Ann. Inst. Statist. Math.
Vol. 53, No. 2, 217–243 (2001)

EMPIRICAL BEST PREDICTION FOR SMALL AREA INFERENCE WITH BINARY DATA

JIMING JIANG[1] AND P. LAHIRI[2]

[1]*Department of Statistics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-7054, U.S.A.*
[2]*Department of Mathematics and Statistics, University of Nebraska at Lincoln, Lincoln, NE 68588-0323, U.S.A.*

(Received December 9, 1998; revised September 27, 1999)

Abstract. The paper introduces a frequentist's alternative to the recently developed hierarchical Bayes methods for small area estimation with binary data. Specifically, the best predictor (BP) and empirical best predictor (EBP) of small area specific random effect are developed in the context of a mixed logistic model and different asymptotic properties of the proposed BP and EBP are studied. An approximation to the mean squared error (MSE) of the proposed EBP correct up to the order $o(m^{-1})$ is obtained, where m denotes the number of small areas. The asymptotic behavior of the relative savings loss (RSL) demonstrates the superiority of the proposed EBP over the usual small area proportion.

Key words and phrases: Asymptotics, composite estimation, empirical best predictor, Laplace approximation, method of moments, mixed models, MSE.

1. Introduction

The surveys are usually designed to produce reliable estimates of various characteristics of interest for large geographic areas. However, for effective planning of health, social and other services and for apportioning government funds, there is a growing demand to produce similar estimates for small geographic areas and subpopulations.

Clearly, the usual design-based estimator which uses only the sample survey data for the particular small-area of interest is unreliable due to relatively small samples that are available from the area. In the absence of reliable small area design-based estimator, one may alternatively use synthetic estimator (see Ghosh and Rao (1994)) which utilizes data from censuses or administrative records to obtain estimates for small geographical areas. Although synthetic estimators have small variances compared to direct survey estimators, they tend to be biased as they do not use the information on the characteristic of interest directly obtainable from sample surveys.

A compromise between the direct survey and the synthetic estimation is the method of the composite estimation (see Holt *et al.* (1979)) which uses sample survey data in conjunction with different census and administrative data. The method uses either *implicit* or *explicit* models which *borrow strength* from related sources. See Ghosh and Rao (1994) for a thorough review of different composite estimation techniques.

The focus of this paper is to develop an efficient composite small area estimation method for binary data when the sampling design is ignorable. There are some interesting research done in this area using hierarchical Bayes methodology (see Malec *et al.* (1997); Ghosh *et al.* (1998); Farrell *et. al.* (1997)). However, the methods are generally computer intensive and could be hard to implement routinely in a large scale operation which are generally handled by people with moderate level of statistics background. For continuous variables for which mixed linear models are appropriate, Prasad and Rao (1990) proposed a frequentist's empirical best linear unbiased prediction (EBLUP) method to estimate small-area characteristics. The Prasad-Rao method is quite appealing to a practitioner since it is very simple to implement for a real life problem. Lahiri and Rao (1995) extended the Prasad-Rao method to a non-normal mixed linear model. However, the Prasad-Rao type simple method is not available for nonlinear mixed models (*e.g.* mixed logistic model).

The purpose of this paper is to develop the Prasad-Rao type frequentist's alternative to the already existing hierarchical Bayes methods. Let y_{ij} denote a binary response (*i.e.*, 0 or 1) corresponding to the j-th observation in the i-th small-area, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$. We assume the following hierarchical model for y_{ij}:

Model. (i) Conditional on p_{ij}, y_{ij}'s are independent Bernoulli random variables with $P(y_{ij} = 1 \mid p_{ij}) = p_{ij}$, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$.

(ii) Conditional on the intercept α_i, $\text{logit}(p_{ij}) = \log[p_{ij}/(1-p_{ij})] = x_{ij}^t\beta + \alpha_i, i = 1, \ldots, m$, $j = 1, \ldots, n_i$, where x_{ij} is a vector of $p \times 1$ known covariates and $\beta = (\beta_k)_{1\le k\le p}$ is the vector of regression coefficients.

(iii) Marginally, $\alpha_1, \ldots, \alpha_m$ are iid $N(0, \sigma^2)$, where σ^2 is an unknown variance.

The total sample size $N = \sum_{i=1}^m n_i$ is generally large. However, in a typical small area problem, n_i could be either very small or moderately large (but relatively much smaller than the total sample size N so one can justify the preference of an estimator which is more sophisticated than the direct survey estimator). Thus, so far as the asymptotics are concerned, it makes sense to study asymptotics under the two situations: (i) n_i is bounded, and (ii) n_i tends to ∞. In all the cases, we shall assume that m, the number of small areas, tends to ∞.

Section 2 presents the best predictor (BP) and empirical best predictor (EBP) of the random effect α_i. Section 3 discusses the behavior of the BP and EBP when σ tends to 0 or ∞. In Section 4, we study the asymptotic behavior of the proposed BP and EBP. In order to have the consistency property of the BP or EBP of α_i, we must have n_i tending to ∞.

An approximation to the MSE of the proposed EBP which is correct up to the order $o(m^{-1})$ is given in Section 5. In this section, we also study the asymptotic properties of the method of moments estimator of $\theta = (\beta^t, \sigma)'$. We would like to emphasize that unlike estimators of variance components in a mixed linear model (e.g., Searle *et al.* (1992)), here the method of moments estimator of σ cannot be negative, and so one never has to truncate the estimator at 0. We then propose an estimator of the MSE for which the bias is of order $o(m^{-1})$.

In Section 6, we extend the results of Section 5 to predict a mixed effect. We note that our proposed EBP is asymptotically better than the sample proportion in terms of relative savings loss introduced by Efron and Morris (1973).

A discussion section is added in Section 7. The technical proofs of lemmas and theorems are deferred to Section 8.

2. An empirical best predictor

The mixed logistic model introduced in the previous section can be written as

$$\text{logit}[P(y_{ij} = 1 \mid \alpha)] = x_{ij}^t\beta + \alpha_i, \tag{2.1}$$

$i = 1, \ldots, m$, $j = 1, \ldots, n_i$ with $\alpha = (\alpha_i)_{1\le i\le m}$. Let $\Theta = \{\theta = (\beta^t\sigma)^t : \beta \in R^p, \sigma \ge 0\}$ be the parameter space.

Let $y = (y_{ij})_{1\le i\le m, 1\le j\le n_i}$, $y_i = (y_{ij})_{1\le j\le n_i}$, $y_{i-} = (y_{i'})_{i'\ne i}$, and $y_{i\cdot} = \sum_{j=1}^{n_i} y_{ij}$. Then (y_i, α_i) is independent of y_{i-}. Therefore, the best predictor (BP) of α_i , in term of the MSE, is given by

$$\begin{aligned} E(\alpha_i \mid y) &= E(\alpha_i \mid y_i) \\ &= \sigma_0 \frac{E\xi \exp(\phi_i(y_{i\cdot}, \sigma_0\xi, \beta_0))}{E\exp(\phi_i(y_{i\cdot}, \sigma_0\xi, \beta_0))} \equiv \psi_i(y_{i\cdot}, \theta_0), \end{aligned} \tag{2.2}$$

where $\theta_0 = (\beta_0^t\sigma_0)^t$ is the vector of true parameters; $\phi_i(k, u, v) = ku - \sum_{j=1}^{n_i} \log(1 + \exp(x_{ij}^t v + u))$; and $\xi \sim N(0, 1)$.

An empirical best predictor (EBP), $\hat{\alpha}_i$, is obtained by replacing the unknown vector θ_0 by a consistent estimator, $\hat{\theta}$, i.e.,

$$\hat{\alpha}_i = \psi_i(y_{i\cdot}, \hat{\theta}). \tag{2.3}$$

3. Asymptotic behavior of the BP (EBP) when $\sigma \to 0$ or ∞

In this section, we consider the asymptotic behavior of the BP when σ^2, the variance of the random effects, goes to 0 or ∞, while the data is held fixed. This means that we are interested in the behavior of $E(\alpha_i \mid y)$ when the data comes from a population described by (2.1), where the random effects have a very small or very large variance. For notational simplicity, we write, in this section, $\theta = \theta_0$.

By (2.2), it is easy to show that when $\sigma \to 0$, $\psi_i(y_{i\cdot}, \theta)/\sigma \to E\xi = 0$. Therefore, as $\sigma \to 0$, $E(\alpha_i \mid y) = o(\sigma)$ and hence $\to 0$.

We now consider the behavior of the BP when $\sigma \to \infty$.

If $1 \le k \le n_i - 1$, both $\exp(\phi_i(k, u, \beta))$ and $u\exp(\phi_i(k, u, \beta)) \in L^1$. Thus, by the dominated convergence theorem, we have

$$\begin{aligned} \psi_i(k, \theta) &= \frac{\int u \exp(\phi_i(k, u, \beta)) \exp(-u^2/2\sigma^2) du}{\int \exp(\phi_i(k, u, \beta)) \exp(-u^2/2\sigma^2) du} \\ &\to \frac{\int u \exp(\phi_i(k, u, \beta)) du}{\int \exp(\phi_i(k, u, \beta)) du}. \end{aligned} \tag{3.1}$$

Also, by the last expression in (2.2) for ψ_i, it is easy to show that

$$\psi_i(0, \theta) \to -\infty \quad \text{and} \quad \psi_i(n_i, \theta) \to \infty. \tag{3.2}$$

We now consider a special case of model (2.1) for which the limit in (3.1) has a closed form expression. Suppose that in model (2.1) one has

$$x_{ij} = x_i, \quad j = 1, \ldots, n_i, \tag{3.3}$$

i.e., the covariates are at the small area (e.g., county) level. A simple case of this model is the following:

$$\text{logit}(P(y_{ij} = 1 \mid \alpha)) = \mu + \alpha_i, \tag{3.4}$$

$i = 1, \ldots, m$, $j = 1, \ldots, n$. For a fixed i, write $\mu = x_i^t\beta$ and $n = n_i$. Then $\phi_i(k, u, \beta) = ku - n\log(1 + \exp(\mu + u))$, which is exactly what one has under (3.4). It can be shown that under (3.3) one has

$$\psi_i(k, \theta) \to \sum_{l=1}^{k-1}\left(\frac{1}{l}\right) - \sum_{l=1}^{n_i-k-1}\left(\frac{1}{l}\right) - x_i^t\beta, \quad 1 \le k \le n_i - 1, \quad n_i \ge 2, \tag{3.5}$$

where $\sum_1^0(\cdot)$ is defined as 0. The proof of (3.5), given in Section 8, is based on the following formulas of integrals which are not normally seen in a calculus book or table of integrals. We derive these formulas in Section 8.

LEMMA 3.1. *For $n \ge 2$ and $1 \le k \le n - 1$,*

$$\int_0^\infty \frac{x^{k-1}}{(1+x)^n}dx = \frac{(k-1)!(n-k-1)!}{(n-1)!}, \quad 1 \le k \le n-1, \quad \text{and}$$

$$\int_0^\infty \log x\frac{x^{k-1}}{(1+x)^n}dx = \frac{(k-1)!(n-k-1)!}{(n-1)!}\left(\sum_{l=1}^{k-1}\left(\frac{1}{l}\right) - \sum_{l=1}^{n-k-1}\left(\frac{1}{l}\right)\right).$$

To see what the right side of (3.5) means, suppose that both k and $n_i - k$ are large. Then, $\sum_{l=1}^{k-1}(1/l) \sim \log k + C$, $\sum_{l=1}^{n_i-k-1}(1/l) \sim \log(n_i - k) + C$, where C is the Euler's constant. Therefore, as $\sigma \to \infty$, we have

$$E(\alpha_i \mid y) \approx \text{logit}(\bar{y}_{i\cdot}) - x_i^t\beta, \tag{3.6}$$

$1 \le y_i \le n_i - 1$, $n_i \ge 2$. In view of (3.2), (3.6) holds even if $y_{i\cdot} = 0$ or n_i.

When β is replaced by an estimator $\hat{\beta}$, $\hat{\alpha}_i = \psi_i(y_{i\cdot}, (\hat{\beta}^t\sigma)^t)$ has the same limiting properties as $\sigma \to 0$ or ∞, with β in the limit replaced by $\hat{\beta}$. For example, in case of (3.3), if both $y_{i\cdot}$ and $n_i - y_{i\cdot}$ are large, we have, as $\sigma \to \infty$,

$$\hat{\alpha}_i \approx \text{logit}(\bar{y}_{i\cdot}) - x_i^t\hat{\beta}. \tag{3.7}$$

It might be interesting to compare (3.7) with the corresponding result for the linear case, as $\sigma \to \infty$. Consider, for example, the linear mixed model $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, where ϵ_{ij} are i.i.d. errors with variance τ^2. Then, as σ^2, the variance of the α_i's, tends to ∞, the best linear unbiased predictor, or BLUP (e.g., Searle *et al.* (1992), §7.4), of α_i goes to $\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}$. Note that in this case, $\hat{\mu} = \bar{y}_{\cdot\cdot}$ is the estimator of μ.

4. Asymptotic behavior of EBP when $m \to \infty$

Let i be a selected index, $1 \le i \le m$. In an asymptotic setting, it makes sense to assume that i depends on m, i.e., $i = i(m)$. For example, $i = m$.

We first consider the case when $n_i \to \infty$ as $m \to \infty$. For the most part, we show that in this case, $\hat{\alpha}_i$ is a consistent estimator of α_i with certain convergence rate. Write

$$\hat{\alpha}_i - \alpha_i = [\psi_i(y_{i\cdot}, \hat{\theta}) - \psi_i(y_{i\cdot}, \theta_0)] + [\psi_i(y_{i\cdot}, \theta_0) - \tilde{\alpha}_i] + [\tilde{\alpha}_i - \alpha_i], \tag{4.1}$$

where $\tilde{\alpha}_i$ is the maximizer of $\phi_i(y_{i\cdot}, u, \beta_0)$ with respect to u. The idea is to show that the three terms on the right side of (4.1) are $O_p(|\hat{\theta} - \theta_0|)$, $O_p(1/n_i)$, and $O_p(1/\sqrt{n_i})$, respectively, as the following three lemmas state. We assume, throughout this section, that there is a constant $A > 0$ such that

$$\max_{1 \le j \le n_i} |x_{ij}| \le A. \tag{4.2}$$

LEMMA 4.1. *There is a constant B_0 such that for any $B \ge B_0$, there is a constant $C = C_B$ and an integer N_B such that for any $n_i \ge N_B$,*

$$P(n_i \mid \psi_i(y_{i\cdot}, \theta_0) - \tilde{\alpha}_i| > C) \le \frac{1}{4\delta^2 n_i} + 2\left(1 - \Phi\left(\frac{\text{logit}(1-2\delta) - A|\beta_0|}{\sigma_0}\right)\right), \tag{4.3}$$

where $\delta = e^{A|\beta_0|}/(e^{A|\beta_0|} + e^B)$, and $\Phi(\cdot)$ is the cdf of $N(0,1)$.

LEMMA 4.2. *For any $B > 2A|\beta_0| + \log(2e^{A|\beta_0|} + 1)$, we have*

$$\begin{aligned} P(\sqrt{n_i}|\tilde{\alpha}_i - \alpha_i| > 4Be^{2B}) &\le \frac{1}{4}\left(\frac{1}{\delta^2 n_i} + \frac{1}{B^2}\right) \\ &+4\left(1 - \frac{1}{2}\left[\Phi\left(\frac{B - A|\beta_0|}{\sigma_0}\right) + \Phi\left(\frac{\text{logit}(1-2\delta) - A|\beta_0|}{\sigma_0}\right)\right]\right), \end{aligned} \tag{4.4}$$

where $\delta = e^{2A|\beta_0|}/(e^{2A|\beta_0|} + e^B)$.

LEMMA 4.3. *There exists a $B_0 > 0$ such that for any $0 < \delta < \sigma_0$ and $B > B_0$, there are constants $C_r = C_r(\delta, B)$, $r = 1, 2$, and an integer $N(\delta, B)$ such that for any $n_i \ge N(\delta, B)$,*

$$\begin{aligned} &P(|\psi_i(y_{i\cdot}, \hat{\theta}) - \psi_i(y_{i\cdot}, \theta_0)| > (C_1 + C_2 n_i^{-1})|\hat{\theta} - \theta_0|) \\ &\le P(|\hat{\theta} - \theta_0| > \delta) + 2(p+2)\left[\frac{1}{4\rho^2 n_i} + 2\left(1 - \Phi\left(\frac{\text{logit}(1-2\rho) - A|\beta_0|}{\sigma_0}\right)\right)\right], \end{aligned} \tag{4.5}$$

where $\rho = \exp(A(|\beta_0| + \delta))/(\exp(A(|\beta_0| + \delta)) + \exp(B(\sigma_0 - \delta)))$.

Combining these lemmas, we obtain the following.

THEOREM 4.1. *Suppose $\sigma_0 > 0$. There exists a constant $B_0 > 0$ such that for any $0 < \delta < \sigma_0$ and $B > B_0$, there are constants $c_1(\delta, B)$, $c_2(B)$, and $d_i(\delta, B)$, $i = 1, 2$, and an integer $N(\delta, B)$ with the following properties.*

i) *$d_2(\delta, B) \longrightarrow 0$, as $\delta \to 0$ and $B \to \infty$.*

ii) *For any $n_i \ge N(\delta, B)$,*

$$\begin{aligned} &P\left(|\hat{\alpha}_i - \alpha_i| \le c_1(\delta, B)|\hat{\theta} - \theta_0| + \frac{c_2(B)}{\sqrt{n_i}}\right) \\ &\ge 1 - P(|\hat{\theta} - \theta_0| > \delta) - \frac{d_1(\delta, B)}{n_i} - d_2(\delta, B). \end{aligned} \tag{4.6}$$

In other words, we have $\hat{\alpha}_i - \alpha_i = O_p(1)|\hat{\theta} - \theta_0| + O_p(1/\sqrt{n_i})$. If, in particular, $|\hat{\theta} - \theta_0| = O_p(1/\sqrt{m})$, then $\hat{\alpha}_i - \alpha_i = O_p(1/\sqrt{m}) + O_p(1/\sqrt{n_i})$.

The proofs of Lemmas 4.1 and 4.3 are based on the following lemma which gives a uniform Laplace approximation to integrals.

LEMMA 4.4. *Suppose $g_n(\cdot)$, $h_n(\cdot)$, and $\pi_n(\cdot)$ are, respectively, three-times, two-times, and one-time continuously differentiable such that $g_n''(\cdot)$, $\pi_n(\cdot) \geq 0$; and $\lambda_n > 0$ is a sequence. For any K, L, ϵ, and $M > 0$, there are constants depending only on these numbers, $C_i = C_i(K, L, \epsilon, M)$, $i = 1, 2$, such that*

$$\int (1 + |h_n(u)|)\pi_n(u)du \leq K, \tag{4.7}$$

$$\int_{|u|>\lambda_n} (1 + |h_n(u)|)\pi_n(u)du \leq Ln^{-3/2}, \tag{4.8}$$

$$n > (6\epsilon^{-1} \log 2)^3, \tag{4.9}$$

$$\inf_{|u-\tilde{u}|\leq n^{-1/3}} \min(g_n''(u), \pi_n(u)) \geq \epsilon, \tag{4.10}$$

and

$$\sup_{|u-\tilde{u}|\leq n^{-1/3}} \max(g_n''(u), |g_n'''(u)|, |h_n(u)|, |h_n'(u)|, |h_n''(u)|, \pi_n(u), |\pi_n'(u)|) \leq M \tag{4.11}$$

imply

$$\left| \frac{\int h_n(u) \exp(-ng_n(u))\pi_n(u)du}{\int \exp(-ng_n(u))\pi_n(u)du} - h_n(\tilde{u}) \right| \leq \left[C_1 + C_2 n^{3/2} \exp\left(-\frac{\rho_n}{2} n^{1/3}\right)\right] n^{-1}, \tag{4.12}$$

where $\tilde{u}$ is the minimizer of $g_n(\cdot)$ and $\rho_n = \inf_{|u|\leq|\tilde{u}|\vee\lambda_n} g_n''(u)$, provided that the integrals on the left side of (4.12) *and $\tilde{u}$ exist.*

As an example, let us point out how the idea of a Laplace approximation is used to deal with the first term on the right side of (4.1). In order to show that this term is $O_p(|\hat{\theta} - \theta_0|)$, it suffices to show that $\partial\psi_i/\partial\theta$ is bounded in probability. Write $g_i(v) = -(1/n)\phi_i(y_{i\cdot}, \sigma v, \beta)$. Let $h = h(y_{i\cdot}, v, \theta)$ be a function. Define

$$T_i(h) = \frac{\int h \exp(-n_i g_i)\pi dv}{\int \exp(-n_i g_i)\pi dv}, \tag{4.13}$$

where $\pi(v)$ is the pdf of $N(0, 1)$. Let $\varphi_i(y_{i\cdot}, \theta) = T_i(v)$, where v represents the identity function $h(v) = v$. Then $\psi_i(y_{i\cdot}, \theta) = \sigma\varphi_i(y_{i\cdot}, \theta)$. It is easy to show that for $1 \leq k \leq p+1$,

$$\frac{\partial\varphi_i}{\partial\theta_k} = n_i \left[T_i(v) T_i\left(\frac{\partial g_i}{\partial\theta_k}\right) - T_i\left(v\frac{\partial g_i}{\partial\theta_k}\right)\right]. \tag{4.14}$$

The equation (4.14) suggests that our goal may be hopeless: as $n_i \to \infty$, how could the right side of (4.14) be bounded? However, the difference inside the square brackets has a special form. According to Laplace approximation to integrals (*e.g.*, De Bruijn (1961), §4),

$$T_i(h) = h(\tilde{v}_i) + O\left(\frac{1}{n_i}\right), \tag{4.15}$$

where $\tilde{v}_i$ is the minimizer of g_i with respect to v. Using (4.15), we see the cancellation of the leading term inside that square brackets:

$$(4.16)\quad \frac{\partial \varphi_i}{\partial \theta_k} = n_i \left[\left(\tilde{v}_i + O\left(\frac{1}{n_i}\right)\right) \left(\frac{\partial g_i}{\partial \theta_k}\bigg|_{\tilde{v}_i} + O\left(\frac{1}{n_i}\right)\right) - \left(\tilde{v}_i \frac{\partial g_i}{\partial \theta_k}\bigg|_{\tilde{v}_i} + O\left(\frac{1}{n_i}\right)\right)\right]$$
$$= (\partial g_i/\partial \theta_k)|_{\tilde{v}_i} O(1) + \tilde{v}_i O(1) - O(1) + O(n_i^{-1}),$$

which may well be bounded, and this is the "proof".

The proofs of Lemmas 4.1–4.4 are given in Section 8.

The results obtained in this section are based on the assumption that $m \to \infty$ and, when considering a particular index i, $n_i \to \infty$. On the other hand, it is easy to show that if n_i is bounded, $\hat{\alpha}_i$ will not be consistent (i.e., $\hat{\alpha}_i - \alpha_i$ does not go to 0 in probability) as $m \to \infty$. In fact, even $E(\alpha_i \mid y)$ is not consistent in such a case, provided that θ_0 is known. Therefore, when n_i is bounded, the prediction error will not go to 0 as the number of small areas increases, hence it is more important to know about the MSE of the predictor. This is the topic of our next section.

5. The MSE of the empirical best predictor

We assume, in this section, that all the n_i's are bounded. A consequence of this assumption is that N, the total sample size, and m, the number of small areas, are of the same order, i.e., $N \sim m$. We have

$$(5.1)\quad \begin{aligned} \text{MSE}(\hat{\alpha}_i) &= E(\hat{\alpha}_i - \alpha_i)^2 \\ &= E(\hat{\alpha}_i - E(\alpha_i \mid y))^2 + E(E(\alpha_i \mid y) - \alpha_i)^2. \end{aligned}$$

The second term on the right side of (5.1) has a close form expression. Namely, by (2.2),

$$(5.2)\quad E(E(\alpha_i \mid y) - \alpha_i)^2 = \sigma_0^2 - E(E(\alpha_i \mid y_i))^2,$$

and

$$(5.3)\quad E(E(\alpha_i \mid y_i))^2 = \sum_{k=0}^{n_i} \psi_i^2(k, \theta_0) p_i(k, \theta_0) \equiv b_i(\theta_0),$$

where

$$(5.4)\quad \begin{aligned} p_i(k, \theta_0) &= P(y_{i\cdot} = k) \\ &= \sum_{z \in S(n_i, k)} \exp\left(\sum_{j=1}^{n_i} z_j x_{ij}^t \beta_0 \right) E \exp(\phi_i(z_\cdot, \sigma_0 \xi, \beta_0)) \end{aligned}$$

with $S(n,k) = \{z = (z_1, \ldots, z_n) \in \{0,1\}^n, z_\cdot = z_1 + \cdots + z_n = k\}$.

For the second term on the right side of (5.1), we use Taylor series expansion:

$$(5.5)\quad \begin{aligned} \hat{\alpha}_i - E(\alpha_i \mid y) &= \psi_i(y_{i\cdot}, \hat{\theta}) - \psi_i(y_{i\cdot}, \theta_0) \\ &= \left(\frac{\partial}{\partial \theta} \psi_i(y_{i\cdot}, \theta_0)\right)^t (\hat{\theta} - \theta_0) \\ &\quad + \frac{1}{2} (\hat{\theta} - \theta_0)^t \left(\frac{\partial^2}{\partial \theta^2} \psi_i(y_{i\cdot}, \theta_0)\right) (\hat{\theta} - \theta_0) + o(|\hat{\theta} - \theta_0|^2). \end{aligned}$$

Suppose that

$$|\hat{\theta} - \theta_0| = O_p(1/\sqrt{N}), \tag{5.6}$$

where $N = \sum_{i=1}^m n_i$ is the total sample size. Note that if some of the n_i's are not bounded, the appropriate rate of convergence in (5.6) may not be $1/\sqrt{N}$ but, instead, $1/\sqrt{m}$. It follows that

$$E(\hat{\alpha}_i - E(\alpha_i \mid y))^2 = \frac{1}{N} E\left(\left(\frac{\partial}{\partial \theta}\psi_i(y_{i\cdot}, \theta_0)\right)^t \sqrt{N}(\hat{\theta} - \theta_0)\right)^2 + o\left(\frac{1}{N}\right). \tag{5.7}$$

Now suppose $\hat{\theta} = \hat{\theta}_{i-}$, an estimator based on y_{i-}, and write $\hat{\alpha}_{i-} = \psi_i(y_{i\cdot}, \hat{\theta}_{i-})$. Then, by independence of y_i and y_{i-}, we have

$$\begin{aligned} (5.8) \qquad & E\left(\left(\frac{\partial}{\partial \theta}\psi_i(y_{i\cdot}, \theta_0)\right)^t \sqrt{N}(\hat{\theta}_{i-} - \theta_0)\right)^2 \\ &= E\left(E\left(\left(\left(\frac{\partial}{\partial \theta}\psi_i(k, \theta_0)\right)^t \sqrt{N}(\hat{\theta}_{i-} - \theta_0)\right)^2 \middle| y_{i\cdot} = k\right)\Big|_{k=y_{i\cdot}}\right) \\ &= E\left(\left(\frac{\partial}{\partial \theta}\psi_i(k, \theta_0)\right)^t V_i(\theta_0)\left(\frac{\partial}{\partial \theta}\psi_i(k, \theta_0)\right)|_{k=y_{i\cdot}}\right) \\ &= E\left(\left(\frac{\partial}{\partial \theta}\psi_i(y_{i\cdot}, \theta_0)\right)^t V_i(\theta_0)\left(\frac{\partial}{\partial \theta}\psi_i(y_{i\cdot}, \theta_0)\right)\right) \\ &= \sum_{k=0}^{n_i}\left(\frac{\partial}{\partial \theta}\psi_i(k, \theta_0)\right)^t V_i(\theta_0)\left(\frac{\partial}{\partial \theta}\psi_i(k, \theta_0)\right)p_i(k, \theta_0) \equiv a_i(\theta_0), \end{aligned}$$

where $V_i(\theta_0) = NE(\hat{\theta}_{i-} - \theta_0)(\hat{\theta}_{i-} - \theta_0)^t$.

Combining (5.1)-(5.3), (5.7) and (5.8), we obtain

$$\text{MSE}(\hat{\alpha}_{i-}) = \sigma_0^2 - b_i(\theta_0) + (1/N)a_i(\theta_0) + o(1/N). \tag{5.9}$$

The above derivation is based on the assumption that $\hat{\theta} = \hat{\theta}_{i-}$. In practice, it might be more convenient to use an estimator, $\hat{\theta}$, which is based on the entire data set, as it will not vary with the small area. A question of interest is: how much does the MSE change? or, to what extend is the formula obtained still valid?

To answer the question, we note that

$$\begin{aligned} (5.10) \qquad \text{MSE}(\hat{\alpha}_i) &= \text{MSE}(\hat{\alpha}_{i-}) + 2E(\hat{\alpha}_i - \hat{\alpha}_{i-})(\hat{\alpha}_{i-} - \alpha_i) + E(\hat{\alpha}_i - \hat{\alpha}_{i-})^2 \\ &= \text{MSE}(\hat{\alpha}_{i-}) + r_i; . \end{aligned}$$

Now we assume, in addition to (5.6), that

$$|\hat{\theta} - \hat{\theta}_{i-}| = o_p(1/\sqrt{N}). \tag{5.11}$$

To see why (5.11) is a reasonable assumption, consider the following simple example in which one estimates the population mean, μ_0, by the sample mean, $\hat{\mu}_m = (1/m)(X_1 + \cdots + X_m)$, where $X_1, \ldots, X_m$ are i.i.d. sample from the population. Then we have, for example, $\hat{\mu}_m - \hat{\mu}_{m-1} = (1/m)(X_m - \hat{\mu}_{m-1}) = O_p(1/m)$, while $\hat{\mu}_m - \mu_0 = O_p(1/\sqrt{m})$. It follows from (5.6) and (5.11) that $r_i = o(1/N)$. This follows by using, again, Taylor expansion and (2.2). Notice that $E(\hat{\alpha}_i - \hat{\alpha}_{i-})(\hat{\alpha}_{i-} - \alpha_i) = E(\hat{\alpha}_i - \hat{\alpha}_{i-})(\hat{\alpha}_{i-} - E(\alpha_i \mid y))$. Therefore, we have, by (5.10) and (5.9),

$$\text{(5.12)} \qquad \begin{aligned} \text{MSE}(\hat{\alpha}_i) &= \text{MSE}(\hat{\alpha}_{i-}) + o(1/N) \\ &= \sigma_0^2 - b_i(\theta_0) + (1/N)c_i(\theta_0) + o(1/N), \end{aligned}$$

where $c_i(\theta_0)$ is $a_i(\theta_0)$ with $V_i(\theta_0)$ replaced by $V(\theta_0) = NE(\hat{\theta} - \theta_0)(\hat{\theta} - \theta_0)^t$.

The above results are not obtained on a rigorous basis. For example, $O(1)$ ($o(1)$) in probability does not necessarily imply $O(1)$ ($o(1)$) in L^1. In the following, we shall give conditions under which the above results can be rigorously established.

We continue to make assumption (4.2), in this section. Also, recall that we assumed at the beginning of this section that the n_i's are bounded. More explicitly, let N_0 be an integer such that

$$\text{(5.13)} \qquad \max_{1 \le i \le m} n_i \le N_0.$$

Let $v = (v_j)$ be a vector and A be a matrix. Define $\|v\| = \max_j |v_j|$ and $\|A\| = (\lambda_{\max}(A^tA))^{1/2}$, where $\lambda_{\max}$ denotes the largest eigenvalue. Also, recall that $|v| = (\sum_j v_j^2)^{1/2}$. Let $0 < \lambda < 1$ and $M_0 > 0$. Define $L_N = M_0(\log N)^\lambda$. Note that L_N depends on λ. Let $\tilde{\theta}$ be an estimator. We define the *truncated estimator*, $\hat{\theta}$, as follows: $\hat{\beta}_k = \tilde{\beta}_k$, $-L_N$, or L_N if $|\tilde{\beta}_k| \le L_N$, $\tilde{\beta}_k < -L_N$, or $\tilde{\beta}_k > L_N$, respectively, $1 \le k \le p$; and $\hat{\sigma} = \tilde{\sigma}$ or $L_N^{1/2}$ if $\tilde{\sigma} \le L_N^{1/2}$ or $\tilde{\sigma} > L_N^{1/2}$, respectively. Note that it is naturally required that an estimator belongs to the parameter space, therefore $\tilde{\sigma} \ge 0$. It is clear that such a truncation will not affect the asymptotic behavior, e.g., consistency and asymptotic efficiency, of the estimator.

THEOREM 5.1. *Let $\Theta_N = \{\theta \in \Theta : \|\beta\| \le L_N, \sigma^2 \le L_N\}$. Suppose*

$$\text{(5.14)} \qquad P(\tilde{\theta} \notin \Theta_N) \vee P(|\tilde{\theta} - \theta_0| > N^{\delta - 1/2}) = O(N^{-1-\epsilon}),$$

where $0 < \delta < 1/6$ and $\epsilon > 0$. Let $\hat{\theta}$ be the truncated estimator. Then, (5.7) *and hence* (5.9) *hold.*

THEOREM 5.2. *Suppose $\tilde{\theta}$ satisfies* (5.14), *where $\epsilon > 2\delta$, and $\tilde{\theta}_{i-}$ be an estimator based on y_{i-} such that*

$$\text{(5.15)} \qquad P(\tilde{\theta}_{i-} \notin \Theta_{N,\lambda}) \vee P(|\tilde{\theta} - \tilde{\theta}_{i-}| > N^{-1/2-\rho}) = O(N^{-1-\epsilon}),$$

where $\rho > \delta$. Let $\hat{\theta}$ and $\hat{\theta}_{i-}$ be the corresponding truncated estimators. Then, (5.12) *holds.*

In the following, we shall consider a special class of estimators for which (5.14) and (5.15) are satisfied. These are the method of moments (MM) estimators. According to Jiang (1998), the MM estimator of θ_0 for model (2.1) is the solution to the following system of equations:

$$\text{(5.16)} \qquad \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ijk} y_{ij} = \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ijk} E_\theta(y_{ij}) \qquad 1 \le k \le p,$$

$$\sum_{i=1}^{m}\sum_{j\neq l} y_{ij}y_{il} = \sum_{i=1}^{m}\sum_{j\neq l} E_\theta(y_{ij}y_{il}). \tag{5.17}$$

We have $E_\theta(y_{ij}) = Eh(x_{ij}^t\beta + \sigma\xi)$, $E_\theta(y_{ij}y_{il}) = Eh(x_{ij}^t\beta + \sigma\xi)h(x_{il}^t\beta + \sigma\xi)$, $j \neq l$, where $h(x) = e^x/(1+e^x)$ and $\xi \sim N(0,1)$. Let $\hat{M}_N$ and $M_N(\theta)$ be the vectors whose components are the left and right sides of (5.16) and (5.17) divided by N. Then, these equations can be expressed by:

$$M_N(\theta) = \hat{M}_N. \tag{5.18}$$

More generally, we define the MM estimator $\tilde{\theta}$ as any vector θ that minimizes the distance between the two sides of (5.18), if such a minimizer exists; otherwise, we define $\tilde{\theta} = (0,\dots,0,1)^t$. Note that as long as such a minimizer exists, it can be chosen such that it belongs to Θ. This is because $M_N(\cdot)$ has the property that $M_N((\beta^t,-\sigma)^t) = M_N((\beta^t,\sigma)^t)$.

The following lemma states that with probability tending to one $\tilde{\theta}$ satisfies (5.18) approximately.

LEMMA 5.1. *For any $\delta > 0$ and integer $q \geq 1$, there is a constant C which may depend on δ and q such that*

$$P(|M_N(\tilde{\theta}) - \hat{M}_N| > N^{\delta-1/2}) \leq CN^{-q\delta}. \tag{5.19}$$

The following lemma shows that with probability tending to one, $\tilde{\theta}$ falls within a compact set.

LEMMA 5.2. *Suppose that there exist $B > 0$ and $\epsilon > 0$ such that for large N,*

$$\inf_{\theta\notin\Theta_B} |M_N(\theta) - M_N(\theta_0)| \geq \epsilon, \tag{5.20}$$

where $\Theta_B = \{\theta : \|\beta\| \leq B, 0 \leq \sigma \leq B\}$. Then for any $0 < \delta < 1/2$ and integer $q \geq 1$, there is a constant C which may depend on δ and q such that

$$P(\tilde{\theta} \notin \Theta_B) \leq CN^{-q\delta}. \tag{5.21}$$

We now consider the convergence rate of $\tilde{\theta} - \theta_0$.

LEMMA 5.3. *Suppose the conditions of Lemma 5.2 are satisfied. Furthermore, suppose that there exists $\epsilon_1 > 0$ such that for large N,*

$$\inf_{\theta\in\Theta_B,\theta\neq\theta_0} \left\{\frac{|M_N(\theta) - M_N(\theta_0)|}{|\theta - \theta_0|}\right\} \geq \epsilon_1. \tag{5.22}$$

Then for any $0 < \delta_1 < \delta < 1/2$ and integer $q \geq 1$,

$$P(|\tilde{\theta} - \theta_0| > N^{\delta-1/2}) = O(N^{-q\delta_1}). \tag{5.23}$$

We now consider a MM estimator which is based on y_{i-}. Such an estimator is the solution to the equations (5.16) and (5.17) with the index i replaced by i' and the summation $\sum_{i=1}^{m}$ replaced by $\sum_{i'\neq i}$. Let $\hat{M}_{N,i-}$ and $M_{N,i-}(\theta)$ be the vectors whose components are the left and right sides of these equations divided by $N - n_i$. Then the MM estimator based on y_{i-}, $\tilde{\theta}_{i-}$ is the solution to

$$M_{N,i-}(\theta) = \hat{M}_{N,i-}, \tag{5.24}$$

or more generally, a vector θ that minimizes the distance between the two sides of (5.24). As pointed out earlier, such a minimizer, if exists, can be chosen such that it belongs to Θ.

Let $M_N(\Theta)$ ($M_{N,i-}(\Theta)$) be the image of Θ under $M_N(\cdot)$ ($M_{N,i-}(\cdot)$). For $z \in R^{p+1}$ and $S \subset R^{p+1}$, let $d(z, S) = \inf_{w\in S} |z - w|$.

LEMMA 5.4. *Suppose that, in addition to the conditions of Lemma* 5.3, (5.20) *and* (5.22) *are satisfied with* $M_N(\cdot)$ *replaced by* $M_{N,i-}(\cdot)$. *Furthermore, suppose*

$$\liminf \lambda_{\min}\left(\left(\frac{\partial M_N(\theta_0)}{\partial\theta}\right)^t\left(\frac{\partial M_N(\theta_0)}{\partial\theta}\right)\right) > 0, \tag{5.25}$$

and there is $\epsilon_2 > 0$ *such that for large* N,

$$d(M_N(\theta_0), M_N^c(\Theta)) \wedge d(M_{N,i-}(\theta_0), M_{N,i-}^c(\Theta)) \geq \epsilon_2. \tag{5.26}$$

Then for any $0 < \delta_1 < 1/4$, $0 < \rho < (1 - 4\delta_1)/2$, *and integer* $q \geq 1$,

$$P(|\tilde{\theta} - \tilde{\theta}_{i-}| > N^{-1/2-\rho}) = O(N^{-q\delta_1}). \tag{5.27}$$

Note. (5.26) states that, asymptotically, $M_N(\theta_0)$ ($M_{N,i-}(\theta_0)$) is an interior point of $M_N(\Theta)$ ($M_{N,i-}(\Theta)$), which is often true when θ_0 itself is an interior point of Θ, as in the following example.

Example 5.1. Consider model (3.4). It is easy to see that $M_N(\theta) = M_{N,i-}(\theta) = (M_1(\theta), (n-1)M_2(\theta))^t$, where $M_r(\theta) = Eh_\theta^r(\xi)$ with $h_\theta(x) = \exp(\mu + \sigma x)/(1 + \exp(\mu + \sigma x))$ and $\xi \sim N(0, 1)$. Let $M(\theta) = (M_1(\theta), M_2(\theta))^t$. Suppose $m \to \infty$, while $n > 1$ is fixed. Then, obviously, (4.2) and (5.13) are satisfied.

It is easy to show that $\sup_{\mu\in R}(M_1(\theta) - M_2(\theta)) \to 0$, as $\sigma \to \infty$. Thus, there is $B_2 > 0$ such that

$$\sup_{\sigma>B_2}\sup_{\mu\in R}(M_1(\theta) - M_2(\theta)) < \epsilon, \tag{5.28}$$

where $\epsilon = \min((1/3)(M_1(\theta_0) - M_2(\theta_0)), (1/2)(1 - M_1(\theta_0)), (1/2)M_1(\theta_0)) > 0$. Also, we have $\sup_{0\leq\sigma\leq B_2}(1 - M_1(\theta)) \to 0$ as $\mu \to \infty$; and $\sup_{0\leq\sigma\leq B_2} M_1(\theta) \to 0$ as $\mu \to -\infty$. Thus, there is $B_1 > 0$ such that

$$\left[\sup_{\mu>B_1}\sup_{0\leq\sigma\leq B_2}(1 - M_1(\theta))\right] \vee \left[\sup_{\mu<-B_1}\sup_{0\leq\sigma\leq B_2} M_1(\theta)\right] < \epsilon. \tag{5.29}$$

Let $B = B_1 \vee B_2$. By (5.28) and (5.29), it is easy to see that

$$\inf_{\theta\notin\Theta_B} |M(\theta) - M(\theta_0)| \geq \epsilon,$$

from which (5.20) easily follows.

Also, it is easy to show, e.g., by the inequality in Jiang (1998, below (3.9) therein), that $|\partial M/\partial\theta| > 0$ ($|A|$ denotes the determinant of matrix A), provided that $\sigma > 0$. Thus, (5.25) is satisfied provided $\sigma_0 > 0$. Furthermore, it is easy to show, by Taylor expansion, that there is $\delta_2 > 0$ and $\epsilon_3 > 0$ such that

$$(5.30) \qquad |M(\theta) - M(\theta_0)| \geq \epsilon_3|\theta - \theta_0|, \qquad |\theta - \theta_0| < \delta_2, \quad \sigma \geq 0.$$

On the other hand, it can be shown that $M(\cdot)$ is injective (Jiang (1998)). Therefore for any $D > \delta_2 \vee |\theta_0|$, the continuous function $g(\theta) = |M(\theta) - M(\theta_0)|/|\theta - \theta_0|$, $\delta_2 \leq |\theta - \theta_0| \leq D$, $\sigma \geq 0$ has a lower bound $\epsilon_4 > 0$. Let $\epsilon_1 = \epsilon_3 \wedge \epsilon_4$. We have, by combining with (5.30), that

$$\inf_{\theta\in\Theta_B, \theta\neq\theta_0} \left\{\frac{|M(\theta) - M(\theta_0)|}{|\theta - \theta_0|}\right\} \geq \epsilon_1,$$

where $B = (D - |\theta_0|)/\sqrt{2}$. (5.22) then easily follows.

Finally, we show that

$$(5.31) \qquad M(\Theta) = \{(u, v) : 0 < u, v < 1, u^2 \leq v < u\} \equiv S.$$

It is easy to see that $M(\Theta) \subset S$. Let $(u, v) \in S$ and $u^2 < v$. Then, for any $\sigma > 0$, there is an unique $\mu = \mu(\sigma)$ such that $M_1((\mu(\sigma), \sigma)^t) = u$. The function $\mu(\cdot)$ is continuous. By an earlier result, $l(\sigma) = M_1((\mu(\sigma), \sigma)^t) - M_2((\mu(\sigma), \sigma)^t) \to 0$ as $\sigma \to \infty$. Therefore, there is $\sigma_2 > 0$ such that $l(\sigma_2) < u - v$, i.e., $M_2((\mu(\sigma_2), \sigma_2)^t) > v$. Also, it is easy to show that $\sup_{\mu\in R}(M_2(\theta) - M_1^2(\theta)) \to 0$ as $\sigma \to 0$. Therefore, there is $\sigma_1 > 0$ such that $M_2((\mu(\sigma_1), \sigma_1)^t) - M_1^2((\mu(\sigma_1), \sigma_1)^t) < v - u^2$, i.e., $M_2((\mu(\sigma_1), \sigma_1)^t) < v$. Thus, by continuity, there is $\sigma > 0$ such that $M_2((\mu(\sigma), \sigma)^t) = v$, i.e., $M((\mu(\sigma), \sigma)^t) = (u, v)^t$. If $(u, v) \in S$ and $u^2 = v$, let μ be such that $e^\mu/(1 + e^\mu) = u$, and $v = 0$. We have $M((\mu, 0)^t) = (u, v)^t$. Therefore, $S \subset M(\Theta)$.

It is now obvious that (5.26) is satisfied, provided that $\sigma_0 > 0$.

We now consider the estimation of the MSE of the EBP. Write $d_i(\theta) = \sigma^2 - b_i(\theta)$. By (5.12), we have

$$(5.32) \qquad \mathrm{MSE}(\hat{\alpha}_i) = d_i(\theta_0) + (1/N)c_i(\theta_0) + o(1/N).$$

To estimate the second term on the right side of (5.32), one may simply replace θ_0 by $\hat{\theta}$, because the difference would be $o(N^{-1})$. However, one cannot do so for the first term, because of the bias $E(d_i(\hat{\theta})) - d_i(\theta_0)$. In our case, the bias term mainly depends on the choice of the estimator $\hat{\theta}$, and it may be of the order N^{-1}, which cannot be ignored. The following theorem gives an evaluation of the bias term when $\hat{\theta}$ is the MM estimator, and hence proposes an appropriate estimator of the MSE.

THEOREM 5.3. *Suppose*(5.20), (5.22), (5.25), *and the first half of* (5.26) (*i.e.*, $d(M_N(\theta_0), M_N^c(\Theta)) \geq \epsilon_2$) *are satisfied. Let* $\hat{\theta}$ *be the truncated MM estimator. Then*

$$(5.33) \qquad E(d_i(\hat{\theta})) = d_i(\theta_0) + (1/N)B_i(\theta_0) + o(1/N),$$

where

$$(5.34) \qquad B_i(\theta_0) = \frac{1}{2}\left(E(\hat{r}_N) - \left(\frac{\partial}{\partial\theta}d_i(\theta_0)\right)^t \left(\frac{\partial}{\partial\theta}M_N(\theta_0)\right)^{-1} E(\hat{q}_N)\right),$$

$$(5.35)\qquad \hat{r}_N = \Delta_N^t \left(\left(\frac{\partial}{\partial\theta} M_N(\theta_0) \right)^{-1} \right)^t \left(\frac{\partial^2}{\partial\theta^2} d_i(\theta_0) \right) \left(\left(\frac{\partial}{\partial\theta} M_N(\theta_0) \right)^{-1} \right) \Delta_N,$$

$\hat{q}_N = (\hat{q}_{N,k})_{1\le k\le p+1}$ with $\hat{q}_{N,k}$ being $\hat{r}_N$ with $(\partial^2/\partial\theta^2)d_i(\theta_0)$ replaced by $(\partial^2/\partial\theta^2)M_{N,k}(\theta_0)$, and $\Delta_N = \sqrt{N}(\hat{M}_N - M_N(\theta_0))$. Furthermore, if we define

$$(5.36)\qquad \widehat{\mathrm{MSE}}(\hat{\alpha}_i) = d_i(\hat{\theta}) + (1/N)(c_i(\hat{\theta}) - B_i(\hat{\theta})),$$

then $E(\widehat{\mathrm{MSE}}(\hat{\alpha}_i) - \mathrm{MSE}(\hat{\alpha}_i)) = o(N^{-1})$.

The proofs of the theorems and lemmas in this section are given in Section 8.

6. Functions of fixed and random effects

In many cases, the problem of interest is a function of the fixed and random effects, say, $h_i = h_i(\beta_0, \alpha_i)$.

Example 6.1. (Linear function) $h_i = x_i^t\beta_0 + \alpha_i$, where x_i may not be the observed vector of covariates.

Example 6.2. (Probability) Consider model (2.1) under the specification (3.3). Let

$$h_i = P(y_{ij} = 1 \mid \alpha) = \frac{\exp(x_i^t\beta_0 + \alpha_i)}{1 + \exp(x_i^t\beta_0 + \alpha_i)}.$$

Example 6.3. (Weighted probability) Suppose the covariates are categorical such that $x_{ij} \in \{v_1, v_2, \dots, v_K\}$. Let

$$h_i = \sum_{k=1}^K w_k \frac{\exp(v_k^t\beta_0 + \alpha_i)}{1 + \exp(v_k^t\beta_0 + \alpha_i)},$$

where $w_1, \dots, w_K$ are a set of weights.

The best predictor for h_i is

$$(6.1)\qquad \begin{aligned} E(h_i|y) &= E(h_i|y_i) \\ &= \frac{Eh_i(\beta_0, \sigma_0\xi)\exp(\phi_i(y_{i\cdot}, \sigma_0\xi, \beta_0))}{E\exp(\phi_i(y_{i\cdot}, \sigma_0\xi, \beta_0))} = \tilde{\psi}_i(y_{i\cdot}, \theta_0). \end{aligned}$$

The development in Section 5 extends almost parallelly to the current case. The empirical best predictor for h_i is

$$(6.2)\qquad \hat{h}_i = \tilde{\psi}_i(y_{i\cdot}, \hat{\theta}).$$

We have $\mathrm{MSE}(\hat{h}_i) = E(\hat{h}_i - E(h_i \mid y))^2 + E(E(h_i \mid y) - h_i)^2$ with

$$(6.3)\qquad \begin{aligned} E(E(h_i \mid y) - h_i)^2 &= Eh_i^2 - E(E(h_i \mid y_i))^2 \\ &= Eh_i^2(\beta_0, \sigma_0\xi) - \sum_{k=0}^{n_i} \tilde{\psi}_i^2(k, \theta_0)p_i(k, \theta_0) \equiv \tilde{d}_i(\theta_0), \end{aligned}$$

and, if $\hat{\theta} = \hat{\theta}_{i-}$, an estimator of θ based on y_{i-},

$$E(\hat{h}_i - E(h_i \mid y))^2 = (1/N)\tilde{a}_i(\theta_0) + o(1/N), \tag{6.4}$$

where $\tilde{a}_i(\theta_0)$ is $a_i(\theta_0)$ with ψ_i replaced by $\tilde{\psi}_i$. Thus, with $\hat{h}_{i-} = \tilde{\psi}_i(y_{i\cdot}, \hat{\theta}_{i-})$,

$$\mathrm{MSE}(\hat{h}_{i-}) = \tilde{d}_i(\theta_0) + (1/N)\tilde{a}_i(\theta_0) + o(1/N). \tag{6.5}$$

Also, one may replace $\hat{\theta}_{i-}$ by $\hat{\theta}$, an estimator of θ based on all data, and may still obtain

$$\begin{aligned} \mathrm{MSE}(\hat{h}_i) &= \mathrm{MSE}(\hat{h}_{i-}) + o(1/N) \\ &= \tilde{d}_i(\theta_0) + (1/N)\tilde{c}_i(\theta_0) + o(1/N), \end{aligned} \tag{6.6}$$

where $\tilde{c}_i(\theta_0)$ is $c_i(\theta_0)$ with ψ_i replaced by $\tilde{\psi}_i$. Finally, if $\hat{\theta}$ is the truncated MM estimator, an estimator of the MSE with bias correction is given by

$$\widehat{\mathrm{MSE}}(\hat{h}_i) = \tilde{d}_i(\hat{\theta}) + (1/N)(\tilde{c}_i(\hat{\theta}) - \tilde{B}_i(\hat{\theta})), \tag{6.7}$$

where $\tilde{B}_i(\hat{\theta})$ is $B_i(\hat{\theta})$ with d_i replaced by $\tilde{d}_i$; and we have $E(\widehat{\mathrm{MSE}}(\hat{h}_i) - \mathrm{MSE}(\hat{h}_i)) = o(N^{-1})$.

Of course, under suitable conditions, all the above results can be established on a rigorous basis.

Example 6.1. (Continued) In this case, the EBP for h_i is $\hat{h}_i = x_i^t\hat{\beta} + \hat{\alpha}_i$, where $\hat{\alpha}_i$ is the EBP for α_i.

Example 6.2. (Continued) It is easy to show that, in this case, the EBP for h_i is

$$\hat{h}_i = \exp(x_i^t\hat{\beta})\frac{E\exp((y_{i\cdot}+1)\hat{\sigma}\xi - (n_i+1)\log(1+\exp(x_i^t\hat{\beta}+\hat{\sigma}\xi)))}{E\exp(y_{i\cdot}\hat{\sigma}\xi - n_i\log(1+\exp(x_i^t\hat{\beta}+\hat{\sigma}\xi)))}, \tag{6.8}$$

where the expectations are taken with respect to $\xi \sim N(0,1)$. Note that the EBP is not $\exp(x_i^t\hat{\beta}+\hat{\alpha}_i)/(1+\exp(x_i^t\hat{\beta}+\hat{\alpha}_i))$, although, according to Lemma 4.4, the two will be very close when n_i is large. A naive predictor of the (conditional) probability would be $\bar{y}_{i\cdot} = y_{i\cdot}/n_i$. Although the EBP given by (6.8) is not difficult to compute (e.g., by Monte Carlo method), it does not have a closed form. So, the question is: just how much better is the EBP than the naive predictor? To answer this question, we consider the relative savings loss (RSL) introduced by Efron and Morris (1973). In the current case, the RSL is given by

$$\mathrm{RSL} = \frac{\mathrm{MSE}(\hat{h}_i) - \mathrm{MSE}(E(h_i \mid y))}{\mathrm{MSE}(\bar{y}_{i\cdot}) - \mathrm{MSE}(E(h_i \mid y))} = \frac{E(\hat{h}_i - E(h_i \mid y))^2}{E(\bar{y}_{i\cdot} - E(h_i \mid y))^2}. \tag{6.9}$$

According to (6.4), the numerator on the right side of (6.9) is of the order $1/N$, while

$$\text{the denominator} = \sum_{k=0}^{n_i}\left(\frac{k}{n_i} - \tilde{\psi}_i(k,\theta_0)\right)^2 p_i(k,\theta_0) \geq (\tilde{\psi}_i(0,\theta_0))^2 p_i(0,\theta_0). \tag{6.10}$$

If n_i is bounded, the right side of (6.10) has a positive lower bound. Therefore the RSL $\to 0$ as $N \to \infty$. In fact, the convergence rate is $O(1/N)$. So, the complication of EBP is worthwhile.

7. Discussion

We have studied the asymptotic properties of the EBP and the problem of estimating the MSE of the EBP when there is not sufficient information about individual random effects. Several extensions of this work seem possible. First, the random effects do not have to be univariate. For example, one may assume that $\alpha_1, \dots, \alpha_m$ are independent and distributed as $N(0, D)$, where the covariance matrix D may depend on some parameters of variance components. Consistent estimators of the fixed effects and variance components in such a model have been obtained by Jiang (1998). It is fairly straightforward to extend most of the results of this paper to such a case. Second, the normality assumption about the random effects is not essential. However, for the EBP to be computable, the distribution of the random effects has to be specified up to a vector of unknown parameters. On the other hand, the normality assumption seems (close to be) realistic in many cases, and it brings computational convenience, especially when the random effects are multivariate. Finally, it is possible to apply methods developed in this paper to other generalized linear mixed models such as models for counts and survival data, which are also encountered in practice. Research is on the way to extend the results obtained for the mixed logistic model to mixed models with general link functions in a complex survey setting.

8. Proofs

Proof of Lemma 3.1. The first equality follows immediately from the substitution $u = x/(1+x)$, because the integral then reduces to $B(k, n-k)$. To prove the second equality, one observes that, with the same transformation $u = x/(1+x)$,

$$\tag{8.1} \begin{aligned} &\int_0^\infty (\log x) x^{k-1}(1+x)^{-n} dx \\ &\quad = \int_0^1 \log(u) u^{k-1}(1-u)^{n-k-1} du - \int_0^1 \log(1-u) u^{k-1}(1-u)^{n-k-1} du. \end{aligned}$$

Next, by Fubini's theorem and the relationship between beta and binomial distributions, it follows that

$$\begin{aligned} &\int_0^1 [\log(u) u^{k-1}(1-u)^{n-k-1}/B(k, n-k)] du \\ &= \int_0^1 \int_u^1 (-dx/x)[u^{k-1}(1-u)^{n-k-1}/B(k, n-k)] du \\ &= \int_0^1 (-dx/x) \int_0^x [u^{k-1}(1-u)^{n-k-1}/B(k, n-k)] du \\ &= -\int_0^1 x^{-1} P(\mathrm{Beta}(k, n-k) \le x) dx \\ &= -\int_0^1 x^{-1} \left[-\sum_{l=1}^{k-1} \frac{(n-1)!}{l!(n-1-l)!} x^l (1-x)^{n-1-l} + \int_0^x (n-1)(1-u)^{n-2} du \right] dx \\ &= \int_0^1 \sum_{l=1}^{k-1} \frac{(n-1)!}{l!(n-1-l)!} x^{l-1}(1-x)^{n-1-l} dx + \int_0^1 \log(u)(n-1)(1-u)^{n-2} du \end{aligned}$$

$$= \sum_{l=1}^{k-1} l^{-1} + \int_0^1 \log(u)(n-1)(1-u)^{n-2} du.$$

Similarly,

$$\int_0^1 [\log(1-u)u^k(1-u)^{n-k-1}/B(k,n-k)]du$$
$$= \int_0^1 [\log(u)u^{n-k-1}(1-u)^{k-1}/B(n-k,k)]du$$
$$= \sum_{l=1}^{n-k-1} l^{-1} + \int_0^1 \log(u)(n-1)(1-u)^{n-2} du.$$

The result thus follows. □

PROOF OF (3.5). As pointed out below (3.4), it suffices to consider (3.4). By (3.1) and a change of variable, it is easy to obtain the result. □

PROOF OF LEMMA 4.4. Let $a_n(u) = g_n(u) - g_n(\tilde{u})$ and $b_n(u) = h_n(u) - h_n(\tilde{u})$. Then

$$\frac{\int h_n(u)\exp(-ng_n(u))\pi_n(u)du}{\int \exp(-ng_n(u))\pi_n(u)du} = h_n(\tilde{u}) + \frac{\int b_n(u)\exp(-na_n(u))\pi_n(u)du}{\int \exp(-na_n(u))\pi_n(u)du}. \tag{8.2}$$

We have

$$\int b_n(u)\exp(-na_n(u))\pi_n(u)du = \int_{|u-\tilde{u}|\le\delta_n} \cdots du + \int_{|u-\tilde{u}|>\delta_n, |u|\le\lambda_n} \cdots du + \int_{|u-\tilde{u}|>\delta_n, |u|>\lambda_n} \cdots du = I_1 + I_2 + I_3, \tag{8.3}$$

where $\delta_n = n^{-1/3}$.

It follows from (4.8) that

$$|I_3| \le L(1 \vee |h_n(\tilde{u})|)n^{-3/2}. \tag{8.4}$$

Suppose $|u-\tilde{u}| > \delta_n$ and $|u| \le \lambda_n$. Then

$$a_n(u) = \frac{1}{2}g_n''(u_1)(u-\tilde{u})^2, \tag{8.5}$$

where u_1 lies between $\tilde{u}$ and u. It follows that

$$|I_2| \le K(1 \vee |h_n(\tilde{u})|)\exp\left(-\frac{\rho_n}{2}n^{1/3}\right). \tag{8.6}$$

Now suppose $|u-\tilde{u}| \le \delta_n$. We have

$$b_n(u) = h_n'(\tilde{u})(u-\tilde{u}) + \frac{1}{2}h_n''(u_2)(u-\tilde{u})^2, \tag{8.7}$$

$$a_n(u) = \frac{1}{2}g_n''(\tilde{u})(u-\tilde{u})^2 + \frac{1}{6}g_n'''(u_3)(u-\tilde{u})^3, \tag{8.8}$$

$$\pi_n(u) = \pi_n(\tilde{u}) + \pi_n'(u_4)(u-\tilde{u}), \tag{8.9}$$

where u_2, u_3, and u_4 lie between $\tilde{u}$ and u. (8.8) implies that

$$\exp(-na_n(u)) = \exp\left(-\frac{n}{2}g_n''(\tilde{u})(u-\tilde{u})^2\right)[1-\exp(-t\gamma(u-\tilde{u})^3)\gamma(u-\tilde{u})^3], \tag{8.10}$$

where $\gamma = (n/6)g_n'''(u_3)$ and $0 \le t \le 1$. By (8.7), we have

$$\begin{aligned} I_1 &= h_n'(\tilde{u})\int_{|u-\tilde{u}|\le\delta_n}(u-\tilde{u})\exp(-na_n(u))\pi_n(u)du \\ &\quad+\frac{1}{2}\int_{|u-\tilde{u}|\le\delta_n}h_n''(u_2)(u-\tilde{u})^2\exp(-na_n(u))\pi_n(u)du \\ &= I_{11}+I_{12}. \end{aligned} \tag{8.11}$$

It follows from (8.5) that

$$|I_{12}| \le \sqrt{\frac{\pi}{2}}\epsilon_1^{-3/2}M_1M_2n^{-3/2}, \tag{8.12}$$

where $\epsilon_1 = \inf_{|u-\tilde{u}|\le\delta_n} g_n''(u)$, $M_1 = \sup_{|u-\tilde{u}|\le\delta_n}|h_n''(u)|$, and $M_2 = \sup_{|u-\tilde{u}|\le\delta_n}\pi_n(u)$. Since

$$\int_{|u-\tilde{u}|\le\delta_n}(u-\tilde{u})\exp\left(-\frac{n}{2}g_n''(\tilde{u})(u-\tilde{u})^2\right)du = 0,$$

we have, by (8.9) and (8.10),

$$\begin{aligned} &\int_{|u-\tilde{u}|\le\delta_n}(u-\tilde{u})\exp(-na_n(u))\pi_n(u)du \\ &\quad= -\frac{n}{6}\pi_n(\tilde{u})\int_{|u-\tilde{u}|\le\delta_n}g_n'''(u_3)\exp(-t\gamma(u-\tilde{u})^3)(u-\tilde{u})^4 \\ &\quad\quad\cdot\exp\left(-\frac{n}{2}g_n''(\tilde{u})(u-\tilde{u})^2\right)du+\int_{|u-\tilde{u}|\le\delta_n}\pi_n'(u_4)(u-\tilde{u})^2\exp(-na_n(u))du. \end{aligned}$$

Thus, we have, similar to (8.12), that

$$|I_{11}| \le \sqrt{\frac{\pi}{2}}|h_n'(\tilde{u})|\left(\pi_n(\tilde{u})(g_n''(\tilde{u}))^{-5/2}M_3\exp\left(\frac{M_3}{6}\right)+2\epsilon_1^{-3/2}M_4\right)n^{-3/2}, \tag{8.13}$$

where $M_3 = \sup_{|u-\tilde{u}|\le\delta_n}|g_n'''(u)|$ and $M_4 = \sup_{|u-\tilde{u}|\le\delta_n}|\pi_n'(u)|$.

On the other hand, we have, again, by (8.5),

$$\int\exp(-na_n(u))\pi_n(u)du \ge \epsilon_2\int_{|u-\tilde{u}|\le\delta_n}\exp\left(-\frac{n}{2}M_5(u-\tilde{u})^2\right)du, \tag{8.14}$$

where $\epsilon_2 = \inf_{|u-\tilde{u}|\le\delta_n}\pi_n(u)$ and $M_5 = \sup_{|u-\tilde{u}|\le\delta_n} g_n''(u)$. Also,

$$\begin{aligned} \int_{|u-\tilde{u}|\le\delta_n}\exp\left(-\frac{n}{2}M_5(u-\tilde{u})^2\right)du &= \int\cdots du-\int_{|u-\tilde{u}|>\delta_n}\cdots du \\ &\ge \sqrt{2\pi}M_5^{-1/2}\left[1-\sqrt{2}\exp\left(-\frac{M_5}{4}n^{1/3}\right)\right] \\ &\quad\cdot n^{-1/2}. \end{aligned} \tag{8.15}$$

Now suppose (4.9) and (4.10) are satisfied. Since $M_5 \geq \epsilon_1 \geq \epsilon$, we have $\sqrt{2}\exp(-(M_5/4)n^{1/3}) < 1/2$. Thus, by combining (8.2)-(8.4), (8.6), and (8.11)-(8.15), we get

$$\left|\frac{\int h_n(u)\exp(-ng_n(u))\pi_n(u)du}{\int \exp(-ng_n(u))\pi_n(u)du} - h_n(\tilde{u})\right| \leq B_1\left[B_2 + B_3 n^{3/2}\exp\left(-\frac{\rho_n}{2}n^{1/3}\right)\right]n^{-1}, \tag{8.16}$$

where $B_1 = \sqrt{2/\pi}\epsilon_2^{-1}M_5^{1/2}$,

$$B_2 = L(1 \vee |h_n(\tilde{u})|) + \sqrt{\frac{\pi}{2}}\left(\epsilon_1^{-3/2}(M_1M_2 + 2|h_n'(\tilde{u})|M_4) + |h_n'(\tilde{u})|\pi_n(\tilde{u})(g_n''(\tilde{u}))^{-5/2}M_3\exp\left(\frac{M_3}{6}\right)\right),$$

and $B_3 = K(1 \vee |h_n(\tilde{u})|)$. The result then follows. □

To prove Lemma 4.1, we need the following result.

LEMMA 8.1. *For any* $B > A|\beta_0| + \log(2e^{A|\beta_0|} + 1)$, $P(|\tilde{\alpha}_i| > B) \leq$ *the right side of* (4.3).

PROOF OF LEMMA 8.1. First, it is easy to show that as long as $y_i \neq (0, \ldots, 0) \equiv 0$ or $(1, \ldots, 1) \equiv 1$, $\tilde{\alpha}_i$ exists, which is the unique solution to the following equation.

$$\sum_{j=1}^{n_i}\frac{\exp(x_{ij}^t\beta_0 + u)}{1 + \exp(x_{ij}^t\beta_0 + u)} = y_{i\cdot}. \tag{8.17}$$

It follows that if $y_i \neq 0$ or 1, then

$$|\tilde{\alpha}_i| \leq A|\beta_0| + |\operatorname{logit}(\bar{y}_{i\cdot})|. \tag{8.18}$$

If we define $\tilde{\alpha}_i$ as $-\infty$ and ∞ when $y_i = 0$ or 1, then (8.18) holds even in those two extreme cases. It is easy to show that

$$P(\operatorname{logit}(\bar{y}_{i\cdot}) > B - A|\beta_0|) \leq P\left(\frac{1}{n_i}\sum_{i=1}^{n_i}(y_{ij} - E(y_{ij}|\alpha_i)) > \delta\right) + P(\alpha_i > \operatorname{logit}(1-2\delta) - A|\beta_0|), \tag{8.19}$$

$$P(\operatorname{logit}(\bar{y}_{i\cdot}) < A|\beta_0| - B) \leq P\left(\frac{1}{n_i}\sum_{i=1}^{n_i}(y_{ij} - E(y_{ij}|\alpha_i)) < -\delta\right) + P(\alpha_i < A|\beta_0| - \operatorname{logit}(1-2\delta)). \tag{8.20}$$

The conclusion follows easily from (8.18)-(8.20). Note that $\operatorname{logit}(1-2\delta) > A|\beta_0|$, and

$$E\left(\sum_{j=1}^{n_i}(y_{ij} - E(y_{ij}|\alpha_i))\right)^2 = \sum_{j=1}^{n_i}E(\operatorname{var}(y_{ij}|\alpha_i)) \leq \frac{n_i}{4}. \tag{8.21}$$

□

PROOF OF LEMMA 4.1. To simplify notation, write $n = n_i$, and $l_i = -(1/n_i)(\phi_i(y_{i\cdot}, u, \beta_0) + \sum_{j=1}^{n_i} y_{ij} x_{ij}^t \beta_0)$. Since $\alpha_i \sim N(0, \sigma_0^2)$, we have $\pi(u) = (\sqrt{2\pi}\sigma_0)^{-1} \exp(-u^2/2\sigma_0^2)$. It follows that

$$\int_{|u|>\lambda_n} (1 + |u|)\pi(u)du \le (\sqrt{2} + 4\sigma_0/\sqrt{2\pi})n^{-3/2},$$

where $\lambda_n = \sigma_0\sqrt{6\log n}$. Also, we have

$$\inf_{|u|\le B+1} \pi(u) = \epsilon_B = (\sqrt{2\pi}\sigma_0)^{-1} \exp(-(B+1)^2/2\sigma_0^2),$$
$$(B+1) \vee \sup_u(\pi(u) \vee |\pi'(u)|) \le M_B = (B+1) \vee (\sqrt{2\pi}\sigma_0(\sigma_0 \wedge 1))^{-1}.$$

Thus, it is easy to show that $(\partial^2 l_i/\partial u^2) \vee |\partial^3 l_i/\partial u^3| \le 1/4 < M_B$; and, by (4.2),

$$\inf_{|u|\le B+1} \frac{\partial^2 l_i}{\partial u^2} \ge \delta_B = \frac{1}{4}\exp(-2(A|\beta_0| + B + 1));$$
$$\inf_{|u|\le B\vee\lambda_n} \frac{\partial^2 l_i}{\partial u^2} \ge \frac{1}{4}\exp(-2(A|\beta_0| + B + \lambda_n)).$$

Note that $\delta_B \ge \epsilon_B$ for large B.

Suppose $|\tilde{\alpha}_i| \le B$. Then, with $\epsilon = \epsilon_B$ and $M = M_B$, the above arguments show that (4.10) and (4.11) (with g_n, $h_n(u)$, and $\pi_n(u)$ replaced by l_i, u, and $\pi(u)$, respectively) are satisfied, if B is large. Furthermore, if n is large such that (4.9) is satisfied and

$$\text{(8.22)} \qquad \frac{3\log n}{n^{1/3}} \le \frac{1}{4}\exp(-2(A|\beta_0| + B + \sigma_0\sqrt{6\log n})),$$

we have $\rho_n = \inf_{|u|\le|\tilde{\alpha}_i|\vee\lambda_n}(\partial^2 l_i/\partial u^2) \ge \inf_{|u|\le B\vee\lambda_n}(\partial^2 l_i/\partial u^2) \ge 3\log n/n^{1/3}$. Thus, by Lemma 4.4, $n|\psi_i(y_{i\cdot}, \theta_0) - \tilde{\alpha}_i| \le C_1 + C_2$, where $C_i = C_i(1 + \sigma_0\sqrt{2/\pi}, \sqrt{2} + 4\sigma_0/\sqrt{2\pi}, \epsilon_B, M_B)$, $i = 1, 2$, are functions defined in Lemma 4.4.

Therefore, $P(n|\psi_i(y_{i\cdot}, \theta_0) - \tilde{\alpha}_i| > C_1 + C_2) \le P(|\tilde{\alpha}_i| > B)$, provided that B is large and n satisfies (4.9) and (8.22). The result then follows from Lemma 8.1. □

PROOF OF LEMMA 4.2. Since $\tilde{\alpha}_i$ is the maximizer of $\phi_i(y_{i\cdot}, u, \beta_0)$, we have

$$\text{(8.23)} \qquad 0 = \frac{\partial\phi_i}{\partial u}\Big|_{\tilde{\alpha}_i} = \frac{\partial\phi_i}{\partial u}\Big|_{\alpha_i} + \frac{\partial^2\phi_i}{\partial u^2}\Big|_{\alpha_i^*}(\tilde{\alpha}_i - \alpha_i),$$

where α_i^* lies between α_i and $\tilde{\alpha}_i$. Thus, $|\alpha_i^*| \le |\alpha_i| \vee |\tilde{\alpha}_i|$. If $|\alpha_i^*| \le B - A|\beta_0|$, then, by (8.23) and (4.2),

$$\text{(8.24)} \qquad \left|\frac{\partial\phi_i}{\partial u}\Big|_{\alpha_i}\right| = |\tilde{\alpha}_i - \alpha_i| \sum_{j=1}^{n_i} \frac{\exp(x_{ij}^t\beta_0 + \alpha_i^*)}{(1 + \exp(x_{ij}^t\beta_0 + \alpha_i^*))^2} \ge \frac{1}{4}e^{-2B} n_i|\tilde{\alpha}_i - \alpha_i|.$$

Therefore,

$$\text{(8.25)} \quad P(\sqrt{n_i}|\tilde{\alpha}_i - \alpha_i| > 4Be^{2B}) \le P(\ldots, |\alpha_i^*| \le B - A|\beta_0|) + P(\ldots, |\alpha_i^*| > B - A|\beta_0|)$$
$$\le P\left(\left|\frac{\partial\phi_i}{\partial u}\Big|_{\alpha_i}\right| > B\sqrt{n_i}\right) + P(|\alpha_i| > B - A|\beta_0|) + P(|\tilde{\alpha}_i| > B - A|\beta_0|).$$

Now, all one has to do is to apply (8.21) and Lemma 8.1 (with B replaced by $B - A|\beta_0|$) to obtain bounds for the first and third terms on the right side of (8.25). □

PROOF OF LEMMA 4.3. Again, we write $n = n_i$ and $g_i = g_i(y_{i\cdot}, v, \theta) = -(1/n_i)\phi_i(y_{i\cdot}, \sigma v, \beta)$ to simplify notation. Let $\Delta_i(h) = T_i(h) - h(\tilde{v}_i)$ (see (4.13)), where $\tilde{v}_i$ is the minimizer of $g_i(y_{i\cdot}, \cdot, \theta)$, then

$$(8.26)\quad \frac{\partial \varphi_i}{\partial \theta_k} = n\left[\frac{\partial g_i}{\partial \theta_k}|_{\tilde{v}_i}\Delta_i(v) + \tilde{v}_i\Delta_i\left(\frac{\partial g_i}{\partial \theta_k}\right) + \Delta_i(v)\Delta_i\left(\frac{\partial g_i}{\partial \theta_k}\right) - \Delta_i\left(v\frac{\partial g_i}{\partial \theta_k}\right)\right].$$

As pointed out earlier, the idea of the proof is to apply Lemma 4.4. To make the latter more convenient to use in the current case, we point out the following obvious consequence.

COROLLARY 8.1. *Let δ, $B > 0$, and K_δ, L_δ, $\epsilon_{\delta,B} \le (2\pi)^{-1/2}\exp(-(B+1)^2/2)$, $M_{\delta,B} \ge (2\pi)^{-1/2}$ satisfy*

$$(8.27)\qquad \sup_{|\theta-\theta_0|\le\delta}\int(1+|h|)\pi dv \le K_\delta,$$

$$(8.28)\qquad \sup_{|\theta-\theta_0|\le\delta} n^{3/2}\int_{|v|>\lambda_n}(1+|h|)\pi dv \le L_\delta,$$

$$(8.29)\qquad \inf_{|v|\le B+1, |\theta-\theta_0|\le\delta}\frac{\partial^2 g_i}{\partial v^2} \ge \epsilon_{\delta,B},$$

and

$$(8.30)\qquad \sup_{|v|\le B+1, |\theta-\theta_0|\le\delta}\max\left(\frac{\partial^2 g_i}{\partial v^2}, \left|\frac{\partial^3 g_i}{\partial v^3}\right|, |h|, \left|\frac{\partial h}{\partial v}\right|, \left|\frac{\partial^2 h}{\partial v^2}\right|\right) \le M_{\delta,B}.$$

Then

$$(8.31)\qquad P\left(\sup_{|\theta-\theta_0|\le\delta} n|\Delta_i(h)| > C_{\delta,B}\right) \le P\left(\sup_{|\theta-\theta_0|\le\delta}|\tilde{v}_i| > B\right),$$

where $C_{\delta,B} = \sum_{i=1}^2 C_i(K_\delta, L_\delta, \epsilon_{\delta,B}, M_{\delta,B})$ (see Lemma 4.4), provided that

$$(8.32)\qquad n > \left(\frac{6\log 2}{\epsilon_{\delta,B}}\right)^3 \quad \textit{and} \quad \frac{3\log n}{n^{1/3}} \le \inf_{|v|\le B\vee\lambda_n, |\theta-\theta_0|\le\delta}\frac{\partial^2 g_i}{\partial v^2}.$$

Let $H = \{v, (\partial g_i/\partial\theta_k), v(\partial g_i/\partial\theta_k), 1 \le k \le p+1\}$. It is easy to show that the following numbers are good enough, in view of Corollary 8.1, for $h \in H$:

$K_\delta = (1 + \sqrt{2/\pi})(1 + A \vee 1)$,
$L_\delta = (\sqrt{2} + 4/\sqrt{2\pi})(1 + A \vee 1)$,
$\epsilon_{\delta,B} = [(2\pi)^{-1/2}\exp(-(B+1)^2/2)] \wedge [4^{-1}(\sigma_0 - \delta)^2\exp(-2(|\theta_0| + \delta)(A + B + 1))]$,
and $M_{\delta,B} = [2 + (\sigma_0 + \delta \vee 2)(B+1) + 4^{-1}(\sigma_0 + \delta \vee 2)^2(B+1)^2](A \vee 1)$,

where it is assumed that $\delta < \sigma_0$. Note that K_δ and L_δ, in fact, do not depend on δ. Also, (8.32) is satisfied if

$$(8.33)\quad n > \left(\frac{6\log 2}{\epsilon_{\delta,B}}\right)^3 \quad \text{and} \quad \frac{3\log n}{n^{1/3}} \le \frac{1}{4}\exp(-2(|\theta_0| + \delta)(A + B + \sqrt{6\log n})).$$

Suppose $|\theta - \theta_0| \le \delta$ such that $|\tilde{v}_i| \le B$, and $\max_{h\in H} n|\Delta_i(h)| \le C_{\delta,B}$. Then it is easy to show that

$$\left| \left.\frac{\partial g_i}{\partial \theta_k}\right|_{\tilde{v}_i} \right| \le A \vee B, \qquad 1 \le k \le p+1.$$

Thus, by (8.26), we have

$$\left|\frac{\partial \varphi_i}{\partial \theta_k}\right| \le (A \vee B + B + 1 + C_{\delta,B} n^{-1}) C_{\delta,B}, \qquad 1 \le k \le p+1.$$

Also, we have

$$\begin{aligned}\left|\frac{\partial \psi_i}{\partial \theta_k}\right| &\le |\varphi_i| + \sigma \left|\frac{\partial \varphi_i}{\partial \theta_k}\right| \\ &\le B + C_{\delta,B} n^{-1} + (\sigma_0 + \delta)\left|\frac{\partial \varphi_i}{\partial \theta_k}\right|, \qquad 1 \le k \le p+1.\end{aligned}$$

If we let $D_1(\delta, B) = B + (\sigma_0+\delta)(A \vee B + B + 1)C_{\delta,B}$, $D_2(\delta, B) = [1 + (\sigma_0+\delta)C_{\delta,B}]C_{\delta,B}$, then the above arguments and Corollary 8.1 imply

$$\begin{aligned}&P\left(\max_{1\le k\le p+1} \sup_{|\theta-\theta_0|\le\delta} \left|\frac{\partial \psi_i}{\partial \theta_k}\right| > D_1(\delta,B) + D_2(\delta,B)n^{-1}\right) \\ &\quad \le P\left(\sup_{|\theta-\theta_0|\le\delta} |\tilde{v}_i| > B\right) + \sum_{h\in H} P\left(\sup_{|\theta-\theta_0|\le\delta} n|\Delta_i(h)| > C_{\delta,B}\right) \\ &\quad \le 2(p+2)P\left(\sup_{|\theta-\theta_0|\le\delta} |\tilde{v}_i| > B\right),\end{aligned}$$

provided $\delta < \sigma_0$ and (8.33) is satisfied.

On the other hand, similar to (8.18), we have $|\tilde{v}_i| \le \sigma^{-1}(A|\beta| + |\operatorname{logit}(\bar{y}_{i\cdot})|)$. Therefore, by (8.19)-(8.21), there is $B_0 > 0$ such that for $\delta < \sigma_0$ and $B > B_0$,

$$\begin{aligned}P\left(\sup_{|\theta-\theta_0|\le\delta} |\tilde{v}_i| > B\right) &\le P(|\operatorname{logit}(\bar{y}_{i\cdot})| > B(\sigma_0 - \delta) - A(|\beta_0| + \delta)) \\ &\le (4\rho^2 n)^{-1} + 2(1 - \Phi(\sigma_0^{-1}(\operatorname{logit}(1-2\rho) - A|\beta_0|))),\end{aligned}$$

where $\rho = \exp(A(|\beta_0| + \delta))/(\exp(A(|\beta_0|+\delta)) + \exp(B(\sigma_0 - \delta)))$. □

PROOF OF THEOREM 5.1. Throughout this proof we will use C to denote a constant whose value may change from time to time. For any $a > 0$, we have $a|u| - (1/4)u^2 \le a^2$ for all u. Therefore, it is easy to show that

$$\frac{\int |u| \exp\left(a|u| - (1/2)u^2\right) du}{\int \exp\left(-a|u| - (1/2)u^2\right) du} \le 2\sqrt{\frac{3}{\pi}} \exp(2a^2). \tag{8.34}$$

On the other hand, it is easy to show that under (4.2) and (5.13),

$$|\phi_i(y_{i\cdot}, \sigma u, \beta)| \le N_0(A|\beta| + \log 2) + 2N_0\sigma|u|. \tag{8.35}$$

Therefore, by (2.2), (8.34), and (8.35), we have

$$|\psi_i(y_{i\cdot},\theta)| \le 2(3/\pi)^{1/2}4^{N_0}|\sigma|\exp(2N_0(A|\beta| + 4N_0\sigma^2)). \tag{8.36}$$

Also, (5.14) implies that

$$P(|\hat{\theta}-\theta_0| > N^{\delta-1/2}) \le P(|\tilde{\theta}-\theta_0| > N^{\delta-1/2}) + P(\tilde{\theta} \notin \Theta_{N,\lambda}) \le CN^{-1-\epsilon}. \tag{8.37}$$

Write

$$\begin{aligned} E(\hat{\alpha}_i - E(\alpha_i|y))^2 &= E(\psi_i(y_{i\cdot},\hat{\theta}) - \psi_i(y_{i\cdot},\theta_0))^2 1_{(|\hat{\theta}-\theta_0| \le N^{\delta-1/2})} \\ &\quad + E(\psi_i(y_{i\cdot},\hat{\theta}) - \psi_i(y_{i\cdot},\theta_0))^2 1_{(|\hat{\theta}-\theta_0| > N^{\delta-1/2})} = I_1 + I_2. \end{aligned} \tag{8.38}$$

If $|\hat{\theta}-\theta_0| \le N^{\delta-1/2}$, then, by Taylor expansion, we have

$$\begin{aligned} \psi_i(y_{i\cdot},\hat{\theta}) - \psi_i(y_{i\cdot},\theta_0) &= \left(\frac{\partial}{\partial\theta}\psi_i(y_{i\cdot},\theta_0)\right)^t (\hat{\theta}-\theta_0) \\ &\quad \frac{1}{2}(\hat{\theta}-\theta_0)\left(\frac{\partial^2}{\partial\theta^2}\psi_i(y_{i\cdot},\theta_*)\right)(\hat{\theta}-\theta_0), \end{aligned} \tag{8.39}$$

where θ_* lies between θ_0 and $\hat{\theta}$. Since $|\hat{\theta}-\theta_0| \le N^{\delta-1/2}$ implies $|\hat{\theta}| \le |\theta_0| + 1$, it is easy to show that

$$\left|(\hat{\theta}-\theta_0)^t\left(\frac{\partial^2}{\partial\theta^2}\psi_i(y_{i\cdot},\theta_*)\right)(\hat{\theta}-\theta_0)\right| \le CN^{2\delta-1},$$

and

$$\left|\left(\frac{\partial}{\partial\theta}\psi_i(y_{i\cdot},\theta_0)\right)^t (\hat{\theta}-\theta_0)\right| \le CN^{\delta-1/2}.$$

Thus, by (8.39), we have

$$I_1 = E\left(\left(\frac{\partial}{\partial\theta}\psi_i(y_{i\cdot},\theta_0)\right)^t (\hat{\theta}-\theta_0)\right)^2 1_{(|\hat{\theta}-\theta_0| \le N^{\delta-1/2})} + o(N^{-1}). \tag{8.40}$$

On the other hand, we have, by (8.36), the definition of $\hat{\theta}$, and (8.37),

$$I_2 \le C(\log N)^\lambda \exp(4N_0(A+4N_0)(\log N)^\lambda)P(|\hat{\theta}-\theta_0| > N^{\delta-1/2}) = o(N^{-1}). \tag{8.41}$$

Finally, by a similar argument as the above, we have

$$\begin{aligned} &E\left(\left(\frac{\partial}{\partial\theta}\psi_i(y_{i\cdot},\theta_0)\right)^t (\hat{\theta}-\theta_0)\right)^2 1_{(|\hat{\theta}-\theta_0| > N^{\delta-1/2})} \\ &\le C(\log N)^{2\lambda}P(|\hat{\theta}-\theta_0| > N^{\delta-1/2}) \\ &= o(N^{-1}). \end{aligned} \tag{8.42}$$

(5.7) then follows by combining (8.38), (8.40)-(8.42). □

PROOF OF THEOREM 5.2. First, we have by (5.14) and (5.15)

$$(8.43)\qquad P(|\hat{\theta}-\hat{\theta}_{i-}|>N^{-1/2-\rho}) \le P(\tilde{\theta}\notin\Theta_{N,\lambda})+P(\tilde{\theta}_{i-}\notin\Theta_{N,\lambda}) + P(|\tilde{\theta}-\tilde{\theta}_{i-}|>N^{-1/2-\rho}) \le CN^{-1-\epsilon}.$$

Here, as before, we use C to denote a constant whose value may change from time to time. We have

$$(8.44)\quad E(\hat{\alpha}_i-\hat{\alpha}_{i-})^2 \le E(\cdots)^2 1_{(|\hat{\theta}-\theta_0|>N^{\delta-1/2})}+E(\cdots)^2 1_{(|\hat{\theta}-\hat{\theta}_{i-}|>N^{-1/2-\rho})} + E(\cdots)^2 1_{(|\hat{\theta}-\theta_0|\le N^{\delta-1/2},|\hat{\theta}-\hat{\theta}_{i-}|\le N^{-1/2-\rho})} = I_1+I_2+I_3.$$

By (8.36) and similar argument as in the previous proof, we have

$$(8.45)\qquad I_i \le C(\log N)^{\lambda}\exp(4N_0(A+4N_0)(\log N)^{\lambda})N^{-1-\epsilon}, \qquad i=1,2.$$

On the other hand, by Taylor expansion, it is easy to show that

$$(8.46)\qquad I_3 \le CN^{-1-2\rho}.$$

Combining (8.44)-(8.46), we see that there is $\tau>\delta$ such that

$$(8.47)\qquad E(\hat{\alpha}_i-\hat{\alpha}_{i-})^2 \le CN^{-1-2\tau}.$$

Also, by the proof of Theorem 5.1, we have

$$(8.48)\qquad E(\hat{\alpha}_{i-}-E(\alpha_i|y))^2 \le CN^{2\delta-1}.$$

Combining (8.47) and (8.48), we have

$$(8.49)\qquad \begin{aligned} |E(\hat{\alpha}_i-\hat{\alpha}_{i-})(\hat{\alpha}_{i-}-\alpha_i)| &= |E(\hat{\alpha}_i-\hat{\alpha}_{i-})(\hat{\alpha}_{i-}-E(\alpha_i|y))| \\ &\le (E(\hat{\alpha}_i-\hat{\alpha}_{i-})^2)^{1/2}(E(\hat{\alpha}_{i-}-E(\alpha_i|y))^2)^{1/2} \\ &\le CN^{\delta-\tau-1}. \end{aligned}$$

The first identity in (5.12) then follows from (5.10), (8.49), and (8.47).
The second identity in (5.12) can be proved by a similar (but easier) way. □

To prove Lemma 5.1-5.4, we need the following lemma.

LEMMA 8.2. *Let $Y_1,\dots,Y_n$ be independent random variables such that $E(Y_i)=0$ and $|Y_i|\le C$. Then*

i) *for any integer $q\ge 1$, there is a constant B_q such that*

$$E\left|\sum_{i=1}^n Y_i\right|^q \le B_q n^{q/2}$$

ii) *for any $\epsilon>0$,*

$$P\left(\left|\sum_{i=1}^n Y_i\right|\ge \epsilon n\right)\le 2\exp\left(-\frac{n}{3}\left(\frac{\epsilon}{C}\right)^2\right).$$

PROOF OF LEMMA 8.2. i) is a special case of the Burkholder's inequality (e.g., Chow and Teicher (1988, §11.2). ii) can be proved by standard argument in large deviation. □

PROOF OF LEMMA 5.1. By the definition of the MM estimator, we have

$$P(|M_N(\tilde{\theta}) - \hat{M}_N| > N^{\delta-1/2}) \leq P(|\hat{M}_N - M_N(\theta_0)| > N^{\delta-1/2}) \leq \sum_{k=1}^{p+1} P(|\hat{M}_{N,k} - M_{N,k}(\theta_0)| > N^{\delta-1/2}/\sqrt{p+1}).$$

For any $1 \leq k \leq p$, we have

$$\hat{M}_{N,k} - M_{N,k}(\theta_0) = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{m}{N}\right)\sum_{j=1}^{n_i} x_{ijk}(y_{ij} - Ey_{ij}) = \frac{1}{m}\sum_{i=1}^{m} Y_i, \tag{8.50}$$

where $Y_1, \ldots, Y_m$ are independent with $E(Y_i) = 0$ and $|Y_i| \leq N_0 A$; and one has a similar expression when $k = p+1$. (5.19) then follows from i) of Lemma 8.2. □

PROOF OF LEMMA 5.2. We have, by Lemma 5.1, that for large N

$$P(\tilde{\theta} \notin \Theta_B) \leq P(|M_N(\tilde{\theta}) - \hat{M}_N| > N^{\delta-1/2}) + P(\tilde{\theta} \notin \Theta_B, |M_N(\tilde{\theta}) - \hat{M}_N| \leq N^{\delta-1/2}) \leq CN^{-q\delta} + P(|\hat{M}_N - M_N(\theta_0)| \geq \epsilon/2).$$

By ii) of Lemma 8.2 and (8.50), it is easy to show that

$$P\left(|\hat{M}_N - M_N(\theta_0)| \geq \frac{\epsilon}{2}\right) \leq 2(p+1)\exp\left(-\frac{m}{12(p+1)}\left(\frac{\epsilon}{N_0 A}\right)^2\right) \leq 2(p+1)e^{-\gamma N}, \tag{8.51}$$

where $\gamma = \epsilon^2/12(p+1)N_0^3 A^2$. The result then follows. □

PROOF OF LEMMA 5.3. Let N_δ be an integer such that $\epsilon_1 N^{\delta-1/2} \geq 2N^{\delta_1-1/2}$, $N \geq N_\delta$. Then, by Lemma 5.2, Lemma 5.1 and its proof, we have that when $N \geq N_\delta$,

$$P(|\tilde{\theta} - \theta_0| > N^{\delta-1/2}) \leq P(\tilde{\theta} \notin \Theta_B) + P(|M_N(\tilde{\theta}) - \hat{M}_N| > N^{\delta_1-1/2}) + P(|\hat{M}_N - M_N(\theta_0)| > N^{\delta_1-1/2}) \leq CN^{-q\delta_1}.$$

Note that $|\tilde{\theta} - \theta_0| > N^{\delta-1/2}$, $\tilde{\theta} \in \Theta_B$, and $|M_N(\tilde{\theta}) - \hat{M}_N| \leq N^{\delta_1-1/2}$ imply that $|\hat{M}_N - M_N(\theta_0)| > N^{\delta_1-1/2}$, when $N \geq N_\delta$. □

PROOF OF LEMMA 5.4. First we note that the components of $\hat{M}_N - \hat{M}_{N,i-}$ is in the form of

$$\frac{1}{N}\sum_{i'=1}^{m}\xi_{i'} - \frac{1}{N-n_i}\sum_{i'\neq i}\xi_{i'} = \frac{1}{N}\xi_i - \frac{n_i}{N(N-n_i)}\sum_{i'\neq i}\xi_{i'},$$

where $\xi_{i'}$, $1 \leq i' \leq m$ are some bounded random variables. It follows that

$$|\hat{M}_N - \hat{M}_{N,i-}| \leq CN^{-1}, \tag{8.52}$$

where and hereafter in the proof C represents a constant whose value may change from time to time. Similarly,

$$|M_N(\tilde{\theta}_{i-}) - M_{N,i-}(\tilde{\theta}_{i-})| \le CN^{-1}. \tag{8.53}$$

Choose δ such that $\delta_1 < \delta < 1/4$ and $0 < \rho < (1-4\delta)/2$. By Taylor expansion, it is easy to show that

$$\begin{aligned} M_{N,k}(\tilde{\theta}) &- M_{N,k}(\tilde{\theta}_{i-}) \\ &= \left(\frac{\partial M_{N,k}(\theta_0)}{\partial\theta}\right)^t (\tilde{\theta} - \tilde{\theta}_{i-}) + \frac{1}{2}(\tilde{\theta} - \theta_0)^t \left(\frac{\partial^2 M_{N,k}(\theta^{(k)})}{\partial\theta^2}\right)(\tilde{\theta} - \theta_0) \\ &\quad -\frac{1}{2}(\tilde{\theta}_{i-} - \theta_0)^t \left(\frac{\partial^2 M_{N,k}(\theta_{(k)})}{\partial\theta^2}\right)(\tilde{\theta}_{i-} - \theta_0) \\ &= \left(\frac{\partial M_{N,k}(\theta_0)}{\partial\theta}\right)^t (\tilde{\theta} - \tilde{\theta}_{i-}) + R_{N,k}, \end{aligned}$$

$1 \le k \le p+1$, where $\theta^{(k)}$ ($\theta_{(k)}$) lies between θ_0 and $\tilde{\theta}$ ($\tilde{\theta}_{i-}$). It is easy to show that all second derivatives of $M_N(\cdot)$ are uniformly bounded. Thus, $|R_{N,k}| \le CN^{2\delta-1}$ if $|\tilde{\theta} - \theta_0| \le N^{\delta-1/2}$ and $|\tilde{\theta}_{i-} - \theta_0| \le N^{\delta-1/2}$. Therefore, by (5.25), $|\tilde{\theta} - \tilde{\theta}_{i-}| > N^{-1/2-\rho}$, $|\tilde{\theta} - \theta_0| \le N^{\delta-1/2}$, and $|\tilde{\theta}_{i-} - \theta_0| \le N^{\delta-1/2}$ imply

$$\begin{aligned} |M_N(\tilde{\theta}) - M_N(\tilde{\theta}_{i-})|^2 &\ge \frac{1}{2}(\tilde{\theta} - \tilde{\theta}_{i-})^t \left(\frac{\partial M_N(\theta_0)}{\partial\theta}\right)^t \left(\frac{\partial M_N(\theta_0)}{\partial\theta}\right)(\tilde{\theta} - \tilde{\theta}_{i-}) \\ &\quad -(p+1)C^2N^{4\delta-2} \\ &\ge 2\epsilon N^{-1-2\rho} - (p+1)C^2N^{4\delta-2} \\ &\ge \epsilon N^{-1-2\rho}, \end{aligned} \tag{8.54}$$

for large N, where ϵ is some positive number. Therefore, when N is large, we have

$$\begin{aligned} P(|\tilde{\theta} - \tilde{\theta}_{i-}| > N^{-1/2-\rho}) &\le P(|\tilde{\theta} - \theta_0| > N^{\delta-1/2}) + P(|\tilde{\theta}_{i-} - \theta_0| > N^{\delta-1/2}) \\ &\quad + P(|M_N(\tilde{\theta}) - M_N(\tilde{\theta}_{i-})| \ge \sqrt{\epsilon}N^{-1/2-\rho}). \end{aligned} \tag{8.55}$$

Since $\rho < 1/2$, and $\hat{M}_N = M_N(\tilde{\theta})$, $\hat{M}_{N,i-} = M_{N,i-}(\tilde{\theta}_{i-})$ when $\hat{M}_N \in M_N(\Theta)$ and $\hat{M}_{N,i-} \in M_{N,i-}(\Theta)$, by (8.52), (8.53), (5.26), and (8.51), there is $\gamma > 0$ such that for large N,

$$\begin{aligned} P(|M_N(\tilde{\theta}) &- M_N(\tilde{\theta}_{i-})| \ge \sqrt{\epsilon}N^{-1/2-\rho}) \\ &\le P(\hat{M}_N \notin M_N(\Theta)) + P(\hat{M}_{N,i-} \notin M_{N,i-}(\Theta)) \\ &\le P(|\hat{M}_N - M_N(\theta_0)| \ge \epsilon_2) \\ &\quad + P(|\hat{M}_{N,i-} - M_{N,i-}(\theta_0)| \ge \epsilon_2) \\ &\le Ce^{-\gamma N}. \end{aligned} \tag{8.56}$$

(5.27) then follows from (8.55), (8.56), and Lemma 5.3. □

To prove Theorem 5.3, we need the following lemma whose proof is straightforward.

LEMMA 8.3. *Let ξ_N, η_N, and ζ_N be random variables such that $|\xi_N| \vee |\eta_N| \le C(\log N)^\lambda$ for some constants C, $\lambda > 0$. Suppose there is a set A_N such that $P(A_N^c) = O(N^{-1-\epsilon})$ for some $\epsilon > 0$, and $\xi_N = \eta_N + \zeta_N$ with $|\zeta_N| \le BN^{-1-\delta}$ on A_N for some constants B, $\delta > 0$. Then $E(\xi_N) = E\eta_N + o(N^{-1-\gamma})$ for any $0 < \gamma < \epsilon \wedge \delta$.*

PROOF OF THEOREM 5.3. By Taylor series expansion, it is easy to show that

$$\hat{\theta} - \theta_0 = \left(\frac{\partial}{\partial\theta}M_N(\theta_0)\right)^{-1}\left(M_N(\hat{\theta}) - M_N(\theta_0) - \frac{1}{2}Q_N\right) + r_N, \tag{8.57}$$

where Q_N is $\hat{q}_N$ with Δ_N replaced by $M_N(\hat{\theta}) - M_N(\theta_0)$, and $|r_N| \le C|\hat{\theta} - \theta_0|^3$ for some constant C. Note that $M_N(\cdot)$ and all its derivatives are uniformly bounded. By (8.57) and, again, Taylor expansion, we have

$$\begin{aligned} d_i(\hat{\theta}) &= d_i(\theta_0) + \left(\frac{\partial}{\partial\theta}d_i(\theta_0)\right)^t\left(\frac{\partial}{\partial\theta}M_N(\theta_0)\right)^{-1}\left(M_N(\hat{\theta}) - M_N(\theta_0) - \frac{1}{2}Q_N\right) \\ &\quad + \frac{1}{2}R_N + s_N, \end{aligned} \tag{8.58}$$

where R_N is $\hat{r}_N$ with Δ_N replaced by $M_N(\hat{\theta}) - M_N(\theta_0)$, and $|s_N| \le C|\hat{\theta} - \theta_0|^3$ for some constant C. Let $\xi_N = d_i(\hat{\theta})$, η_N be the sum of the first three terms on the right side of (8.58) with $M_N(\hat{\theta})$ replaced by $\hat{M}_N$, and $\zeta_N = s_N$. Then ξ_N and η_N are bounded random variables. Let $A_N = \{\tilde{\theta} \in \Theta_B, |\tilde{\theta} - \theta_0| \le N^{\delta-1/2}$, and $\hat{M}_N \in M_N(\Theta)\}$, where $0 < \delta < 1/6$. Then, by Lemma 5.2, Lemma 5.3, and the proof of Lemma 5.4, we have, for some $\gamma > 0$ and any $0 < \delta_1 < \delta$ and integer $q \ge 1$,

$$\begin{aligned} P(A_N^c) &\le P(\tilde{\theta} \notin \Theta_B) + P(|\tilde{\theta} - \theta_0| > N^{\delta-1/2}) + P(\hat{M}_N \notin M_N(\Theta)) \\ &= O(N^{-q\delta}) + O(N^{-q\delta_1}) + O(e^{-\gamma N}) = O(N^{-1-\epsilon}) \end{aligned}$$

for some $\epsilon > 0$, if q is large. Also, if N is large, then on A_N we have $\hat{\theta} = \tilde{\theta}$, $|\hat{\theta} - \theta_0| \le N^{\delta-1/2}$, and $M_N(\hat{\theta}) = \hat{M}_N$, which, combined with (8.58), implies that $\xi_N = \eta_N + \zeta_N$ with $|\zeta_N| \le CN^{3\delta-3/2}$. (5.33) then follows from Lemma 8.3.

By a similar argument, one can show that $E(\widehat{\mathrm{MSE}}(\hat{\alpha}_i) - \mathrm{MSE}(\hat{\alpha}_i)) = o(N^{-1})$. □

Acknowledgements

The research of the two authors was supported, in part, by National Science Foundation Grant SBR-9705574. The research of the first author was also supported by National Security Agency Grant MDA904-98-1-0038. The authors wish to thank an Associate Editor and two referees for their careful review of the manuscript. One of the referees also provides an alternative and more intuitive proof for Lemma 3.1.

REFERENCES

Chow, Y. S. and Teicher, H. (1988). *Probability Theory*, (2nd ed), Springer, New York.

De Bruijn, N. G. (1961). *Asymptotic Methods in Analysis*, North-Holland, Amsterdam.

Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach, *J. Amer. Statist. Assoc.*, **68**, 117–130.

Farrell, P., MacGibbon, B. and Tomberlin, T. J. (1997). Empirical Bayes estimators of small area proportions in multistage designs, *Statist. Sinica*, **7**, 1065–1084.

Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal (with discussion), *Statist. Sci.*, 9, 55–93.

Ghosh, M., Natarajan, K., Stroud, T. W. F. and Carlin, B. P. (1998). Generalized linear models for small-area estimation, *J. Amer. Statist. Assoc.*, **93**, 273–282.

Holt, D., Smith, T. M. F. and Tomberlin, T. J. (1979). A model-based approach to estimation for small subgroups of a population, *J. Amer. Statist. Assoc.*, **74**, 405–410.

Jiang, J. (1998). Consistent estimators in generalized linear mixed models, *J. Amer. Statist. Assoc.*, **93**, 720–729.

Lahiri, P. and Rao, J. N. K. (1995). Robust estimation of mean squared error of small area estimators, *J. Amer. Statist. Assoc.*, 90, 758–766.

Malec, D., Sedransk, J., Moriarity, C. L. and LeClere, F. B. (1997). Small area inference for binary variables in the national health interview survey, *J. Amer. Statist. Assoc.*, **92**, 815–826.

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators, *J. Amer. Statist. Assoc.*, **85**, 163–171.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance Components*, Wiley, New York.