Prospective and retrospective analyses under logistic regression models

Biao Zhang*

Department of Mathematics, The University of Toledo, 2801 W. Bancroft Street, Toledo, OH 43606, USA

Received 11 July 2004
Available online 23 March 2005

Abstract


© 2005 Elsevier Inc. All rights reserved.

AMS 1991 subject classification: primary 62G05; 62G20

Keywords: Case–control data; Chi-squared; Correlation coefficient; Discriminant function statistic; Fisher information matrix; Fisher’s z-transformation; Maximum semiparametric likelihood; Moment generating function; Moore–Penrose generalized inverse; Odds ratio; Score equation; Score statistic; Unbiased estimating equation; Wald statistic
1. Introduction

Logistic regression models have become the most widely used statistical tool for modeling binary response variables and for analyzing case–control data [4]. Let $Y$ be a binary response variable and let $X$ be the associated $p \times 1$ vector of explanatory variables. Then the standard logistic regression model assumes that

$$ P(Y = 1|X = x) = \frac{\exp(x^* + \beta^T x)}{1 + \exp(x^* + \beta^T x)} \equiv \phi(x; x^*, \beta), \tag{1.1} $$

where $x^*$ is a scale parameter and $\beta$ is a $p \times 1$ vector of odds-ratio parameters. Under case–control sampling as described by Prentice and Pyke [10], data are collected retrospectively in the sense that the value of $X$ is observed for samples of subjects having $Y = 1$, ‘cases’, and having $Y = 0$, ‘controls’. Specifically, let $X_1, \ldots, X_{n_0}$ be a random sample from $P(x|Y = 0)$ and, independently of the $X_i$, let $Z_1, \ldots, Z_{n_1}$ be a random sample from $P(x|Y = 1)$. Let $g(x) = f(x|Y = 0)$ and $h(x) = f(x|Y = 1)$ denote, respectively, the conditional density or frequency functions of $X$ given $Y = 0$ and 1. Qin and Zhang [11] showed that model (1.1) is equivalent to the following two-sample semiparametric model:

$$ X_1, \ldots, X_{n_0} \text{ are independent with density } g(x), $$

$$ Z_1, \ldots, Z_{n_1} \text{ are independent with density } h(x) = \exp(x^* + \beta^T x)g(x), \tag{1.2} $$

where $x = x^* + \log((1 − \pi)/\pi)$ with $\pi = P(Y = 1) = 1 − P(Y = 0)$.

Anderson [2,3] showed for discrete covariates that the odds ratio estimators and their asymptotic covariance matrices may be obtained by applying the prospective logistic regression model (1.1) to the case–control study as if the data had been obtained in a prospective study. Prentice and Pyke [10] generalized these results to allow for continuous and mixed discrete and continuous covariates. Scott and Wild [12] showed that likelihood ratio tests obtained from fitting the prospective logistic regression model (1.1) is also valid in case–control studies. For robust estimation in logistic case–control studies, Wang and Carroll [13,14] demonstrated that the prospective robust analysis leads not only to valid robust estimates for the odds-ratio parameters but also to their asymptotically correct standard errors. Weinberg and Wacholder [15] extended Prentice and Pyke’s [10] results to the broad class of multiplicative-intercept risk models described by Hsieh et al. [9]. Zhang [17] showed that information matrix tests and their asymptotic distributions may be obtained by fitting the prospective logistic regression model (1.1) to case–control data. On the other hand, the prospective analysis under model (1.1) and the retrospective analysis under model (1.2) do not always produce the same result; for example, $x^*$ in model (1.1) is not estimable under model (1.2). This paper has a twofold purpose. The first purpose of this paper is to develop an extension of the aforementioned results to a class of statistics whose prospective analysis under the logistic regression model (1.1) and retrospective analysis under the two-sample semiparametric model (1.2) yield identical conclusions on point estimators and their standard errors. With this extension, Prentice and Pyke’s [10] results are generalized to this wide class of statistics whose prospective analysis under the logistic regression model (1.1) is valid under case–control sampling or under model (1.2). The second purpose of this paper is to extend the results of Prentice and Pyke [10] and Wang and Carroll [13,14] to a class of unbiased estimating equations whose prospective and retrospective analyses give
rise to the same conclusions on odds-ratio parameter estimators and their standard errors. This extension generalizes the score equation of Prentice and Pyke [10] and the robust score equation of Wang and Carroll [13,14] to a wide class of unbiased estimating equations under the case–control sampling scheme in model (1.2).

This paper is organized as follows. In Section 2, we formulate our problem and propose a class of random vectors under models (1.1) and (1.2) along with an important assumption about the class of random vectors. In Section 3, we explore the prospective and retrospective analyses of a class of statistics, whereas in Section 4, we study a class of unbiased estimating equations under models (1.1) and (1.2). Also in Sections 3 and 4, we present five examples related to logistic case–control studies. In Section 5, we consider prospective and retrospective inferences for the odds-ratio parameter \( \beta \) and report results of simulation studies and two real data problems. Finally, proofs of the main theoretical results are provided in the appendix.

2. Approach

Let \((T_1, Y_1), \ldots, (T_n, Y_n)\) be independent bivariate random vectors from a population and let \((t_1, y_1), \ldots, (t_n, y_n)\) be their corresponding observed values. The joint distribution of \((T_i, Y_i)\) is unspecified. We assume, however, that the conditional distribution of \(Y_i\) given \(T_i\) is specified by the prospective logistic regression model (1.1). As discussed in [11], this assumption implies that the conditional distribution of \(Y_i\) given \(T_i\) is specified by the prospective two-sample semiparametric model (1.2). Both the marginal distribution of \(T_i\) and that of \(Y_i\) are unspecified. Under prospective sampling stipulated by model (1.1), the value of \(Y_i\) is observed for given \(T_i = t_i\) so that \(Y_i|T_i = t_i\) has a binomial \(B \{ 1, \phi(t_i; x^*, \beta) \}\) distribution for \(i = 1, \ldots, n\) and \(Y_1, \ldots, Y_n\) are independent for given \((t_1, \ldots, t_n)\). In prospective analysis, statistical inferences are based on the conditional distribution of \((Y_1, \ldots, Y_n)\) given \((T_1, \ldots, T_n)\). On the other hand, under case–control sampling stipulated by model (1.2), the value of \(T_i\) is observed for given \(Y_i = y_i\). For notational convenience, let \(\{T_1, \ldots, T_n\}\) denote, under model (1.2), the combined sample \(\{X_1, \ldots, X_{n_0}; Z_1, \ldots, Z_{n_1}\}\) with \(n = n_0 + n_1\) so that \(t_i = x_i\) and \(y_i = 0\) for \(i = 1, \ldots, n_0\) and \(t_i = z_{i-n_0}\) and \(y_i = 1\) for \(i = n_0 + 1, \ldots, n\). In retrospective analysis, statistical inferences are based on the conditional distribution of \((T_1, \ldots, T_n)\) given \((Y_1, \ldots, Y_n)\).

For \(k = 1, \ldots, q\), let \(d_k(t, y, \theta, \beta)\) be a real function of data \((t, y)\) and parameter vector \((\theta, \beta)\) and let \(d(t, y, \theta, \beta) = (d_1(t, y, \theta, \beta), \ldots, d_q(t, y, \theta, \beta))^T\). On the basis of \(n\) bivariate random vectors \((T_1, Y_1), \ldots, (T_n, Y_n)\), we define \(D_n(\theta, \beta)\) to be the \(q \times 1\) random vector given by

\[
D_n(\theta, \beta) = \frac{1}{n} \sum_{i=1}^{n} d(T_i, Y_i, \theta, \beta). \quad (2.1)
\]

In (2.1), the parameter \(\theta\) is related to \(x^*\) and \(z\) in two different ways according as we perform a prospective analysis or a retrospective analysis on \(D_n(\theta, \beta)\). Under the prospective logistic regression model (1.1), we define \(\theta = x^*\), whereas under the retrospective two-sample semiparametric model (1.2), we define \(\theta = z + \log \rho\) with \(\rho = n_1/n_0\). The latter choice
for $\theta$ reflects the fact that $x^*$ is not estimable under model (1.2) based on case–control data $\{X_1, \ldots, X_{n_0}; Z_1, \ldots, Z_{n_1}\}$. Indeed, under model (1.2), we can only estimate $x = x^* + \log\{(1 - \pi)/\pi\}$ instead of $x^*$.

Throughout this paper, let $(\theta_0, \beta_0) = (x_0^*, \beta_0^*)$ be the true value of $(\theta, \beta^*)$ under model (1.1) and $(\theta_0, \beta_0^*) = (z_0 + \log \rho, \beta_0^*)$ be the true value of $(\theta, \beta^*)$ under model (1.2). Furthermore, if $v$ is a $p \times 1$ vector, we define $v' = (1, v^T)$. Moreover, we assume that $\rho = n_1/n_0$ remains fixed as $n \to \infty$ and that $d_k(t, 1, \theta, \beta)$ and $d_k(t, 0, \theta, \beta)$ are related by

$$d_k(t, 0, \theta, \beta) = -\exp(\theta + \beta^T t) d_k(t, 1, \theta, \beta), \quad k = 1, \ldots, q. \tag{2.2}$$

Assumption (2.2) is naturally satisfied in several important applications of logistic regression models as shown in Sections 3.3 and 4.3. On the basis of $D_n(\theta, \beta)$ satisfying assumption (2.2), we study in Sections 3 and 4 the prospective and retrospective analyses of a class of statistics and a class of unbiased estimating equations, respectively.

3. A class of statistics based on $D_n(\theta, \beta)$

Let $(\hat{x}^*, \hat{\beta})$ be the (prospective) maximum likelihood estimator of $(x_0^*, \beta_0)$ under model (1.1) and let $(\tilde{x}, \tilde{\beta})$ be the (retrospective) maximum semiparametric likelihood estimator of $(x_0, \beta_0)$ under model (1.2). According to Prentice and Pyke [10] and Qin and Zhang [11], $(\hat{x}^*, \hat{\beta})$ and $(\tilde{x}, \tilde{\beta})$ are related by $\hat{x}^* = \tilde{x} + \log \rho$ and $\hat{\beta} = \tilde{\beta}$. As a result, the (prospective) maximum likelihood estimator $\hat{\theta}$ of $\theta_0$ under model (1.1) and the (retrospective) maximum semiparametric likelihood estimator $\tilde{\theta}$ of $\theta_0$ under model (1.2) are identical in that $\hat{\theta} = \tilde{x}^* = \tilde{x} + \log \rho = \tilde{\theta}$. Consequently, $D_n(\hat{\theta}, \hat{\beta}) = D_n(\tilde{\theta}, \tilde{\beta})$. This fact indicates that the prospective analysis under model (1.1) and the retrospective analysis under model (1.2) give rise to the same point estimator for $D_n(\theta, \beta)$. Moreover, the prospective and retrospective analyses in the next two subsections demonstrate that under assumption (2.2), $D_n(\hat{\theta}, \hat{\beta})$ and $D_n(\tilde{\theta}, \tilde{\beta})$ are, respectively, asymptotically normal with mean zero under models (1.1) and (1.2) and have the same estimated asymptotic covariance matrix.

3.1. Prospective analysis of $D_n(\hat{\theta}, \hat{\beta})$

Under the prospective logistic regression model (1.1), we have $\theta = x^*$ so that $D_n(\theta, \beta) = n^{-1} \sum_{i=1}^{n} d(T_i, Y_i, x^*, \beta)$. In prospective analysis, assumption (2.2) becomes $d_k(t, 0, x^*, \beta) = -\exp(x^* + \beta^T t) d_k(t, 1, x^*, \beta)$ for $k = 1, \ldots, q$. For $j, k = 1, \ldots, q$, write

$$U_n(x^*, \beta) = \sum_{i=1}^{n} \left\{ y_i - \frac{\exp(x^* + \beta^T t_i)}{1 + \exp(x^* + \beta^T t_i)} \right\} t_i' ,$$

$$b_{nk}(x^*, \beta) = \frac{1}{n} \sum_{i=1}^{n} \frac{d_k(t_i, 0, x^*, \beta)}{1 + \exp(x^* + \beta^T t_i)} t_i' ,$$

$$b_k(x^*, \beta) = \lim_{n \to \infty} b_{nk}(x^*, \beta), \quad B(x^*, \beta) = (b_1(x^*, \beta), \ldots, b_q(x^*, \beta))^T, $$
Theorem 1. Suppose that model (1.1) and assumption (2.2) hold and that \( I(\mathbf{x}, \beta) \) is positive definite. Suppose further that \( b_k(\mathbf{x}, \beta) \) and \( c_{ij}(\mathbf{x}, \beta) \) are all finite. Let \((\mathbf{z}^*, \hat{\beta})\) be, under model (1.1), the (prospective) maximum likelihood estimator of \((\mathbf{z}_0^*, \beta_0)\) given by the solution to the system of score equations: \( U_n(\mathbf{x}, \beta) = 0 \). Let \( \hat{D}_n = D_n(\hat{\theta}, \hat{\beta}) = D_n(\mathbf{z}^*, \hat{\beta}). \) Then a standard prospective analysis under model (1.1) yields the following theorem.

**Theorem 1.** Suppose that model (1.1) and assumption (2.2) hold and that \( I(\mathbf{x}, \beta) \) is positive definite. Suppose further that \( d(t, \mathbf{y}, \theta, \beta) \) is continuously differentiable in \((\theta, \beta)\) for every \((t, \mathbf{y})\). Then conditional on \((T_1 = t_1, \ldots, T_n = t_n)\), one can write \( \hat{D}_n = D_n(\mathbf{z}_0^*, \beta_0) + n^{-1}B(\mathbf{z}_0^*, \beta_0)I^{-1}(\mathbf{z}_0^*, \beta_0)U_n(\mathbf{z}_0^*, \beta_0) + o_p(n^{-1/2}). \) As a result, \( \sqrt{n}\hat{D}_n \xrightarrow{d} N_q(0, \Gamma) \) as \( n \to \infty \), where \( \Gamma = (\gamma_{ij}), i, j = 1, \ldots, q \) with \( \gamma_{ij} = c_{ij}(\mathbf{z}_0^*, \beta_0) - b_i^*(\mathbf{z}_0^*, \beta_0)I^{-1}(\mathbf{z}_0^*, \beta_0)b_j(\mathbf{z}_0^*, \beta_0). \)

Let \( \hat{c}_{ij} = c_{ni}(\mathbf{z}^*, \hat{\beta}), \hat{b}_i = b_{ni}(\mathbf{z}^*, \hat{\beta}), \) and \( \hat{I} = I_n(\mathbf{z}^*, \hat{\beta}). \) Then the asymptotic covariance matrix \( \Gamma \) can be estimated by \( \hat{\Gamma} = (\hat{\gamma}_{ij}), i, j = 1, \ldots, q \) with \( \hat{\gamma}_{ij} = \hat{c}_{ij} - \hat{b}_i^*\hat{\Gamma}^{-1}\hat{b}_j. \) Moreover, we have \( \hat{K}_n = n\hat{D}_n^\dagger\hat{\Gamma}^\dagger\hat{D}_n \xrightarrow{d} \chi_q^2 \) under model (1.1) as \( n \to \infty \), where \( \hat{\Gamma}^\dagger \) is the Moore–Penrose generalized inverse of \( \hat{\Gamma} \) and \( r \) is the rank of \( \hat{\Gamma} \). In particular, if \( \Gamma \) is nonsingular, we have \( \hat{K}_n = n\hat{D}_n^\dagger\hat{\Gamma}^\dagger\hat{D}_n \xrightarrow{d} \chi_q^2 \) under model (1.1) as \( n \to \infty \).

### 3.2. Retrospective analysis of \( D_n(\hat{\theta}, \hat{\beta}) \)

Under the retrospective two-sample semiparametric model (1.2), we have \( \hat{\theta} = \mathbf{z} + \log \rho \) and \( D_n(\theta, \beta) = n^{-1} \sum_{i=1}^{n} d(T_i, \mathbf{Y}_i, \theta, \beta) = n^{-1} \sum_{i=1}^{n} d(X_i, 0, \theta, \beta) + n^{-1} \sum_{j=1}^{n} d(Z_j, 1, \theta, \beta). \) In retrospective analysis, assumption (2.2) becomes \( d_k(t, 0, \theta, \beta) = -\rho \exp(\mathbf{z} + \beta^\dagger t)dk(t, 1, \theta, \beta) \) for \( k = 1, \ldots, q. \) For \( j, k = 1, \ldots, q, \) write

\[
\begin{align*}
    b_k(\theta, \beta, G) &= \frac{1}{1 + \rho} \int d_k(t, 0, \theta, \beta)t' dG(t), \\
    B(\theta, \beta, G) &= (b_1(\theta, \beta, G), \ldots, b_q(\theta, \beta, G))^\dagger, \\
    c_{jk}(\theta, \beta, G) &= \frac{1}{1 + \rho} \int d_j(t, 1, \theta, \beta)dk(t, 1, \theta, \beta) \exp(\theta + \beta^\dagger t) \\
    &\times \{1 + \exp(\theta + \beta^\dagger t)\} dG(t).
\end{align*}
\]
\begin{equation}
S(\theta, \beta, G) = \frac{1}{1 + \rho} \int \frac{\exp(\theta + \beta^T t)}{1 + \exp(\theta + \beta^T t)} t'(t')^\tau \, dG(t),
\end{equation}

\begin{equation}
Q_n(\theta, \beta) = \sum_{k=1}^{n} \left\{ I(k > n_0) - \frac{\exp(\theta + \beta^T T_k)}{1 + \exp(\theta + \beta^T T_k)} \right\} T_k'.
\end{equation}

Let \((\tilde{\theta}, \tilde{\beta}) = (\tilde{x} + \log \rho, \tilde{\beta})\) be, under model (1.2), the (retrospective) maximum semiparametric likelihood estimator of \((\theta_0, \beta_0)\) given by the solution to the system of score equations: \(Q_n(\theta, \beta) = 0\). Let \(\tilde{D}_n = D_n(\tilde{\theta}, \tilde{\beta})\). The following theorem establishes the asymptotic distribution of \(\tilde{D}_n\) under model (1.2).

**Theorem 2.** Suppose that model (1.2) and assumption (2.2) hold and that \(d(t, 1, \theta, \beta)\) is continuously differentiable in \((\theta, \beta)\) for every \(t\). Suppose further that \(S(\theta_0, \beta_0, G)\) is positive definite, \(\|B(\theta_0, \beta_0, G)\| < \infty\), \(\int \|t\|^2 \, dG(t) < \infty\), and \(\int \|d(t, 1, \theta_0, \beta_0)\|^2 \exp(\theta_0 + \beta_0^T t) [1 + \exp(\theta_0 + \beta_0^T t)] \, dG(t) < \infty\). Then one can write \(\tilde{D}_n = D_n(\theta_0, \beta_0) + n^{-1} B(\theta_0, \beta_0, G) S^{-1}(\theta_0, \beta_0, G) Q_n(\theta_0, \beta_0) + o_p(n^{-1/2})\). As a result, \(\sqrt{n} \tilde{D}_n \overset{d}{\rightarrow} N_q(0, \Sigma)\) as \(n \rightarrow \infty\), where \(\Sigma = (\sigma_{ij})_{i,j=1,...,q}\) with \(\sigma_{ij} = c_{ij}(\theta_0, \beta_0, G) - b_i^\tau(\theta_0, \beta_0, G) S^{-1}(\theta_0, \beta_0, G) b_j (\theta_0, \beta_0, G)\).

The proof of Theorem 2 is given in the appendix. According to Qin and Zhang [11],

\[ \tilde{G}(t) = \frac{1}{n_0} \sum_{i=1}^{n} \frac{I_{[T_i \leq t]}}{1 + \rho \exp(\tilde{\theta} + \tilde{\beta}^T T_i)} \]  

and

\[ \tilde{H}(t) = \frac{1}{n_0} \sum_{i=1}^{n} \frac{\exp(\tilde{\theta} + \tilde{\beta}^T T_i) I_{[T_i \leq t]}}{1 + \rho \exp(\tilde{\theta} + \tilde{\beta}^T T_i)} \]

are the maximum semiparametric likelihood estimators of \(G\) and \(H\) under model (1.2). Let

\[ \tilde{b}_k = b_k(\tilde{\theta}, \tilde{\beta}, \tilde{G}) = \frac{1}{n} \sum_{i=1}^{n} \frac{d_k(t_i, 0, \tilde{\theta}, \tilde{\beta})}{1 + \exp(\tilde{\theta} + \tilde{\beta}^T t_i)} t'_i, \]

\[ \tilde{S} = S(\tilde{\theta}, \tilde{\beta}, \tilde{G}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\exp(\tilde{\theta} + \tilde{\beta}^T t_i)}{1 + \exp(\tilde{\theta} + \tilde{\beta}^T t_i)} t'_i(t'_i)^\tau, \]

\[ \tilde{c}_{jk} = c_{jk}(\tilde{\theta}, \tilde{\beta}, \tilde{G}) = \frac{1}{n} \sum_{i=1}^{n} d_j(t_i, 1, \tilde{\theta}, \tilde{\beta}) d_k(t_i, 1, \tilde{\theta}, \tilde{\beta}) \times \exp(\tilde{\theta} + \tilde{\beta}^T t_i), \quad j, k = 1, \ldots, q. \]

Then \((\tilde{b}_k, \tilde{c}_{jk}, \tilde{S})\) is the corresponding empirical version of \((b_k(\theta_0, \beta_0, G), c_{jk}(\theta_0, \beta_0, G), S(\theta_0, \beta_0, G))\) with \((\theta_0, \beta_0, G)\) replaced by \((\tilde{\theta}, \tilde{\beta}, \tilde{G})\). It can be shown that \((\tilde{b}_k, \tilde{c}_{jk}, \tilde{S})\) is a consistent estimator of \((b_k(\theta_0, \beta_0, G), c_{jk}(\theta_0, \beta_0, G), S(\theta_0, \beta_0, G))\) under model (1.2). Consequently, the asymptotic covariance matrix \(\Sigma\) can be consistently estimated by \(\tilde{\Sigma} = (\tilde{\sigma}_{ij})_{i,j=1,...,q}\) with \(\tilde{\sigma}_{ij} = \tilde{c}_{ij} - \tilde{b}_i^\tau \tilde{S}^{-1} \tilde{b}_j\). If \(\tilde{\Sigma}^+\) is the Moore–Penrose generalized inverse of \(\tilde{\Sigma}\), then \(\tilde{K}_n = n^{\tilde{D}_n^\tau \tilde{\Sigma}^+ \tilde{D}_n} \overset{d}{\rightarrow} \chi_r^2\) under model (1.2) as \(n \rightarrow \infty\), where \(r\) is the rank of \(\tilde{\Sigma}\). In particular, if \(\Sigma\) is nonsingular, we have \(\tilde{K}_n = n^{\tilde{D}_n^\tau \Sigma^{-1} \tilde{D}_n} \overset{d}{\rightarrow} \chi_q^2\) under model (1.2) as \(n \rightarrow \infty\).
As indicated at the beginning of this section, we have \((\tilde{\theta}, \tilde{\beta}) = (\hat{\mathbf{x}}^*, \hat{\beta})\) and \(\tilde{D}_n = \hat{D}_n\). It is seen from (3.1) and (3.3) that \(\tilde{b}_k = \hat{b}_k, \tilde{c}_{ij} = \hat{c}_{ij}\), and \(\hat{S} = \hat{I}\). As a result, \(\hat{\Sigma} = \hat{\Gamma}\) and \(\hat{K}_n = \hat{K}_n\). Consequently, the prospectively estimated asymptotic covariance matrix \(\hat{\Gamma}\) of \(\tilde{D}_n\) under model (1.1) and the retrospectively estimated asymptotic covariance matrix \(\hat{\Sigma}\) of \(\tilde{D}_n\) under model (1.2) are identical. Therefore, there is no need under assumption (2.2) to distinguish between the prospective and retrospective analyses on \(D_n(\theta, \beta)\) in that the prospective point estimator \(\tilde{D}_n = D_n(\hat{\theta}, \hat{\beta})\) and the retrospective point estimator \(\hat{D}_n = D_n(\theta, \beta)\) possess the same observed value, the same asymptotic mean 0, and the same estimated asymptotic covariance matrix. In addition, the prospective and retrospective Wald-type statistics \(\hat{K}_n\) and \(\tilde{K}_n\) have the same observed value and the same asymptotic \(\chi^2_r\) or \(\chi^2_q\) distribution.

In logistic case–control studies, sampling data are generated from the retrospective model (1.2) instead of from the prospective model (1.1). We would, however, like to ignore the case–control structure under model (1.2) and to analyze case–control data under model (1.1) as if they had been obtained in a prospective study, partly because it is more natural to model disease status for given covariates than to model covariates for given disease status and partly because standard logistic regression programs are readily available for fitting the prospective logistic regression model (1.1). Prentice and Pyke [10] showed that it is valid to obtain the odds-ratio parameter estimators and their estimated asymptotic covariance matrices with case–control sampling by fitting the prospective logistic regression model (1.1) to case–control data. The prospective and retrospective analyses on \(D_n(\theta, \beta)\) reveal that when assumption (2.2) holds, \(\tilde{D}_n = D_n(\hat{\theta}, \hat{\beta})\) can be constructed either from the prospective analysis under model (1.1) or from the retrospective analysis under model (1.2). Furthermore, the asymptotic distribution of \(\tilde{D}_n\) is normal with mean \(\theta\) and estimated asymptotic covariance matrix \(\hat{\Sigma}\) whether it is derived from the conditional distribution of \(Y\) given \(X\) under model (1.1) or from the conditional distribution of \(X\) given \(Y\) under model (1.2). Moreover, the asymptotic distribution of \(\tilde{T}_n\) is the \(\chi^2_r\) or \(\chi^2_q\) distribution whether it is deduced from the prospective analysis under model (1.1) or from the retrospective analysis under model (1.2). These results demonstrate that \(\tilde{D}_n\) and its estimated asymptotic covariance matrix may be obtained from case–control data under model (1.2) by applying the prospective logistic regression model (1.1) directly as if the case–control structure in model (1.2) were completely ignored and as if the case–control data had arisen from a prospective study, thus generalizing Prentice and Pyke’s [10] results to a wide class of statistics \(\tilde{D}_n = D_n(\hat{\theta}, \hat{\beta})\) satisfying assumption (2.2).

3.3. Examples

In this subsection, we provide two illustrative examples to demonstrate that Theorems 1 and 2 generalize several results in logistic case–control studies.

**Example 1.** Let \(d(t, y, \theta, \beta) = \eta(t, \theta, \beta)(y + (y - 1) \exp(\theta + \beta^T t))\), where \(\eta(t, \theta, \beta)\) is a \(q \times 1\) vector-valued function. Then assumption (2.2) is satisfied. It is seen that under model (1.2), \(D_n(\tilde{x} + \log \rho, \tilde{\beta})\) is the generalized moments specification test statistic of Fokianos et al. [6]. For \(p = q = 2\), Fokianos et al. [6] considered four choices for \(\eta(t, \theta, \beta)\) given
Theorems 1 and 2 extend the results of Zhang [17] to a wide class of statistics (1.1) and (1.2). Although the unbiased estimating equation 

\[ E[\frac{d_k(t, y, \theta, \beta)}{1 + \exp(\theta + \beta^t t)}] \] 

be a solution to the unbiased estimating equation model (1.1) and a retrospective unbiased estimating equation under model (1.2). Let

\[ d_k(t, y, \theta, \beta) = \frac{\exp(\theta + \beta^t t) - y - (1 - y)\{\exp(\theta + \beta^t t)\}^2}{\{1 + \exp(\theta + \beta^t t)\}^2} w_k(t), \]

where \( w_k(t) \) is a function of \( t \). It is easy to verify that assumption (2.2) is satisfied. For the particular choice of \( w_k(t) \) as given in [17], \( D_n(\tilde{z}^*, \tilde{\beta}) \) and \( D_n(\tilde{z} + \log \rho, \tilde{\beta}) \) are, respectively, the prospective and retrospective information-matrix test statistics \( \tilde{Q}_n \) and \( \tilde{Q}_n \) of Zhang [17]. Theorems 1 and 2 extend the results of Zhang [17] to a wide class of statistics \( D_n(\tilde{\theta}, \tilde{\beta}) \) satisfying assumption (2.2).

4. A class of unbiased estimating equations based on \( D_n(\tilde{\theta}, \tilde{\beta}) \)

Throughout this section, let \( q = p + 1 \). It can be shown under assumption (2.2) that

\[ E[D_n(\tilde{\theta}, \tilde{\beta})|T_1 = t_1, \ldots, T_n = t_n] = 0 \] under model (1.1) and \( E[D_n(\tilde{\theta}, \tilde{\beta})|Y_1 = y_1, \ldots, Y_n = y_n] = 0 \) under model (1.2). These two facts imply that the estimating equation \( D_n(\tilde{\theta}, \tilde{\beta}) \) cannot be a solution to the unbiased estimating equation \( D_n(\tilde{\theta}, \tilde{\beta}) = 0 \). Based on the unbiased estimating function \( D_n(\tilde{\theta}, \tilde{\beta}) \), we propose to estimate \((x^*, \beta)\) under model (1.1) by \((\tilde{x}^*, \tilde{\beta})\) with \( \tilde{x}^* = \tilde{\theta} \) and to estimate \((x, \beta)\) under model (1.2) by \((\tilde{x}, \beta)\) with \( \tilde{x} = \tilde{\theta} - \log \rho \). Clearly, \( \tilde{x}^* \) and \( \tilde{x} \) are related by \( \tilde{x}^* = \tilde{x} + \log \rho \), resembling the connection between the (prospective) maximum likelihood estimator \( \hat{x}^* \) and the (retrospective) maximum semiparametric likelihood estimator \( \tilde{x} \). Moreover, the value of the point estimator \( \tilde{\beta} \) for \( \beta \) is the same under models (1.1) and (1.2). Although the unbiased estimating equation \( D_n(\tilde{\theta}, \tilde{\beta}) = 0 \) is the same in prospective sampling under model (1.1) and retrospective sampling under model (1.2), this does not imply that \( (\tilde{\theta}, \tilde{\beta}) \) has the same sampling distribution under models (1.1) and (1.2). The prospective and retrospective analyses on \( (\tilde{\theta}, \tilde{\beta}) \) in the next two subsections demonstrate under assumption (2.2) that \( \tilde{\beta} \) is asymptotically normal with mean \( \beta_0 \) under both models (1.1) and (1.2) and that the prospectively estimated asymptotic covariance matrix of \( \tilde{\beta} \) under model (1.1) is identical to the retrospectively estimated asymptotic covariance matrix of \( \tilde{\beta} \) under model (1.2).

4.1. Prospective analysis of \((\tilde{\theta}, \tilde{\beta})\)

In prospective analysis under model (1.1), we have \( \theta = x^* \) and \( \tilde{\theta} = \tilde{x}^* \) so that \((\tilde{x}^*, \tilde{\beta})\) is a solution to \( D_n(x^*, \beta) = n^{-1} \sum_{i=1}^{n} d(T_i, Y_i, x^*, \beta) = 0 \). Let \( C(x^*, \beta) \) be the \((p+1) \times (p+1)\) matrix defined by \( C(x^*, \beta) = (c_{ij}(x^*, \beta))_{i,j=1,\ldots,p+1} \) with \( c_{ij}(x^*, \beta) \) given in (3.1). A standard prospective analysis under model (1.1) produces the following result regarding the prospective asymptotic distribution of \((\tilde{x}^*, \tilde{\beta})\).
Theorem 3. Suppose that model (1.1) and assumption (2.2) hold and that \( B(z_0^*,\beta_0) \) is nonsingular. Suppose further that \( d(t, y, z^*, \beta) \) is twice continuously differentiable in \((z^*, \beta)\) for every \((t, y)\) and the second-order partial derivatives of \( d(t, y, z^*, \beta) \) are dominated by a finite function \( \psi(t, y) \) for every \((z^*, \beta)\) in a neighborhood of \((z_0^*, \beta_0)\). Then

(a) As \( n \to \infty \), with probability tending to 1 there exists a sequence of roots \((\tilde{z}^*, \tilde{\beta})\) of the system of equations \( D_n(z^*, \beta) = 0 \) such that \((\tilde{z}^*, \tilde{\beta})\) is consistent for estimating \((z_0^*, \beta_0)\).

(b) Conditional on \((T_1 = t_1, \ldots, T_n = t_n)\), one can write \((\tilde{z}^* - z_0^*, \tilde{\beta} - \beta_0) = -B^{-1}(z_0^*,\beta_0) D_n(z^*,\beta)+o_p\left(n^{-1/2}\right)\) as \( n \to \infty \). As a result, \( \sqrt{n}(\tilde{z}^* - z_0^*, \tilde{\beta} - \beta_0)^\tau \xrightarrow{d} N_{p+1}(0, \Lambda) \) as \( n \to \infty \), where \( \Lambda = B^{-1}(z_0^*,\beta_0) C(z_0^*,\beta_0)\{B^{-1}(z_0^*,\beta_0)\}^\tau \) with \( B(z^*, \beta) \) defined in (3.1).

Let \( \hat{B} = (b_{n1}(\tilde{z}^*, \bar{\beta}), \ldots, b_{np+1}(\tilde{z}^*, \bar{\beta}))^\tau \) and \( \hat{C} = (c_{nij}(\tilde{z}^*, \bar{\beta}))_{i,j=1,\ldots,p+1} \) with \( b_{n1}(\tilde{z}^*, \bar{\beta}) \) and \( c_{nij}(\tilde{z}^*, \bar{\beta}) \) defined in (3.1). Then the asymptotic covariance matrix \( \Lambda \) can be estimated by \( \hat{\Lambda} = \hat{B}^{-1} \hat{C} (\hat{B}^{-1})^\tau \). Suppose we are interested in testing a hypothesis of the form \( H_0 : A\beta = a \), where \( A \) is a \( d \times p \) matrix with \( d \leq p \). Then a Wald-type statistic for testing \( H_0 \) under model (1.1) is given by \( \hat{W}_n = n(A\beta_0 - a)^\tau (A\hat{\Lambda}_{22}A^\tau)^{-1}(A\beta_0 - a) \), where \( \hat{\Lambda}_{22} \) is the \( p \times p \) matrix obtained from \( \hat{\Lambda} \) by excluding its first row and first column.

According to Theorem 3, \( \hat{W}_n \xrightarrow{d} \chi_p^2 \) under model (1.1) as \( n \to \infty \).

4.2. Retrospective analysis of \((\tilde{\theta}, \tilde{\beta})\)

In retrospective analysis under model (1.2), \((\tilde{\theta}, \tilde{\beta})\) is a solution to \( D_n(\theta, \beta) = n^{-1} \sum_{i=1}^{n_0} d(X_i, 0, \theta, \beta) + n^{-1} \sum_{j=1}^{n_1} d(Z_j, 1, \theta, \beta) = 0 \) with \( \theta = \alpha + \log \rho \) and \( \bar{\theta} = \tilde{\alpha} + \log \rho \). Let \( C(\theta, \beta, G) \) be the \((p+1) \times (p+1)\) matrix defined by \( C(\theta, \beta, G) = (c_{ij}(\theta, \beta, G))_{i,j=1,\ldots,p+1} \) with \( c_{ij}(\theta, \beta, G) \) given in (3.2). The asymptotic distribution of \((\tilde{\theta}, \tilde{\beta})\) under model (1.2) is given in the following theorem.

Theorem 4. Suppose that model (1.2) and assumption (2.2) hold, and that \( d(t, 1, \theta, \beta) \) is twice continuously differentiable in \((\theta, \beta)\) for every \( t \) and the second-order partial derivatives of \( d(t, 1, \theta, \beta) \) are dominated by a function \( \psi(t) \) satisfying \( \int \psi(t) dG(t) < \infty \) for every \((\theta, \beta)\) in a neighborhood of \((0, 0)\). Suppose further that \( B(0, 0, G) \) is nonsingular, \( \|B(\theta_0, 0, G)\| < \infty \), \( \int \|t\|^2 dG(t) < \infty \), and \( \int \|d(t, 1, \theta, \beta_0)\|^2 \exp(\theta_0 + \beta_0^\tau t)[1 + \exp(\theta_0 + \beta_0^\tau t)] dG(t) < \infty \). Then

(a) As \( n \to \infty \), with probability tending to 1 there exists a sequence of roots \((\tilde{\theta}, \tilde{\beta})\) of the system of equations \( D_n(\theta, \beta) = 0 \) such that \((\tilde{\theta}, \tilde{\beta})\) is consistent for estimating \((0, 0)\).

(b) As \( n \to \infty \), one can write \((\tilde{\theta} - \theta_0, \tilde{\beta} - \beta_0)^\tau = -B^{-1}(0, 0, G)D_n(\theta_0, \beta_0) + o_p\left(n^{-1/2}\right)\). As a result, \( \sqrt{n}(\tilde{\theta} - \theta_0, \tilde{\beta} - \beta_0)^\tau \xrightarrow{d} N_{p+1}(0, \Omega) \) as \( n \to \infty \), where \( \Omega = B^{-1}(0, 0, G) C(0, 0, G)\{B^{-1}(0, 0, G)\}^\tau - M \) with \( B(\theta, \beta, G) \) defined in (3.2) and \( M = (m_{ij})_{i,j=1,\ldots,p+1} \) having \( m_{11} = (1 + \rho)^2 / \rho \) and all other \( m_{ij} = 0 \).
The proof of Theorem 4 is given in the appendix. Based on the point estimator \((\hat{\theta}, \hat{\beta})\), we propose to estimate \(G\) under model (1.2) by \(\hat{G}(t) = n_0^{-1} \sum_{i=1}^{n} [I(T_i \leq t) \{1 + \exp(\hat{\theta} + \hat{\beta}^T T_i)\}]\). With \((\hat{\theta}, \hat{\beta})\) in place of \(\hat{\theta}, \hat{\beta}\), \(\hat{G}(t)\) is an alternative to the maximum semiparametric likelihood estimator \(\tilde{G}(t)\) of Qin and Zhang [11] for estimating \(G\) under model (1.2). Let

\[
\begin{align*}
\tilde{B} &= B(\hat{\theta}, \hat{\beta}, \tilde{G}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{1 + \exp(\tilde{\theta} + \tilde{\beta}^T t_i)} d(t_i, 0, \tilde{\theta}, \tilde{\beta})(t_i')^\tau, \\
\tilde{C} &= C(\tilde{\theta}, \tilde{\beta}, \tilde{G}) = \frac{1}{n} \sum_{i=1}^{n} d(t_i, 1, \tilde{\theta}, \tilde{\beta}) d^\tau(t_i, 1, \tilde{\theta}, \tilde{\beta}) \exp(\tilde{\theta} + \tilde{\beta}^T t_i).
\end{align*}
\]

Then \((\tilde{B}, \tilde{C})\) is the corresponding empirical version of \((B(0_0, 0_0, G), C(0_0, 0_0, G))\) with \((0_0, 0_0, G)\) replaced by \((\hat{\theta}, \hat{\beta}, \tilde{G})\). It can be shown that \((\tilde{B}, \tilde{C})\) is a consistent estimator of \((B(0_0, 0_0, G), C(0_0, 0_0, G))\) under model (1.2). Consequently, the asymptotic covariance matrix \(\Omega\) can be consistently estimated by \(\hat{\Omega} = \tilde{B}^{-1} \tilde{C}(\tilde{B}^{-1})^\tau = M\). A Wald-type statistic for testing \(H_0 : A\beta = a\) under model (1.2) is given by \(\hat{W}_n = n(A\tilde{\beta} - a)^\tau(\hat{A}\Omega_{22} A^\tau)^{-1}(A\tilde{\beta} - a)\), where \(\hat{A}\Omega_{22}\) is the \(p \times p\) matrix obtained from \(\hat{\Omega}\) by excluding its first row and first column. According to Theorem 4, \(\hat{W}_n \overset{d}{\rightarrow} \chi^2_p\) under model (1.2) as \(n \rightarrow \infty\).

It is seen from (3.1) and (4.1) that \(\hat{B} = \tilde{B}\) and \(\hat{C} = \tilde{C}\) since \(\tilde{\theta} = \tilde{\theta}\). As a result, \(\hat{\Lambda} = \hat{B}^{-1} \hat{C}(\hat{B}^{-1})^\tau = \hat{B}^{-1} \hat{C}(\hat{B}^{-1})^\tau = M + M.\) If \(\{\tilde{B}^{-1} \tilde{C}(\hat{B}^{-1})^\tau\}_{22}\) is the \(p \times p\) matrix obtained from \(\tilde{B}^{-1} \tilde{C}(\hat{B}^{-1})^\tau\) by excluding its first row and first column, then \(\hat{\Lambda}_{22} = \{\hat{B}^{-1} \hat{C}(\hat{B}^{-1})^\tau\}_{22} = \hat{\Omega}_{22}\) and \(\hat{W}_n = n(A\tilde{\beta} - a)^\tau(\hat{A}\hat{\Omega}_{22} A^\tau)^{-1}(A\tilde{\beta} - a) = n(A\tilde{\beta} - a)^\tau(\hat{A}\hat{\Omega}_{22} A^\tau)^{-1}(A\tilde{\beta} - a) = \hat{W}_n\). Consequently, the prospectively estimated asymptotic covariance matrix \(\hat{\Lambda}\) of \(\tilde{\beta}\) under model (1.1) and the retrospectively estimated asymptotic covariance matrix \(\hat{\Omega}_{22}\) of \(\tilde{\beta}\) under model (1.2) are identical. Therefore, when we solve \(D_n(\theta, \beta) = 0\) for an estimator \(\tilde{\beta}\) of \(\beta\), we do not need to distinguish under assumption (2.2) between the prospective sampling under model (1.1) and the retrospective sampling under model (1.2) in order to determine the asymptotic distribution of \(\hat{\beta}\). This is because \(\sqrt{n}(\beta - \beta_0)\) is asymptotically normal with the same mean 0 and the same estimated asymptotic covariance matrix whether conditional on \((T_1 = t_1, \ldots, T_n = t_n)\) under model (1.1) or conditional on \((Y_1 = y_1, \ldots, Y_n = y_n)\) under model (1.2). In addition, the prospective and retrospective Wald-type statistics \(\tilde{W}_n\) and \(\hat{W}_n\) have the same observed value and the same asymptotic \(\chi^2_p\) distribution.

Theorems 3 and 4 indicate that under case–control sampling scheme, \(\tilde{\beta}\) and its estimated asymptotic covariance matrix may be obtained by solving \(D_n(\theta, \beta) = 0\) under the prospective logistic regression model (1.1) as if the case–control data had been obtained from prospective sampling. Because the class of unbiased estimating equations \(D_n(\theta, \beta) = 0\) includes, as shown in the next subsection, the score equation of Prentice and Pyke [10] and the robust score equation of Wang and Carroll [13,14], Theorems 3 and 4 generalizes the results of Prentice and Pyke [10] and Wang and Carroll [13,14] to a wide class of estimating equations \(D_n(\theta, \beta) = 0\) satisfying assumption (2.2).
4.3. Examples

In this subsection, we supply three illustrative examples to demonstrate that Theorems 3 and 4 generalize several well-known results in logistic case–control studies.

**Example 3.** Let \( d(t, y, \theta, \beta) = [y - \exp(\theta + \beta^* t)]/[1 + \exp(\theta + \beta^* t)] \)\( y \)' Then assumption (2.2) is satisfied. For this choice of \( d(t, y, \theta, \beta) \), \( D_n(x^*, \beta) = 0 \) is the set of prospective score equations, whereas \( D_n(x + \log \rho, \beta) = 0 \) is the set of retrospective score equations of Prentice and Pyke [10] and Qin and Zhang [11]. Thus, Theorems 3 and 4 generalize the results of Prentice and Pyke [10] to a wide class of unbiased estimating equations of Prentice and Pyke [10] and Qin and Zhang [11]. Thus, Theorems 3 and 4 generalize to a wide class of unbiased estimating equations \( D_n(\theta, \beta) = 0 \) satisfying assumption (2.2).

**Example 4.** Let \( d(t, y, \theta, \beta, \xi) = w(t, \theta, \beta, \xi)[y - \exp(\theta + \beta^* t)]/[1 + \exp(\theta + \beta^* t)] \)\( y \) and \( D_n(\theta, \beta, \xi) = n^{-1} \sum_{i=1}^{n} d(T_i, Y_i, \theta, \beta, \xi) \), where \( \xi \) is a nuisance parameter and \( w(t, \theta, \beta, \xi) \) is a weight function not depending on \( y \). We can show that assumption (2.2) holds for each fixed \( \xi \). Let \( \hat{\xi}_0 \) be the true value of \( \xi \) and \( \hat{\xi} \) be an estimator of \( \hat{\xi}_0 \) based on the case–control sample \( \{T_1, \ldots, T_n\} \) under model (1.2). Then \( D_n(\theta, \beta, \hat{\xi}) = 0 \) is the set of retrospective robust score equations of Wang and Carroll [13]. Let \( (\hat{\theta}(\hat{\xi}), \hat{\beta}(\hat{\xi})) \) be a solution to \( D_n(\theta, \beta, \xi) = 0 \) under model (1.2) for each fixed \( \xi \). It can be shown that if \( \hat{\xi}_0 = \alpha_p (n^{-1/4}) \), then \( (\hat{\theta}(\hat{\xi}_0), \hat{\beta}(\hat{\xi}_0)) \) have the same asymptotic distribution under model (1.2) as if \( \hat{\xi}_0 \) were known. Consequently, Theorems 3 and 4 provide an extension of Wang and Carroll’s [13] results to a wide class of unbiased estimating equations \( D_n(\theta, \beta, \xi) = 0 \) satisfying assumption (2.2) for each fixed \( \xi \).

**Example 5.** Let \( d(t, y, \theta, \beta, \xi) = w(t, y, \theta, \beta, \xi)[y - \exp(\theta + \beta^* t)]/[1 + \exp(\theta + \beta^* t)] - c(t, y, \theta, \beta, \xi) t' \) and \( D_n(\theta, \beta, \xi) = n^{-1} \sum_{i=1}^{n} d(T_i, Y_i, \theta, \beta, \xi) \), where \( \xi \) is a nuisance parameter and \( w(t, \theta, \beta, \xi) \) is a response-dependent weight function as given in [14]. Here \( c(t, 0, \theta, \beta, \xi) \) and \( c(t, 1, \theta, \beta, \xi) \) are chosen such that assumption (2.2) holds for each fixed \( \xi \). Wang and Carroll [14] presented an example for the choices of \( w(t, y, \theta, \beta, \xi) \) and \( c(t, y, \theta, \beta, \xi) \).

5. Prospective and retrospective inferences for \( \beta \)

In this section, we consider the problem of testing \( H_0 : \beta = 0 \) versus \( H_1 : \beta \neq 0 \) under model (1.1) or (1.2). Under model (1.2), if \( \beta = 0 \) then \textit{a priori} \( x = 0 \) since \( h(x) \) is a density function. Prentice and Pyke [10] showed that the maximum likelihood estimator of \( \beta \) and its estimated asymptotic covariance matrix with case–control sampling may be obtained by applying the prospective logistic regression model (1.1) to the case–control study as if the data had been obtained in a prospective study. Thus, the theory of Wald tests in a prospective study may be employed to perform a significance test of a null hypothesis \( H_0 : \beta = 0 \) under model (1.1) based on case–control data. Throughout this section, suppose
that the conditions of Theorems 1 and 2 hold. Write
\[ \mu_0 = \int t \, dG(t), \quad \mu_1 = \int t \exp(\alpha + \beta^\top t) \, dG(t), \]
\[ \Sigma = \int (t - \mu_0)(t - \mu_0)^\top \, dG(t), \]
\[ S_{TT} = \sum_{i=1}^{n} (T_i - \bar{T})(T_i - \bar{T})^\top, \quad S_{YY} = \sum_{i=1}^{n} (Y_i - \bar{Y})(Y_i - \bar{Y})^\top, \]
\[ S_{TY} = \sum_{i=1}^{n} (T_i - \bar{T})(Y_i - \bar{Y}), \]
\[ \tilde{\beta}_r = nS_{TT}^{-1}(\bar{Z} - \bar{X}), \quad M_n = \frac{\rho}{(1 + \rho)^2} n^2 (\bar{Z} - \bar{X})^\top S_{TT}^{-1} (\bar{Z} - \bar{X}). \tag{5.1} \]

It can be shown as in the proof of Theorem 1 of Qin and Zhang [11] that under model (1.2) with \((\alpha, \beta) = (0, 0)\), one can write \( \tilde{\beta} = \tilde{\beta}_r + o_p(n^{-1/2}) \) as \( n \to \infty \). Thus, the maximum semiparametric likelihood estimator \( \tilde{\beta} \) is asymptotically proportional to the difference between the case and control sample means \( \bar{Z} \) and \( \bar{X} \).

Let \( z_0^g = \log \rho, x_0 = 0, \) and \( \beta_0^g = 0 \). Then \( \theta_0 = \log \rho \) under both models (1.1) and (1.2). Furthermore, let \( d(t, y, \theta_0, \beta_0) = t[y - \exp(\theta_0 + \beta_0^g t)(1 - y)] \). Then assumption (2.2) holds with \((\theta, \beta) = (\theta_0, \beta_0) = (\log \rho, 0)\) and \( k = 1 \). The random variable \( D_n(\theta_0, \beta_0) \) in (2.1) becomes
\[ D_n(\theta_0, \beta_0) = \frac{1}{n} \sum_{i=1}^{n} d(T_i, Y_i, \theta_0, \beta_0) = \sum_{i=1}^{n} T_i[Y_i - \rho(1 - Y_i)]. \]
The prospective and retrospective versions of \( D_n(\theta_0, \beta_0) \) are, respectively, given by
\[ \hat{D}_n = D_n(x_0^g, \beta_0) = \frac{S_{TY}}{n(1 - \bar{Y})} \quad \text{and} \quad \tilde{D}_n = D_n(x_0, \beta_0) = \frac{\rho}{1 + \rho}(\bar{Z} - \bar{X}). \]

It follows from Theorems 1 and 2 that under \( H_0: \beta = 0 \), \( \hat{M}_n = \rho^{-1} n^2 \hat{\beta} \hat{D}_n S_{TT}^{-1} \hat{D}_n \overset{d}{\to} \chi_p^2 \) under model (1.1) and \( \tilde{M}_n = \rho^{-1} n^2 \tilde{\beta} \tilde{D}_n S_{TT}^{-1} \tilde{D}_n \overset{d}{\to} \chi_p^2 \) under model (1.2) as \( n \to \infty \). Moreover, \( \hat{M}_n = \tilde{M}_n = M_n \).

The Wald statistic and the score statistic for testing \( H_0: \beta = 0 \) are respectively given by
\[ W_n = \frac{\rho}{(1 + \rho)^2} \bar{\beta} S_{TT} \bar{\beta} \quad \text{and} \quad M_n = \frac{\rho}{(1 + \rho)^2} \tilde{\beta}_r S_{TT} \tilde{\beta}_r. \]

As \( n \to \infty \), \( W_n \overset{d}{\to} \chi_p^2, M_n \overset{d}{\to} \chi_p^2, \) and \( W_n - M_n = o_p(1) \) under \( H_0: \beta = 0 \) in model (1.1) or (1.2). Notice that the Wald statistic \( W_n \) employs a quadratic form in the deviations between \( \bar{\beta} \) and 0, whereas the score statistic \( M_n \) uses a quadratic form in the deviations between \( \tilde{\beta}_r \) and 0 or between \( \bar{Z} \) and \( \bar{X} \). As demonstrated in the next subsection, \( \tilde{\beta}_r \) and \( M_n \) are closely related to the discriminant function approach to estimation and test of \( \beta \) under model (1.2).
5.1. Discriminant function approach

It is well known that if \( T \mid Y = j \sim N_p(\mu_j, \Sigma) \) for \( j = 0, 1 \), then the prospective logistic regression model (1.1) holds with

\[
\gamma^* = \log \left( \frac{\pi}{1 - \pi} \right) - \frac{1}{2} \left( \mu_1^* \Sigma^{-1} \mu_1^* - \mu_0^* \Sigma^{-1} \mu_0^* \right), \quad \beta = \Sigma^{-1}(\mu_1 - \mu_0).
\]

Thus, the odds-ratio parameter \( \beta \) is proportional to the difference between the case and control population means \( \mu_1 \) and \( \mu_0 \). This property also holds under model (1.2) when \( \beta \) is small. Indeed, let \( M(s) = \int e^{sx} \, dG(x) \) be the moment generating function of \( G \) and \( N(\beta) = M'(\beta)/M(\beta) \), where \( M'(\beta) = \partial M(\beta)/\partial \beta \). Then \( N(0) = \mu_0 \) and \( N'(0) = \partial N(\beta)/\partial \beta \big|_{\beta=0} = \partial^2 M(\beta)/\partial \beta \partial \beta \big|_{\beta=0} = \Sigma \). For small \( \beta \), an application of the first-order Taylor-expansion gives

\[
\mu_1 - \mu_0 = E_H(Z) - E_G(X) = e^\beta M'(\beta) - E_G(X) = N(\beta) - E_G(X) = N(0) + N'(0)\beta - E_G(X) + o(\|\beta\|) = \Sigma \beta + o(\|\beta\|),
\]

thus yielding \( \beta \approx \Sigma^{-1}(\mu_1 - \mu_0) \) under model (1.2) when \( \beta \) is small. This property reveals one attractive feature of model (1.2): although the density functions \( g(x) \) and \( h(x) \) are modeled nonparametrically, they are linked by an “exponential tilt” \( \exp(\alpha + \beta \ell \alpha x) \) with the odds ratio parameter \( \beta \) quantifying the difference between the case and control population means.

According to Hosmer and Lemeshow [8, p. 44], the discriminant function estimator of \( \beta = \Sigma^{-1}(\mu_1 - \mu_0) \) is given by \( \tilde{\beta}_d = \tilde{\Sigma}^{-1}(\tilde{Z} - \tilde{X}) \), where

\[
\tilde{\Sigma} = \frac{1}{n - 2} \left[ \sum_{i=1}^{n_0} (X_i - \tilde{X})^2 + \sum_{j=1}^{n_1} (Z_j - \tilde{Z})^2 \right]
\]

is the multivariate extension of the pooled sample variance. Under the multivariate normal assumption, \( \tilde{\beta}_d \) is the maximum likelihood estimator of \( \beta \).

The discriminant function statistic for testing \( H_0 : \beta = 0 \) under model (1.2) is given by

\[
V_n = \frac{\rho}{(1 + \rho)^2} n \tilde{\beta}_d \tilde{\Sigma} \tilde{\beta}_d = \frac{\rho}{(1 + \rho)^2} n (\tilde{Z} - \tilde{X})' \tilde{\Sigma}^{-1} (\tilde{Z} - \tilde{X}). \tag{5.2}
\]

Under \( H_0 : \beta = 0 \), we have that \( V_n \xrightarrow{d} \chi^2_p \) as \( n \to \infty \). Notice that the discriminant function statistic \( V_n \) uses a quadratic form in the deviations between \( \tilde{Z} \) and \( \tilde{X} \). It is seen from (1.1) and (1.2) that \( \tilde{\beta}_d \) and \( V_n \) employ the pooled sample variance \( \tilde{\Sigma} \), whereas \( \tilde{\beta}_r \) and \( M_n \) utilize the sample variance \( n^{-1} S_{TT} \) of \( T_1, \ldots, T_n \) under \( H_0 : \beta = 0 \).

5.2. Correlation coefficient approach

Throughout this subsection, we assume that \( p = 1 \). The aforementioned statistics \( \tilde{\beta}_r, \tilde{\beta}_d, M_n, \) and \( V_n \) pertain to the difference \( \tilde{Z} - \tilde{X} \), which can be expressed as \( \tilde{Z} - \tilde{X} = S_{TY}/S_{YY} \) after some algebra. This latter fact motivates us to examine the population correlation
coefficient between $T$ and $Y$. On the basis of the joint distribution of $(T, Y)$, we can show that the square of the population correlation coefficient $R$ between $T$ and $Y$ is given by

$$R^2 = \frac{[\text{Corr}(T, Y)]^2}{\text{Var}(T) \text{Var}(Y)} = \frac{\pi(1 - \pi)[E_H(Z) - E_G(X)]^2}{\text{Var}(T)} = \frac{\pi(1 - \pi)[E_H(Z) - E_G(X)]^2}{(1 - \pi) \text{Var}_G(X) + \pi \text{Var}_H(Z) + \pi(1 - \pi)[E_H(Z) - E_G(X)]^2} = \frac{C}{1 + C'},$$

where

$$C = \frac{\pi(1 - \pi)[E_H(Z) - E_G(X)]^2}{(1 - \pi) \text{Var}_G(X) + \pi \text{Var}_H(Z)}.$$

It can be further shown that $\beta = 0$ under model (1.2) if and only if $R^2 = 0$ or $C = 0$ unless $G$ is a degenerate distribution function. Thus, we can alternatively construct test statistics for testing $H_0 : \beta = 0$ under model (1.2) based on estimators of $R^2$ and $C$.

Based on the data $(T_1, Y_1), \ldots, (T_n, Y_n)$ generated from the cross-sectional sampling plan stipulated by model (1.2) and $\pi = P(Y = 1)$, we propose to estimate $R^2$ and $C$, respectively, by

$$R_n^2 = \frac{\frac{1}{n} \sum_{i=1}^{n_0} E_H(Z^*_i) - E_G(X^*_i))^2}{1/n \text{STT}}$$

and

$$C_n = \frac{n_0 \frac{1}{n} \sum_{i=1}^{n_0} E_H(Z^*_i) - E_G(X^*_i))^2}{n \text{Var}_G(X^*) + \frac{n_1}{n} \text{Var}_H(Z^*)},$$

where $X^* \sim \tilde{G}$ and $Z^* \sim \tilde{H}$. According to Remark 1 of Zhang [16], the sample moments of the control sample $X_1, \ldots, X_{n_0}$ and the case sample $Z_1, \ldots, Z_{n_1}$ match the moments of $\tilde{G}$ and $\tilde{H}$, respectively. As a result, we have

$$R_n^2 = \frac{\rho}{(1 + \rho)^2} \frac{n(\bar{Z} - \bar{X})^2}{\text{STT}^2} = \frac{S_{TY}^2}{S_{TT} S_{YY}},$$

$$C_n = \frac{\rho}{(1 + \rho)^2} \frac{n(\bar{Z} - \bar{X})^2}{\sum_{i=1}^{n_0} (X_i - \bar{X})^2 + \sum_{j=1}^{n_1} (Z_j - \bar{Z})^2} = \frac{\rho}{(1 + \rho)^2} \frac{n(\bar{Z} - \bar{X})^2}{(n - 2) \Sigma}$$

$$= \frac{\rho}{(1 + \rho)^2} \frac{n}{n - 2} \sum \hat{p}^2 d.$$
This connection between $\tilde{\beta}$ and $R_n$ is a reflection of that in the simple linear regression model, apart from the proportional constant $(1 + \rho)^2 / \rho$.

The relation between $M_n$ and $R_n^2$ motivates us to alternatively test $H_0 : \beta = 0$ under model (1.1) or (1.2) by employing the Fisher’s $z$-transformation statistic [5, p. 43]

$$F_n = \frac{n - 3}{4} \left[ \log \left( \frac{1 + R_n}{1 - R_n} \right) \right]^2.$$

With the aforesaid cross-sectional sampling scheme, $F_n \xrightarrow{d} \chi^2_1$ under $H_0 : \beta = 0$ as $n \to \infty$. The test statistic $F_n$ can be applied to both the prospective and the retrospective studies.

5.3. Examples

In this subsection, we consider the problem of testing $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$ under the logistic regression model (1.1) based on case–control data by applying the four test statistics $W_n$, $M_n$, $V_n$, and $F_n$ to two real data sets.

**Example 6.** Gramenzi et al. [7] reported results of a northern Italy case–control study on the relationship between cigarette smoking and myocardial infarction in women. The sample consisted of young and middle-aged women admitted to the coronary care units of 30 hospitals in northern Italy with acute myocardial infarction, as cases, and controls admitted to the same hospitals with other acute disorders. The dataset is also listed in Table 5.11 of Agresti [1, p.137]. The table classifies the cases and the controls by their smoking histories, measured in terms of average number of cigarettes per day. We shall use scores $\{0, 7.5, 19.5, 30\}$ for smoking level. Let $X$ denote the level of cigarette smoking per day and $Y = 1$ or 0 represent the presence or absence of myocardial infarction in women. Zhang [17] reported a good fit of model (1.1) to this data set based on an information matrix test.

Under model (1.2), we find that $(\tilde{z}, \tilde{\beta}) = (-0.572, 0.077)$ and that the four test statistics are $W_n = 82.3915$, $M_n = 94.2362$, $V_n = 107.3906$, and $F_n = 102.5050$ with one degree of freedom. The observed $P$-values of these four test statistics are all close to 0, indicating strong evidence of a positive smoking effect on myocardial infarction.

**Example 7.** Hosmer and Lemeshow [8] used the logistic regression model (1.1) to analyze the relationship between age and the status of coronary heart disease among 100 subjects participating in a study. The complete dataset is listed on page 3 in their book. Qin and Zhang [11] reported a good fit of model (1.1) to this data set by using a Kolmogorov–Smirnov-type statistic. Let $X$ denote age and $Y = 1$ or 0 represent the presence or absence of coronary heart disease. Since the data $(Y_i, X_i)$, $i = 1, \ldots, 100$, can be thought as being drawn independently and identically from the joint distribution of $(Y, X)$, we can apply the four test statistics $W_n$, $M_n$, $V_n$, and $F_n$ to this data set.

Under model (1.2), we have $(\tilde{z}, \tilde{\beta}) = (-5.028, 0.111)$, $W_n = 21.2541$, $M_n = 26.3989$, $V_n = 35.1502$, and $F_n = 31.2810$. With one degree of freedom, the observed $P$-values of
these four test statistics are all close to 0, indicating strong evidence of a positive age effect on coronary heart disease.

5.4. A simulation study

In this subsection, we present a simulation study to compare the performances of the four test statistics $W_n, M_n, V_n,$ and $F_n$ for testing $H_0 : \beta = 0$ by examining their powers against some local alternatives $H_1 : \beta \neq 0$ under model (1.2). In our simulation study, we consider two different sets of cases and control population distributions. In the first place, we assume that $g(x)$ and $h(x)$ are, respectively, the normal density functions of the $N(\mu_0, 1)$ and $N(\mu_1, 1)$ distributions. Then model (1.2) holds with $x = (\mu_0^2 - \mu_1^2)/2$ and $\beta = \mu_1 - \mu_0$. Let $\mu_0 = 1$ be fixed so that $x = (1 - \mu_1^2)/2$ and $\beta = \mu_1 - 1$. In the second place, we suppose that $g(x)$ and $h(x)$ are, respectively, the exponential density functions of the $E(\mu_0)$ and $E(\mu_1)$ distributions. Here $E(\mu)$ denotes an exponential distribution with density function given by $\mu \exp(-\mu x)$ for $x > 0$. Then model (1.2) holds with $x = \log \mu_1 - \log \mu_0$ and $\beta = \mu_1 - \mu_0$. Let $\mu_0 = 1$ be fixed so that $x = \log \mu_1$ and $\beta = \mu_1 - 1$. In both normal and exponential cases, the problem of testing $H_0 : \mu_1 = \mu_0$ versus $H_1 : \mu_1 \neq \mu_0$ is equivalent to that of testing $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$ under model (1.2). Let $\mu_{1n} = \mu_0 + n^{-1/2}\gamma$ and $\beta_n = \mu_{1n} - \mu_0 = n^{-1/2}\gamma$. Our aim is to compare the performances of $W_n, M_n, V_n,$ and $F_n$ by examining their powers against some local alternatives $H_1 : \beta = \beta_n$ under model (1.2).

In our simulations, we considered $\gamma = 0, 1.0, 2.0$ and sample sizes of $(n_0, n_1) = (40, 60)$ and $(n_0, n_1) = (60, 40)$. Note that for $\gamma = 0, 1.0, 2.0$, we have $\beta_n = 0, 0.1, 0.2$ when $n = 100$. For each pair $(n_0, n_1)$ and each value of $\gamma$, we generated 1000 independent sets of combined random samples from the $N(1, 1)$ and $N(\mu_{1n}, 1)$ distributions in the normal case and 1000 independent sets of combined random samples from the $E(1)$ and $E(\mu_{1n})$ distributions in the exponential case. Since $p = 1$, all four test statistics $W_n, M_n, V_n,$ and $F_n$ have the same asymptotic chi-squared distribution with one degree of freedom.

The simulation results are summarized in Tables 1 and 2. It is seen that the achieved significance levels of $W_n, M_n, V_n,$ and $F_n$ are all quite close to the corresponding nominal significance levels and the powers of $W_n, M_n, V_n,$ and $F_n$ are becoming progressively larger as $\gamma$ moves away from 0. Our simulation results also reveal that in all cases, the powers of $V_n$ are slightly greater than those of $M_n$ and $F_n$, which are in turn greater than those of $W_n$ except for a tie between $M_n$ and $V_n$ in the exponential case with $(n_0, n_1) = (40, 60), \gamma = 1.0,$ and nominal significance level equal to 0.1. In summary, our simulation study indicates that the test statistics $M_n, V_n,$ and $F_n$ are quite comparable to each other in power and are superior to the other test statistic $W_n$ in terms of their power performances.

Appendix. Proofs

Proof of Theorem 2. For $k = 1, \ldots, q$, let $\xi_k(\theta, \beta) = n^{-1} \sum_{i=1}^n d_k(T_i, Y_i, \theta, \beta) = n^{-1} \sum_{i=1}^{n_0} d_k(X_i, 0, \theta, \beta) + n^{-1} \sum_{j=1}^{n_1} d_k(Z_j, 1, \theta, \beta)$. Under assumption (2.2), we have
Table 1
Achieved significance levels and powers in the normal case

<table>
<thead>
<tr>
<th>( \gamma )</th>
<th>((n_0, n_1))</th>
<th>( \mu_{1n} )</th>
<th>Nominal level</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>( W_n )</td>
<td>( M_n )</td>
</tr>
<tr>
<td>0</td>
<td>(40, 60)</td>
<td>1.0</td>
<td>0.10</td>
<td>0.089</td>
</tr>
<tr>
<td>0</td>
<td>(40, 60)</td>
<td>1.0</td>
<td>0.05</td>
<td>0.044</td>
</tr>
<tr>
<td>0</td>
<td>(40, 60)</td>
<td>1.0</td>
<td>0.01</td>
<td>0.005</td>
</tr>
<tr>
<td>1</td>
<td>(40, 60)</td>
<td>1.1</td>
<td>0.10</td>
<td>0.129</td>
</tr>
<tr>
<td>1</td>
<td>(40, 60)</td>
<td>1.1</td>
<td>0.05</td>
<td>0.057</td>
</tr>
<tr>
<td>1</td>
<td>(40, 60)</td>
<td>1.1</td>
<td>0.01</td>
<td>0.010</td>
</tr>
<tr>
<td>2</td>
<td>(40, 60)</td>
<td>1.2</td>
<td>0.10</td>
<td>0.252</td>
</tr>
<tr>
<td>2</td>
<td>(40, 60)</td>
<td>1.2</td>
<td>0.05</td>
<td>0.140</td>
</tr>
<tr>
<td>2</td>
<td>(40, 60)</td>
<td>1.2</td>
<td>0.01</td>
<td>0.034</td>
</tr>
<tr>
<td>0</td>
<td>(60, 40)</td>
<td>1.0</td>
<td>0.10</td>
<td>0.097</td>
</tr>
<tr>
<td>0</td>
<td>(60, 40)</td>
<td>1.0</td>
<td>0.05</td>
<td>0.042</td>
</tr>
<tr>
<td>0</td>
<td>(60, 40)</td>
<td>1.0</td>
<td>0.01</td>
<td>0.009</td>
</tr>
<tr>
<td>1</td>
<td>(60, 40)</td>
<td>1.1</td>
<td>0.10</td>
<td>0.139</td>
</tr>
<tr>
<td>1</td>
<td>(60, 40)</td>
<td>1.1</td>
<td>0.05</td>
<td>0.068</td>
</tr>
<tr>
<td>1</td>
<td>(60, 40)</td>
<td>1.1</td>
<td>0.01</td>
<td>0.014</td>
</tr>
<tr>
<td>2</td>
<td>(60, 40)</td>
<td>1.2</td>
<td>0.10</td>
<td>0.254</td>
</tr>
<tr>
<td>2</td>
<td>(60, 40)</td>
<td>1.2</td>
<td>0.05</td>
<td>0.157</td>
</tr>
<tr>
<td>2</td>
<td>(60, 40)</td>
<td>1.2</td>
<td>0.01</td>
<td>0.042</td>
</tr>
</tbody>
</table>

Table 2
Achieved significance levels and powers in the exponential case

<table>
<thead>
<tr>
<th>( \gamma )</th>
<th>((n_0, n_1))</th>
<th>( \mu_{1n} )</th>
<th>Nominal level</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>( W_n )</td>
<td>( M_n )</td>
</tr>
<tr>
<td>0</td>
<td>(40, 60)</td>
<td>1.0</td>
<td>0.10</td>
<td>0.093</td>
</tr>
<tr>
<td>0</td>
<td>(40, 60)</td>
<td>1.0</td>
<td>0.05</td>
<td>0.033</td>
</tr>
<tr>
<td>0</td>
<td>(40, 60)</td>
<td>1.0</td>
<td>0.01</td>
<td>0.003</td>
</tr>
<tr>
<td>1</td>
<td>(40, 60)</td>
<td>1.1</td>
<td>0.10</td>
<td>0.133</td>
</tr>
<tr>
<td>1</td>
<td>(40, 60)</td>
<td>1.1</td>
<td>0.05</td>
<td>0.064</td>
</tr>
<tr>
<td>1</td>
<td>(40, 60)</td>
<td>1.1</td>
<td>0.01</td>
<td>0.014</td>
</tr>
<tr>
<td>2</td>
<td>(40, 60)</td>
<td>1.2</td>
<td>0.10</td>
<td>0.229</td>
</tr>
<tr>
<td>2</td>
<td>(40, 60)</td>
<td>1.2</td>
<td>0.05</td>
<td>0.134</td>
</tr>
<tr>
<td>2</td>
<td>(40, 60)</td>
<td>1.2</td>
<td>0.01</td>
<td>0.028</td>
</tr>
<tr>
<td>0</td>
<td>(60, 40)</td>
<td>1.0</td>
<td>0.10</td>
<td>0.083</td>
</tr>
<tr>
<td>0</td>
<td>(60, 40)</td>
<td>1.0</td>
<td>0.05</td>
<td>0.034</td>
</tr>
<tr>
<td>0</td>
<td>(60, 40)</td>
<td>1.0</td>
<td>0.01</td>
<td>0.006</td>
</tr>
<tr>
<td>1</td>
<td>(60, 40)</td>
<td>1.1</td>
<td>0.10</td>
<td>0.118</td>
</tr>
<tr>
<td>1</td>
<td>(60, 40)</td>
<td>1.1</td>
<td>0.05</td>
<td>0.054</td>
</tr>
<tr>
<td>1</td>
<td>(60, 40)</td>
<td>1.1</td>
<td>0.01</td>
<td>0.008</td>
</tr>
<tr>
<td>2</td>
<td>(60, 40)</td>
<td>1.2</td>
<td>0.10</td>
<td>0.201</td>
</tr>
<tr>
<td>2</td>
<td>(60, 40)</td>
<td>1.2</td>
<td>0.05</td>
<td>0.107</td>
</tr>
<tr>
<td>2</td>
<td>(60, 40)</td>
<td>1.2</td>
<td>0.01</td>
<td>0.014</td>
</tr>
</tbody>
</table>
\( \hat{d}_k(t, 0, \theta, \beta)/\eta + \exp(\theta + \beta^T t) \{ \hat{d}_k(t, 1, \theta, \beta)/\eta \} = d_k(t, 0, \theta, \beta)t' \) with \( \eta = (\theta, \beta)^T \).

Applying a first-order Taylor expansion yields

\[
\xi_k(\tilde{\theta}, \tilde{\beta}) = \xi_k(\theta_0, \beta_0) + \frac{\partial \xi_k(\theta_0, \beta_0)}{\partial \theta}(\tilde{\theta} - \theta_0) + \frac{\partial \xi_k(\theta_0, \beta_0)}{\partial \beta^T} (\tilde{\beta} - \beta_0) + o_p(\delta_n)
\]

\[
= \xi_k(\theta_0, \beta_0) + b^*_k(\theta_0, \beta_0, G) \left( \tilde{\theta} - \theta_0 \right) + o_p(\delta_n)
\]

\[
= \xi_k(\theta_0, \beta_0) + \frac{1}{n} b^*_k(\theta_0, \beta_0, G) S^{-1}(\theta_0, \beta_0, G) \times Q_n(\theta_0, \beta_0) + o_p(\delta_n), \quad k = 1, \ldots, q,
\]

where \( \delta_n = ||\tilde{\theta} - \theta_0|| + ||\tilde{\beta} - \beta_0|| = O_p \left( n^{-1/2} \right) \). It can be shown after very extensive algebra that under model (1.2) and assumption (2.2), we have

\[
\sqrt{n} E \left\{ \xi_k(\theta_0, \beta_0) + \frac{1}{n} b^*_k(\theta_0, \beta_0, G) S^{-1}(\theta_0, \beta_0, G) Q_n(\theta_0, \beta_0) \right\} = 0, \quad k = 1, \ldots, q,
\]

\[
n \text{Cov} \left\{ \xi_i(\theta_0, \beta_0) + \frac{1}{n} b^*_i(\theta_0, \beta_0, G) S^{-1}(\theta_0, \beta_0, G) Q_n(\theta_0, \beta_0), \right\}
\]

\[
= c_{ij}(\theta_0, \beta_0, G) - \sigma_{ij}, \quad 1 \leq i, j \leq q.
\]

It now follows from the multivariate central limit theorem and Slutsky’s theorem that

\[
\sqrt{n} \tilde{D}_n = \sqrt{n} D_n(\theta_0, \beta_0) + \frac{1}{\sqrt{n}} B^*(\theta_0, \beta_0, G) S^{-1}(\theta_0, \beta_0, G) Q_n(\theta_0, \beta_0)
\]

\[
+ o_p(1) \to N_q(0, \Sigma)
\]

in distribution as \( n \to \infty \), thus establishing Theorem 2. \( \square \)

**Proof of Theorem 4.** Part (a) can be proved by employing a similar approach as in the proof of the consistency of \((\hat{\theta}, \hat{\beta})\) obtained from the logistic score equations in Prentice and Pyke [10]. For part (b), applying a first-order Taylor expansion gives

\[
0 = \xi_k(\bar{\theta}, \bar{\beta}) = \xi_k(\theta_0, \beta_0) + \frac{\partial \xi_k(\theta_0, \beta_0)}{\partial \theta}(\bar{\theta} - \theta_0) + \frac{\partial \xi_k(\theta_0, \beta_0)}{\partial \beta^T} (\bar{\beta} - \beta_0) + o_p(\delta_n)
\]

\[
= \xi_k(\theta_0, \beta_0) + b^*_k(\theta_0, \beta_0, G) \left( \bar{\theta} - \theta_0 \right) + o_p(\delta_n), \quad k = 1, \ldots, p + 1.
\]
As a result, it follows from Slutsky’s theorem that

\[ \sqrt{n}\left(\bar{\theta} - \theta_0\right) = -B^{-1}(\theta_0, \beta_0, G) \sqrt{n} D_n(\theta_0, \beta_0) + o_p(1) \to N_{p+1}(0, \Omega) \] in distribution as \( n \to \infty \),

where

\[ \Omega = B^{-1}(\theta_0, \beta_0, G)\Psi \{B^{-1}(\theta_0, \beta_0, G)\}^\tau \]

\[ = B^{-1}(\theta_0, \beta_0, G)C(\theta_0, \beta_0, G)\{B^{-1}(\theta_0, \beta_0, G)\}^\tau - \frac{(1 + \rho)^2}{\rho} B^{-1}(\theta_0, \beta_0, G)A_0\{B^{-1}(\theta_0, \beta_0, G)A_0\}^\tau \]

\[ = B^{-1}(\theta_0, \beta_0, G)C(\theta_0, \beta_0, G)\{B^{-1}(\theta_0, \beta_0, G)\}^\tau - \frac{(1 + \rho)^2}{\rho} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \]

The proof is completed. \( \square \)

Acknowledgments

The author wishes to thank a referee for several constructive and helpful comments, which led to an improved presentation. This research was supported in part by a grant from the National Security Agency.

References