_n(x, y)| \leq K_n + \frac{2M_f}{m_n}. \tag{6.3}$$

Inequality (6.3) further implies that, for large n ,

$$\begin{aligned} \sup_{G^{-1}(a) \leq x \leq G^{-1}(b)} \sup_{|y-x| \leq \delta_n} |\tilde{G}(x) - \tilde{G}(y) - G(x) + G(y)| &\leq \max_{0 \leq i \leq d_n} \sup_{x \in I_i} \max_{-b_n \leq j \leq b_n-1} \sup_{t_{ij} \leq y \leq t_{i(j+1)}} |H_n(x, y)| \\ &\leq K_n + \frac{2M_f}{m_n} = K_n + o(n^{-1/2}). \end{aligned} \tag{6.4}$$

Thus, in order to prove part (d), it is enough to show that $K_n = o_p(n^{-1/2})$. For this purpose, by (6.2), (2.6), and (2.8), we have

$$\begin{aligned} K_n &= \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} |[H_1(t_i) - H_2(t_i) - G(t_i)] - [H_1(t_{ij}) - H_2(t_{ij}) - G(t_{ij})] + R_n(t_i) - R_n(t_{ij})| \\ &\leq \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} |Z_{ij}| + 2 \sup_{-\infty < x < \infty} R_n(x) = \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} |Z_{ij}| + o_p(n^{-1/2}). \end{aligned} \tag{6.5}$$

It now suffices to show that $\max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} |Z_{ij}| = o_p(n^{-1/2})$. For this purpose, it can be shown after extensive algebra that

$$\begin{aligned} \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} \text{var} \left(\frac{I_{[X_1 \leq t_i]} - I_{[X_1 \leq t_{ij}]}}{1 + \rho \exp[\alpha_0 + r(X_1)\beta_0]} \right) &\leq M_f \left(\delta_n + \frac{1}{m_n} \right) = O(n^{-1/2}), \\ \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} \text{var} \left(\frac{\rho I_{[Z_1 \leq t_i]} - I_{[Z_1 \leq t_{ij}]}}{1 + \rho \exp[\alpha_0 + r(Z_1)\beta_0]} \right) &\leq \rho M_f \left(\delta_n + \frac{1}{m_n} \right) = O(n^{-1/2}), \\ \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} |\rho A_{11}(t_i) - \rho A_{11}(t_{ij})| &\leq M_f \left(\delta_n + \frac{1}{m_n} \right) = O(n^{-1/2}), \\ \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} |\rho A_{12}^k(t_i) - \rho A_{12}^k(t_{ij})| &\leq M_r M_f^{1/2} \left(\delta_n + \frac{1}{m_n} \right)^{1/2} = O(n^{-1/4}), \end{aligned} \tag{6.6}$$

where $M_r = \max_{1 \leq k \leq \rho} E_G[r_k(X_1)^2]^{1/2}$ and $A_{12}^k(\cdot)$ for $k = 1, \dots, p$ are defined in (6.10). According to (6.2) and (2.7), we can write

$$\begin{aligned}
 & \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} |Z_{ij}| \\
 &= \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} \left| \frac{1}{n_0} \sum_{k=1}^n \frac{I_{[T_k \leq t_i]} - I_{[T_k \leq t_{ij}]}}{1 + \rho \exp[\alpha_0 + r(T_k)\beta_0]} - [G(t_i) - G(t_{ij})] \right. \\
 & \quad \left. - \frac{\rho}{n} [A_{11}(t_i) - A_{11}(t_{ij}), A_{12}^T(t_i) - A_{12}^T(t_{ij})] S^{-1} \begin{pmatrix} \frac{\partial l(\alpha_0, \beta_0)}{\partial \alpha} \\ \frac{\partial l(\alpha_0, \beta_0)}{\partial \beta} \end{pmatrix} \right| \\
 &= \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} \left| \frac{1}{n_0} \sum_{k=1}^{n_0} \frac{I_{[X_k \leq t_i]} - I_{[X_k \leq t_{ij}]}}{1 + \rho \exp[\alpha_0 + r(X_k)\beta_0]} - [G(t_i) - G(t_{ij})] \right. \\
 & \quad \left. + \frac{1}{n_1} \sum_{k=1}^{n_1} \frac{\rho [I_{[Z_k \leq t_i]} - I_{[Z_k \leq t_{ij}]}]}{1 + \rho \exp[\alpha_0 + r(Z_k)\beta_0]} - [G_2(t_i) - G_2(t_{ij})] \right. \\
 & \quad \left. - \frac{\rho}{n} [A_{11}(t_i) - A_{11}(t_{ij}), A_{12}^T(t_i) - A_{12}^T(t_{ij})] S^{-1} \begin{pmatrix} \frac{\partial l(\alpha_0, \beta_0)}{\partial \alpha} \\ \frac{\partial l(\alpha_0, \beta_0)}{\partial \beta} \end{pmatrix} \right| \\
 &\leq \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} \left| \frac{1}{n_0} \sum_{k=1}^{n_0} \frac{I_{[X_k \leq t_i]} - I_{[X_k \leq t_{ij}]}}{1 + \rho \exp[\alpha_0 + r(X_k)\beta_0]} - [G(t_i) - G(t_{ij})] \right| \\
 & \quad + \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} \left| \frac{1}{n_1} \sum_{k=1}^{n_1} \frac{\rho [I_{[Z_k \leq t_i]} - I_{[Z_k \leq t_{ij}]}]}{1 + \rho \exp[\alpha_0 + r(Z_k)\beta_0]} - [G_2(t_i) - G_2(t_{ij})] \right| \\
 & \quad + \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} \left| \frac{\rho}{n} [A_{11}(t_i) - A_{11}(t_{ij}), A_{12}^T(t_i) - A_{12}^T(t_{ij})] S^{-1} \begin{pmatrix} \frac{\partial l(\alpha_0, \beta_0)}{\partial \alpha} \\ \frac{\partial l(\alpha_0, \beta_0)}{\partial \beta} \end{pmatrix} \right|, \tag{6.7}
 \end{aligned}$$

where $G_1(t) = \int_{-\infty}^t 1/\{1 + \rho \exp[\alpha_0 + r(x)\beta_0]\} g(x) dx$ and $G_2(t) = G(t) - G_1(t)$. To deal with the first term in (6.7), for any $\lambda > 0$, using (6.6) and Bernstein's inequality (Serfling 1980, p. 95) gives

$$\begin{aligned}
 &P\left(\sqrt{n} \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} \left| \frac{1}{n_0} \sum_{k=1}^{n_0} \left[\frac{I_{[X_k \leq t_i]} - I_{[X_k \leq t_{ij}]}}{1 + \rho \exp[\alpha_0 + r(X_k)\beta_0]} - [G_1(t_i) - G_1(t_{ij})] \right] \right| > \lambda \right) \\
 &\leq \sum_{i=0}^{d_n} \sum_{j=-b_n}^{b_n} P\left(\left| \sum_{k=1}^{n_0} \left[\frac{I_{[X_k \leq t_i]} - I_{[X_k \leq t_{ij}]}}{1 + \rho \exp[\alpha_0 + r(X_k)\beta_0]} - [G_1(t_i) - G_1(t_{ij})] \right] \right| > \delta_n n_0 \lambda \right) \\
 &\leq 2(d_n + 1)(2b_n + 1) \exp\left(-\frac{n_0 \delta_n^2 \lambda^2}{2M_1(\delta_n + m_n^{-1}) + \frac{4}{3}\delta_n \lambda}\right) = o(1), \quad \text{as } n \rightarrow \infty.
 \end{aligned}$$

Since λ is arbitrary, we can conclude that

$$\max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} \left| \frac{1}{n_0} \sum_{k=1}^{n_0} \frac{I_{[X_k \leq t_i]} - I_{[X_k \leq t_{ij}]}}{1 + \rho \exp[\alpha_0 + r(X_k)\beta_0]} - [G_1(t_i) - G_1(t_{ij})] \right| = o_p(n^{-1/2}). \quad (6.8)$$

Similarly, we can show that

$$\max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} \left| \frac{1}{n_1} \sum_{k=1}^{n_1} \frac{\rho[I_{[Z_k \leq t_i]} - I_{[Z_k \leq t_{ij}]}]}{1 + \rho \exp[\alpha_0 + r(Z_k)\beta_0]} - [G_2(t_i) - G_2(t_{ij})] \right| = o_p(n^{-1/2}). \quad (6.9)$$

To deal with the third term in (6.7), note from part (a) of Theorem 2.1 that

$$\frac{1}{n} S^{-1} \begin{pmatrix} \frac{\partial l(\alpha_0, \beta_0)}{\partial \alpha} \\ \frac{\partial l(\alpha_0, \beta_0)}{\partial \beta} \end{pmatrix} = O_p(n^{-1/2}).$$

Moreover, define

$$\begin{aligned}
 A_{12}^k(t) &= \int_{-\infty}^t \frac{\exp[\alpha_0 + r(x)\beta_0]}{1 + \rho \exp[\alpha_0 + r(x)\beta_0]} r_k(x) dG(x), \quad k = 1, \dots, p, \\
 B_{0ij} &= \rho[A_{11}(t_i) - A_{11}(t_{ij})], \quad B_{kij} = \rho[A_{12}^k(t_i) - A_{12}^k(t_{ij})], \quad k = 1, \dots, p, \\
 B_{ij} &= (B_{0ij}, B_{1ij}, \dots, B_{pij}),
 \end{aligned} \tag{6.10}$$

$$L = (L_0, L_1, \dots, L_p)^T = \frac{1}{n} S^{-1} \begin{pmatrix} \frac{\partial l(\alpha_0, \beta_0)}{\partial \alpha} \\ \frac{\partial l(\alpha_0, \beta_0)}{\partial \beta} \end{pmatrix}.$$

According to (6.6) and part (a) of Theorem 2.1, we have

$$\max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} |B_{kij}| = O(n^{-1/4}), \quad L_k = O_p(n^{-1/2}), \quad k = 0, 1, \dots, p.$$

Thus

$$\begin{aligned} & \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} \left| \frac{\rho}{n} [A_{11}(t_i) - A_{11}(t_{ij}), A_{12}^T(t_i) - A_{12}^T(t_{ij})] S^{-1} \begin{pmatrix} \frac{\partial I(\alpha_0, \beta_0)}{\partial \alpha} \\ \frac{\partial I(\alpha_0, \beta_0)}{\partial \beta} \end{pmatrix} \right| \\ &= \max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} \left| \sum_{k=1}^p B_{kij} L_k \right| \leq \sum_{k=0}^p |L_k| \left[\max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} |B_{kij}| \right] = O_p(n^{-3/4}). \end{aligned} \tag{6.11}$$

Combining (6.7), (6.8), (6.9) and (6.11) yields

$$\max_{0 \leq i \leq d_n} \max_{-b_n \leq j \leq b_n} |Z_{ij}| = o_p(n^{-1/2}) + o_p(n^{-1/2}) + O_p(n^{-3/4}) = o_p(n^{-1/2}). \tag{6.12}$$

Combine (6.5) and (6.12) gives $K_n = o_p(n^{-1/2})$. This completes the proof of Lemma 6.1 according to (6.4). \square

Proof of Theorem 4.1. For convenience, we use $G^{-1}\tilde{G}G^{-1}(\cdot)$ etc. to denote composite functions. Let

$$\begin{aligned} r_{n1}(s) &= -[G^{-1}\tilde{G}G^{-1}(s) - G^{-1}GG^{-1}(s) + Q_n(s)], \\ r_{n2}(s) &= G^{-1}G\tilde{G}^{-1}(s) - G^{-1}\tilde{G}\tilde{G}^{-1}(s) + G^{-1}\tilde{G}G^{-1}(s) - G^{-1}GG^{-1}(s), \\ r_{n3}(s) &= G^{-1}\tilde{G}\tilde{G}^{-1}(s) - G^{-1}(s); \end{aligned}$$

then it is easy to see that

$$\tilde{G}^{-1}(s) - G^{-1}(s) = Q_n(s) + r_{n1}(s) + r_{n2}(s) + r_{n3}(s). \tag{6.13}$$

Let $\tilde{x}(s) = \tilde{G}G^{-1}(s)$; then $\sup_{a \leq s \leq b} |\tilde{x}(s) - s| = O_p(n^{-1/2})$ by part (a) of Lemma 6.1. Since G has continuous positive density g on $[G^{-1}(a), G^{-1}(b)]$, $g(G^{-1}(s))$ is bounded away from zero on $[a, b]$. As a result, applying Taylor expansion and part (b) of Theorem 2.1 gives

$$\begin{aligned} \sup_{a \leq s \leq b} |r_{n1}(s)| &= \sup_{a \leq s \leq b} \left| G^{-1}(\tilde{x}(s)) - G^{-1}(s) + Q_n(s) \right| \\ &\leq \sup_{a \leq s \leq b} \left| \frac{1}{g(G^{-1}(s))} [\tilde{G}(G^{-1}(s)) - G(G^{-1}(s))] + Q_n(s) \right| + o_p(n^{-1/2}) \\ &= \sup_{a \leq s \leq b} \left| \frac{R_n(G^{-1}(s))}{g(G^{-1}(s))} \right| + o_p(n^{-1/2}) = o_p(n^{-1/2}). \end{aligned} \tag{6.14}$$

Let $\phi_n = \sup_{a \leq s \leq b} |\tilde{G}^{-1}(s) - G^{-1}(s)|$; then $\phi_n = O_p(n^{-1/2})$ by part (b) of Lemma 6.1. Applying parts (a) and (d) of Lemma 6.1 gives

$$\begin{aligned} \sup_{a \leq s \leq b} |r_{n2}(s)| &\leq \sup_{a \leq s \leq b} \left| \frac{[\tilde{G}G^{-1}(s) - GG^{-1}(s)] - [\tilde{G}\tilde{G}^{-1}(s) - G\tilde{G}^{-1}(s)]}{g(G^{-1}(s))} \right| + o_p(n^{-1/2}) \\ &\leq \left[\sup_{a \leq s \leq b} \frac{1}{g(G^{-1}(s))} \right] \sup_{G^{-1}(a) \leq x \leq G^{-1}(b)} \sup_{|y-x| \leq \phi_n} |[\tilde{G}(x) - G(x)] - [\tilde{G}(y) - G(y)]| \\ &\quad + o_p(n^{-1/2}) = o_p(n^{-1/2}). \end{aligned} \tag{6.15}$$

Furthermore, by part (c) of Lemma 6.1, we have

$$\sup_{a \leq s \leq b} |r_{n3}(s)| \leq \sup_{a \leq s \leq b} \left| \frac{\tilde{G}(\tilde{G}^{-1}(s)) - s}{g(G^{-1}(s))} \right| + o_p(n^{-1}) = O_p(n^{-1}). \tag{6.16}$$

Let $r_n(s) = r_{n1}(s) + r_{n2}(s) + r_{n3}(s)$; then both (4.1) and (4.3) hold by combining (6.13)–(6.16). To prove (4.4), according to (4.1) and (4.3), it is enough to show that

$$\sqrt{n}Q_n \xrightarrow{\mathcal{D}} V(G^{-1})/g(G^{-1}) \quad \text{on } D[a, b]. \tag{6.17}$$

Let

$$V_n(t) = H_1(t) - H_2(t) - G(t); \tag{6.18}$$

then

$$\sqrt{n}V_n(t) = \sqrt{n}(H_1(t) - H_2(t) - \tilde{G}(t)) + \sqrt{(1 + \rho)}\sqrt{n_0}(\tilde{G}(t) - G(t)).$$

It is easy to see that $EV_n(t) = EH_1(t) - G(t) = 0$. For $s \leq t$, it can be shown after very extensive algebra that $E[\{\sqrt{n}V_n(s)\}\{\sqrt{n}V_n(t)\}] = EV(s)V(t)$, which, along with the central limit theorem for sample means and the Cramér–Wold device, implies that the finite-dimensional distributions of $\sqrt{n}V_n$ converge weakly to those of V . By employing the tightness criteria in Billingsley (1968), we can show that the process $\{\sqrt{n}V_n(t), -\infty \leq t \leq \infty\}$ is tight in $D[-\infty, \infty]$. As a result, we have

$$\sqrt{n}V_n \xrightarrow{\mathcal{D}} V \quad \text{in } D[-\infty, \infty]. \tag{6.19}$$

Therefore, applying (4.2), (6.18), and (6.19) yields

$$\sqrt{n}Q_n = -\frac{\sqrt{n}V_n(G^{-1})}{g(G^{-1})} \xrightarrow{\mathcal{D}} -\frac{V(G^{-1})}{g(G^{-1})} \stackrel{\mathcal{D}}{=} \frac{V(G^{-1})}{g(G^{-1})} \quad \text{on } D[a, b],$$

thus establishing (6.17). The proof is completed. □

Acknowledgement

The author wishes to thank an associate editor and a referee for several helpful comments on an earlier version of the paper, which improved the results and presentation.

References

- Anderson, J.A. (1972) Separate sample logistic discrimination. *Biometrika*, **59**, 19–35.
- Anderson, J.A. (1979) Multivariate logistic compounds, *Biometrika*, **66**, 17–26.
- Bickel, P.J. and Doksum, K.A. (1977) *Mathematical Statistics: Basic Ideas and Selected Topics*. Englewood Cliffs, NJ: Prentice Hall.
- Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.
- Breslow, N. and Day, N.E. (1980) *Statistical Methods in Cancer Research, Vol. 1: The Analysis of Case–Control Studies*. Lyon: IARC.
- Drescher, K. and Becher, H. (1997) Estimating the generalized impact fraction from case–control data. *Biometrics*, **53**, 1170–1176.
- Efron, B. and Tibshirani, R. (1996) Using specially designed exponential families for density estimation. *Ann. Statist.*, **24**, 2431–2461.
- Farewell, V. (1979) Some results on the estimation of logistic models based on retrospective data. *Biometrika*, **66**, 27–32.
- Fokianos, K., Kedem, B., Qin, J., Haferman, J.L. and Short, D.A. (1998) On combining instruments. *J. Appl. Meteorology*, **37**, 220–226.
- Gilbert, P.B., Self, S.R. and Ashby, M. (1998) Statistical methods for assessing differential vaccine protection against human immunodeficiency virus types. *Biometrics*, **54**, 799–814.
- Gilbert, P.B., Lele, S.R. and Vardi, Y. (1999) Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials, *Biometrika*, **86**, 27–43.
- Gill, R.D., Vardi, Y. and Wellner, J.A. (1988) Large sample theory of empirical distributions in biased sampling models. *Ann. Statist.*, **16**, 1069–1112.
- Hall, P. and La Scala, B. (1990) Methodology and algorithms of empirical likelihood. *Internat. Statist. Rev.*, **58**, 109–127.
- Hosmer, D.J. and Lemeshow, S. (1989) *Applied Logistic Regression*. New York: Wiley.
- Owen, A.B. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.
- Owen, A.B. (1990) Empirical likelihood confidence regions. *Ann. Statist.*, **18**, 90–120.
- Owen, A.B. (1991) Empirical likelihood for linear models. *Ann. Statist.*, **19**, 1725–1747.
- Prentice, R.L. and Pyke, R. (1979) Logistic disease incidence models and case–control studies. *Biometrika*, **66**, 403–411.
- Qin, J. (1993) Empirical likelihood in biased sample problems. *Ann. Statist.*, **21**, 1182–1196.
- Qin, J. and Lawless, J. (1994) Empirical likelihood and general estimating equations. *Ann. Statist.*, **22**, 300–325.
- Qin, J. and Zhang, B. (1997) A goodness of fit test for logistic regression models based on case–control data. *Biometrika*, **84**, 609–618.
- Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*. New York: Wiley.
- Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.
- Vardi, Y. (1982) Nonparametric estimation in presence of length bias. *Ann. Statist.*, **10**, 616–620.
- Vardi, Y. (1985) Empirical distribution in selection bias models. *Ann. Statist.*, **13**, 178–203.

Received January 1998 and revised February 1999