

## M-ESTIMATION OF THE MIXED-TYPE GENERALIZED LINEAR MODEL

BY

YING DONG\* (DALIAN), LIXIN SONG\*\* (DALIAN),  
MINGQIU WANG (QUFU), AND MUHAMMAD AMIN (PESHAWAR)

*Abstract.* To investigate the features of the individual from the mixed-type model, a novel model, named the mixed-type generalized linear model, is proposed firstly in this work, which is verified to be realistic and useful. We consider the robustness of M-estimation to estimate the unknown parameters of the mixed-type generalized linear model. By applying the law of large numbers and the central limit theorem, the consistency and asymptotic normality of the M-estimation for the mixed-type generalized linear model are proved with regularity assumptions. At last, in order to evaluate the finite sample performance of the estimator for the new model, several applied instances are presented, which show the good performance of the estimator.

**2010 AMS Mathematics Subject Classification:** Primary: 62F12, 62J05; Secondary: 62J07.

**Key words and phrases:** Asymptotic normality, consistency, M-estimation, mixed-type generalized linear model.

### 1. INTRODUCTION

The usual methods of parameter estimation are the least squares method, the maximum likelihood method, the quasi-likelihood method, etc. The nonrobustness of the maximum likelihood estimator and quasi-likelihood estimator for parameters are extensively reported in the literature, such as McCullagh and Nelder [11], Heyde [3], and so on. As an alternative robust method, M-estimation has been researched extensively. See Huber [5]–[9], Yohai and Maronna [16] and Bai et al. [1], [2] for details.

The generalized linear model (GLM) is a class of widely used statistical models, which was proposed by Nelder and Wedderburn [13] as a way of unifying var-

---

\* The first author is supported by the NSFC Special Fund of Tianyuan Visiting Scholar Fund (11726616), the Doctor Startup Fund of Liaoning Province (201501164), and the Fundamental Research Funds for the Central Universities (DC 201501040, DCPY 2016064, DCPY 2018056).

\*\* Corresponding author.

ious other statistical models. The GLM is a direct extension of the general linear model. Its response variable follows any probability distribution in the exponential family of distributions, such as logistic model, Poisson model, logarithmic model, survival data model, and so on. The parameters in these models can be estimated using a common calculation methodology. It is the common feature that motivates us to study the GLM as a unified theory.

Much progress has been made in terms of the GLM and M-estimation. Preisser and Qaqish [14] considered a class of robust estimators in the generalized framework of generalized estimating equation. Sinha [15] studied robust analysis in generalized linear mixed models. Mancini et al. [10] and Muler and Yohai [12] proposed a robust M-estimator that assigns a much lower weight to outliers than does the Gaussian maximum likelihood estimator. Huang and Song [4] discussed M-estimator of a generalized linear model with measurement errors.

The mixture phenomenon appears more and more commonly, so we should pay much attention to investigate the features of the individual from the mixed-type model. As the generalized linear model is widely used in the actuarial area, the insurance companies can consider to set up insurance premiums according to the male drivers' claim pattern (male drivers belong to one kind of distribution), and also to the female drivers' claim pattern (female drivers belong to another kind of distribution). So the compensation scheme could be determined by considering on both male and female drivers. These problems cannot be solved by the existing models.

Motivated by this situation, a new model, named the mixed-type generalized linear model, is proposed in this work, and the M-estimation of the mixed-type generalized linear model is discussed. Furthermore, it is proved that the estimator is consistent and asymptotically normal with the appropriate assumptions. The simulation studies demonstrate the good performance of the proposed model.

This paper is organized as follows. In Section 2, the preliminaries and assumptions including the basic structure of the mixed-type GLM and the concept of M-estimation are given. The main results with full proofs are presented in Section 3. Some applied instances to verify the performance of the estimator are given in Section 4. At last, a conclusion is given in Section 5.

## 2. PRELIMINARIES AND ASSUMPTIONS

**2.1. Mixed-type generalized linear model and M-estimation.** The response variable  $Y$  in the model

$$(2.1) \quad E(Y|\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = pF_1(\beta_{10}^T \mathbf{X}^{(1)}) + qF_2(\beta_{20}^T \mathbf{X}^{(2)})$$

is considered, where  $p + q = 1$ . It comes from the two-type generalized linear

model

$$(2.2) \quad \begin{cases} Y^{(1)} = F_1(\boldsymbol{\beta}_{10}^T \mathbf{X}^{(1)}) + \varepsilon^{(1)}, \\ Y^{(2)} = F_2(\boldsymbol{\beta}_{20}^T \mathbf{X}^{(2)}) + \varepsilon^{(2)}, \end{cases}$$

where  $F_1(\cdot)$  and  $F_2(\cdot)$  are given strictly monotonic and continuous functions. Also,  $E(\varepsilon^{(1)}) = 0$ ,  $\text{Var}(\varepsilon^{(1)}) = V_1(\boldsymbol{\beta}_{10}^T \mathbf{X}^{(1)})$ ,  $E(\varepsilon^{(2)}) = 0$  and  $\text{Var}(\varepsilon^{(2)}) = V_2(\boldsymbol{\beta}_{20}^T \mathbf{X}^{(2)})$ , where  $V_1(\cdot)$  and  $V_2(\cdot)$  are nonnegative, continuous and bounded functions. It can be observed that (2.1) is the mixture of the two-types  $Y^{(1)}$  and  $Y^{(2)}$ . The proportion of the first type  $Y^{(1)}$  is  $p$ , while the proportion of the second type  $Y^{(2)}$  is  $q$ . Meanwhile, a whole data  $\{\mathbf{X}_i^{(1)}, \mathbf{X}_i^{(2)}, Y_i\}$  ( $i = 1, 2, \dots, N$ ) can be obtained from this model. Then the mixed-type generalized linear model can be defined as

$$(2.3) \quad Y = pF_1(\boldsymbol{\beta}_{10}^T \mathbf{X}^{(1)}) + qF_2(\boldsymbol{\beta}_{20}^T \mathbf{X}^{(2)}) + \varepsilon,$$

where  $p + q = 1$  and  $E(\varepsilon) = 0$ ,  $\text{Var}(\varepsilon) < \infty$ .  $\Theta = \Theta_1 \otimes \Theta_2$  is the parameter space which is a bounded closed set. The true parameter  $\boldsymbol{\beta}_{10}$  is a  $d_1$ -dimension in  $\Theta_1$ , and the true parameter  $\boldsymbol{\beta}_{20}$  is a  $d_2$ -dimension in  $\Theta_2$ .  $\{\mathbf{X}_i^{(\delta)}\}$  ( $i = 1, 2, \dots, N$ ) are i.i.d. random variables with  $\delta = 1, 2$ , respectively.

A robust method provides a useful and stable alternative that is not sensitive to outliers. Huber [5] introduced M-estimation of  $\boldsymbol{\beta}$ , which is defined as any value of  $\hat{\boldsymbol{\beta}}$  that minimizes

$$\sum_{i=1}^N \rho(Y_i - pF_1(\boldsymbol{\beta}_1^T \mathbf{X}_i^{(1)}) - qF_2(\boldsymbol{\beta}_2^T \mathbf{X}_i^{(2)})),$$

with a suitable choice of the function  $\rho(\cdot)$  which is the function of  $\boldsymbol{\beta}$ . In general,  $\rho(\cdot)$  is a given nonnegative function. The least absolute deviation and ordinary least squares are the special cases of M-estimation.

In fact, the exact form of  $\varepsilon$  in (2.3) can be

$$\varepsilon = F_1(\boldsymbol{\beta}_{10}^T \mathbf{X}^{(1)})(I_{\{\delta=1\}} - p) + F_2(\boldsymbol{\beta}_{20}^T \mathbf{X}^{(2)})(I_{\{\delta=2\}} - q) + \varepsilon^{(1)}I_{\{\delta=1\}} + \varepsilon^{(2)}I_{\{\delta=2\}},$$

the events of  $\{\delta = 1\}$  and  $\{\delta = 2\}$  are unobservable, but the  $\Pr\{\delta = 1\} = p$  and  $\Pr\{\delta = 2\} = q$  are given. By further calculation, we obtain

$$E(\varepsilon) = 0,$$

$$\text{Var}(\varepsilon) = pq(F_1(\boldsymbol{\beta}_{10}^T \mathbf{X}^{(1)}) - F_2(\boldsymbol{\beta}_{20}^T \mathbf{X}^{(2)}))^2 + pV_1(\boldsymbol{\beta}_{10}^T \mathbf{X}^{(1)}) + qV_2(\boldsymbol{\beta}_{20}^T \mathbf{X}^{(2)}).$$

**2.2. Assumptions.** To achieve the asymptotic results of the estimator, the following assumptions are made.

(A1) The order of the integral and limit can be changed in the proofs.

(A2)  $\rho(u)$  is the symmetric convex function and monotonously increases on  $[0, \infty)$ , and  $\rho(0) = 0$ . There exists  $C_1 > 0$  such that  $0 \leq \rho''(u) \leq C_1$ . Let  $L = E\{(\rho'(\varepsilon))^2 | \mathbf{X}^{(1)}, \mathbf{X}^{(2)}\}$ . For any  $a$  and  $b$ ,  $E[\rho(\tilde{\varepsilon} + a(I_{\{\delta=1\}} - p) + b(I_{\{\delta=2\}} - q) + A)]$  reaches its unique minimum when  $A=0$ , where  $\tilde{\varepsilon} = \varepsilon^{(1)}I_{\{\delta=1\}} + \varepsilon^{(2)}I_{\{\delta=2\}}$ .

(A3)  $pF_1(\beta_{10}^T \mathbf{X}^{(1)}) + qF_2(\beta_{20}^T \mathbf{X}^{(2)}) = pF_1(\beta_1^T \mathbf{X}^{(1)}) + qF_2(\beta_2^T \mathbf{X}^{(2)})$  if and only if  $(\beta_1 = \beta_{10}$  and  $\beta_2 = \beta_{20})$ .

(A4) For any  $\beta_1 \in \Theta_1$  and  $\beta_2 \in \Theta_2$ , there exists  $C_2 > 0$  such that

$$E|\rho'(Y - pF_1(\beta_1^T \mathbf{X}^{(1)}) - qF_2(\beta_2^T \mathbf{X}^{(2)}))| < C_2.$$

(A5) Let  $\mathbf{X}^{(1)}$  and  $\mathbf{X}^{(2)}$  be bounded a.s., which means that there exists  $C_3 > 0$  such that  $\max\{\|\mathbf{X}^{(1)}\|, \|\mathbf{X}^{(2)}\|\} \leq C_3$ .

(A6) The two functions  $F_1(\cdot)$  and  $F_2(\cdot)$  have the second order derivatives.

(A7)  $F_1(\cdot)$  and  $F_2(\cdot)$  are not logarithmic functions (in other words, the first order derivatives of  $F_1(\cdot)$  and  $F_2(\cdot)$  are not the inverse functions).

(A8) Let us put

$$S_N(\beta_1, \beta_2) = \frac{1}{N} \sum_{i=1}^N \rho(Y_i - pF_1(\beta_1^T \mathbf{X}_i^{(1)}) - qF_2(\beta_2^T \mathbf{X}_i^{(2)}))$$

and

$$S(\beta_1, \beta_2) = E\{\rho(Y - pF_1(\beta_1^T \mathbf{X}^{(1)}) - qF_2(\beta_2^T \mathbf{X}^{(2)}))\}.$$

Suppose  $S(\beta_1, \beta_2)$  is a continuous function, and  $S_N(\beta_1, \beta_2)$  has the third order derivative in the neighborhood of the true value parameters.

### 3. MAIN RESULTS

**3.1. Consistency for the M-estimator of the mixed-type GLM.** The consistency for the M-estimator of the mixed-type GLM can be proved by using Theorem 3.1.

**THEOREM 3.1.** *Let  $(\hat{\beta}_{1N}, \hat{\beta}_{2N})$  be the minimizer of  $S_N(\beta_1, \beta_2)$  in  $\Theta$ . Then we have*

$$(i) \lim_{N \rightarrow \infty} \hat{\beta}_{1N} = \beta_{10} \text{ a.s.};$$

$$(ii) \lim_{N \rightarrow \infty} \hat{\beta}_{2N} = \beta_{20} \text{ a.s.}$$

In order to prove Theorem 3.1, we need the following propositions.

**PROPOSITION 3.1.** *Under conditions (A1)–(A7), for any  $\alpha_1, \beta_1 \in \Theta_1$  and  $\alpha_2, \beta_2 \in \Theta_2$ , there exists  $N_0 = N_0(\alpha_1, \alpha_2)$  such that when  $N > N_0$ , we have*

$$(i) |S_N(\beta_1, \beta_2) - S_N(\alpha_1, \alpha_2)| \leq C_2 C_3 M(\|\beta_1 - \alpha_1\| + \|\beta_2 - \alpha_2\|) + C_1 C_3^2 M^2(\|\beta_1 - \alpha_1\|^2 + \|\beta_2 - \alpha_2\|^2);$$

$$(ii) \quad |S(\beta_1, \beta_2) - S(\alpha_1, \alpha_2)| \leq C_2 C_3 M (\|\beta_1 - \alpha_1\| + \|\beta_2 - \alpha_2\|) + C_1 C_3^2 M^2 (\|\beta_1 - \alpha_1\|^2 + \|\beta_2 - \alpha_2\|^2).$$

Proof. (i) For  $i = 1, 2, \dots, N$ , there exists  $\xi_i$  between  $Y_i - pF_1(\alpha_1^T X_i^{(1)}) - qF_2(\alpha_2^T X_i^{(2)})$  and  $Y_i - pF_1(\beta_1^T X_i^{(1)}) - qF_2(\beta_2^T X_i^{(2)})$ . According to the conditions (A5) and (A6), there exist real numbers  $L_1$  and  $L_2$  such that  $|F_1'(\beta^T X^{(1)})| \leq L_1$  and  $|F_2'(\beta^T X^{(2)})| \leq L_2$ . Let  $M = \max\{pL_1, qL_2\}$ ; we have

$$\begin{aligned} & |\rho(Y_i - pF_1(\beta_1^T X_i^{(1)}) - qF_2(\beta_2^T X_i^{(2)})) - \rho(Y_i - pF_1(\alpha_1^T X_i^{(1)}) - qF_2(\alpha_2^T X_i^{(2)}))| \\ & \leq \rho'(Y_i - pF_1(\alpha_1^T X_i^{(1)}) - qF_2(\alpha_2^T X_i^{(2)})) \\ & \quad \times (L_1 \cdot p \cdot \|\beta_1 - \alpha_1\| \cdot \|X^{(1)}\| + L_2 \cdot q \cdot \|\beta_2 - \alpha_2\| \cdot \|X^{(2)}\|) \\ & \quad + \frac{\rho''(\xi_i)}{2} [pL_1(\beta_1 - \alpha_1)^T X_i^{(1)} + qL_2(\beta_2 - \alpha_2)^T X_i^{(2)}]^2 \\ & \leq \rho'(Y_i - pF_1(\alpha_1^T X_i^{(1)}) - qF_2(\alpha_2^T X_i^{(2)})) \\ & \quad \times M \cdot (\|\beta_1 - \alpha_1\| \cdot \|X^{(1)}\| + \|\beta_2 - \alpha_2\| \cdot \|X^{(2)}\|) \\ & \quad + \frac{C_1}{2} \cdot M^2 \cdot 2 \cdot (\|\beta_1 - \alpha_1\|^2 \cdot \|X^{(1)}\|^2 + \|\beta_2 - \alpha_2\|^2 \cdot \|X^{(2)}\|^2) \\ & \leq C_3 M \rho'(Y_i - pF_1(\alpha_1^T X_i^{(1)}) - qF_2(\alpha_2^T X_i^{(2)})) (\|\beta_1 - \alpha_1\| + \|\beta_2 - \alpha_2\|) \\ & \quad + C_1 C_3^2 M^2 (\|\beta_1 - \alpha_1\|^2 + \|\beta_2 - \alpha_2\|^2). \end{aligned}$$

According to the Kolmogorov law of large numbers, we have

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \{ \rho'(Y_i - pF_1(\alpha_1^T X_i^{(1)}) - qF_2(\alpha_2^T X_i^{(2)})) \} \\ = E \{ \rho'(Y_i - pF_1(\alpha_1^T X_i^{(1)}) - qF_2(\alpha_2^T X_i^{(2)})) \} \text{ a.s.} \end{aligned}$$

Then there exists  $N_0 = N_0(\alpha_1, \alpha_2)$  such that when  $N > N_0$ , by the condition (A4), we have

$$\frac{1}{N} \sum_{i=1}^N \{ \rho'(Y_i - pF_1(\alpha_1^T X_i^{(1)}) - qF_2(\alpha_2^T X_i^{(2)})) \} \leq C_2.$$

So, we obtain

$$\begin{aligned} & |S_N(\beta_1, \beta_2) - S_N(\alpha_1, \alpha_2)| \\ & \leq \frac{1}{N} \sum_{i=1}^N \{ \rho(Y_i - pF_1(\beta_1^T X_i^{(1)}) - qF_2(\beta_2^T X_i^{(2)})) \\ & \quad - \rho(Y_i - pF_1(\alpha_1^T X_i^{(1)}) - qF_2(\alpha_2^T X_i^{(2)})) \} \end{aligned}$$

$$\begin{aligned}
&\leq \frac{C_3}{N} \cdot M \cdot \sum_{i=1}^N \rho'(Y_i - pF_1(\alpha_1^T \mathbf{X}_i^{(1)}) - qF_2(\alpha_2^T \mathbf{X}_i^{(2)})) (\|\beta_1 - \alpha_1\| + \|\beta_2 - \alpha_2\|) \\
&\quad + C_1 \cdot M^2 \cdot C_3^2 (\|\beta_1 - \alpha_1\|^2 + \|\beta_2 - \alpha_2\|^2) \\
&\leq C_2 C_3 M (\|\beta_1 - \alpha_1\| + \|\beta_2 - \alpha_2\|) + C_1 C_3^2 M^2 (\|\beta_1 - \alpha_1\|^2 + \|\beta_2 - \alpha_2\|^2).
\end{aligned}$$

(ii) Similarly, we have

$$\begin{aligned}
&|S(\beta_1, \beta_2) - S(\alpha_1, \alpha_2)| \\
&= |E\{\rho(Y - pF_1(\beta_1^T \mathbf{X}^{(1)}) - qF_2(\beta_2^T \mathbf{X}^{(2)})) \\
&\quad - \rho(Y - pF_1(\alpha_1^T \mathbf{X}^{(1)}) - qF_2(\alpha_2^T \mathbf{X}^{(2)}))\}| \\
&\leq C_3 M \cdot E\{\rho'(Y - pF_1(\alpha_1^T \mathbf{X}^{(1)}) - qF_2(\alpha_2^T \mathbf{X}^{(2)}))\} (\|\beta_1 - \alpha_1\| + \|\beta_2 - \alpha_2\|) \\
&\quad + C_1 C_3^2 M^2 (\|\beta_1 - \alpha_1\|^2 + \|\beta_2 - \alpha_2\|^2) \\
&\leq C_2 C_3 M (\|\beta_1 - \alpha_1\| + \|\beta_2 - \alpha_2\|) + C_1 C_3^2 M^2 (\|\beta_1 - \alpha_1\|^2 + \|\beta_2 - \alpha_2\|^2).
\end{aligned}$$

This completes the proof of Proposition 3.1. ■

PROPOSITION 3.2. *Under conditions (A1)–(A7), we have*

$$\lim_{N \rightarrow \infty} \sup_{\substack{\beta_1 \in \Theta_1 \\ \beta_2 \in \Theta_2}} |S_N(\beta_1, \beta_2) - S(\beta_1, \beta_2)| = 0 \text{ a.s.}$$

Proof.  $\Theta$  is a bounded closed set; for any  $\omega > 0$ , there exists an  $\eta(\omega)$ -net, where

$$\eta(\omega) = \frac{\omega}{6(C_2 C_3 M + C_1 C_3^2 M^2)}.$$

We take  $\{\beta_1^{(1)}, \beta_1^{(2)}, \dots, \beta_1^{(s)}\}$  and  $\{\beta_2^{(1)}, \beta_2^{(2)}, \dots, \beta_2^{(t)}\}$  from the  $\eta(\omega)$ -net. Then, for any  $\beta_1 \in \Theta_1$  and  $\beta_2 \in \Theta_2$ , there exist  $j$  and  $k$  such that

$$\max\{\|\beta_1 - \beta_1^{(j)}\|, \|\beta_2 - \beta_2^{(k)}\|\} \leq \min\{\eta(\omega), 1\}.$$

By the Kolmogorov law of large numbers, we have

$$\lim_{N \rightarrow \infty} (S_N(\beta_1^{(j)}, \beta_2^{(k)}) - S(\beta_1^{(j)}, \beta_2^{(k)})) = 0,$$

i.e. for the above-mentioned  $\omega > 0$ , there exists  $\tilde{N}_{j,k} > 0$  such that, when  $N > \tilde{N}_{j,k}$ , we have

$$(3.1) \quad |S_N(\beta_1^{(j)}, \beta_2^{(k)}) - S(\beta_1^{(j)}, \beta_2^{(k)})| < \frac{\omega}{3}.$$

By Proposition 3.1(i), there exist finite numbers  $N_1, N_2, \dots, N_{t \times s}$ , and we take  $N^* = \max\{N_1, N_2, \dots, N_{t \times s}\}$ . When  $N > N^*$ , we have

$$\begin{aligned}
 (3.2) \quad & |S_N(\beta_1, \beta_2) - S_N(\beta_1^{(j)}, \beta_2^{(k)})| \\
 & \leq C_2 C_3 M (\|\beta_1 - \beta_1^{(j)}\| + \|\beta_2 - \beta_2^{(k)}\|) \\
 & \quad + C_1 C_3^2 M^2 (\|\beta_1 - \beta_1^{(j)}\|^2 + \|\beta_2 - \beta_2^{(k)}\|^2) \\
 & \leq (C_2 C_3 M + C_1 C_3^2 M^2) (\|\beta_1 - \beta_1^{(j)}\| + \|\beta_2 - \beta_2^{(k)}\|) < \frac{\omega}{3}.
 \end{aligned}$$

By Proposition 3.1(ii), we obtain

$$\begin{aligned}
 (3.3) \quad & |S(\beta_1^{(j)}, \beta_2^{(k)}) - S(\beta_1, \beta_2)| \\
 & \leq C_2 C_3 M (\|\beta_1^{(j)} - \beta_1\| + \|\beta_2^{(k)} - \beta_2\|) \\
 & \quad + C_1 C_3^2 M^2 (\|\beta_1^{(j)} - \beta_1\|^2 + \|\beta_2^{(k)} - \beta_2\|^2) \\
 & \leq (C_2 C_3 M + C_1 C_3^2 M^2) (\|\beta_1 - \beta_1^{(j)}\| + \|\beta_2 - \beta_2^{(k)}\|) < \frac{\omega}{3}.
 \end{aligned}$$

Then, combining (3.1), (3.2) and (3.3), let

$$\tilde{N} = \max_{1 \leq j \leq s, 1 \leq k \leq t} \{N^*, \tilde{N}_{j,k}\},$$

when  $N > \tilde{N}$ , and we have

$$\begin{aligned}
 & |S_N(\beta_1, \beta_2) - S(\beta_1, \beta_2)| \\
 & \leq |S_N(\beta_1, \beta_2) - S_N(\beta_1^{(j)}, \beta_2^{(k)})| + |S_N(\beta_1^{(j)}, \beta_2^{(k)}) - S(\beta_1^{(j)}, \beta_2^{(k)})| \\
 & \quad + |S(\beta_1^{(j)}, \beta_2^{(k)}) - S(\beta_1, \beta_2)| \\
 & < \frac{\omega}{3} + \frac{\omega}{3} + \frac{\omega}{3} = \omega.
 \end{aligned}$$

So, we conclude that

$$\lim_{N \rightarrow \infty} \sup_{\substack{\beta_1 \in \Theta_1 \\ \beta_2 \in \Theta_2}} |S_N(\beta_1, \beta_2) - S(\beta_1, \beta_2)| = 0.$$

This completes the proof of Proposition 3.2. ■

**Proof of Theorem 3.1.** It can be shown that

$$\begin{aligned}
 & S(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \\
 &= E\{\rho(Y - pF_1(\boldsymbol{\beta}_1^T \mathbf{X}^{(1)}) - qF_2(\boldsymbol{\beta}_2^T \mathbf{X}^{(2)}))\} \\
 &= E\{\rho(pF_1(\boldsymbol{\beta}_{10}^T \mathbf{X}^{(1)}) + qF_2(\boldsymbol{\beta}_{20}^T \mathbf{X}^{(2)}) - pF_1(\boldsymbol{\beta}_1^T \mathbf{X}^{(1)}) - qF_2(\boldsymbol{\beta}_2^T \mathbf{X}^{(2)}) + \varepsilon)\} \\
 &= E\left\{E\left\{\rho(pF_1(\boldsymbol{\beta}_{10}^T \mathbf{X}^{(1)}) + qF_2(\boldsymbol{\beta}_{20}^T \mathbf{X}^{(2)}) - pF_1(\boldsymbol{\beta}_1^T \mathbf{X}^{(1)}) \right. \right. \\
 &\quad \left. \left. - qF_2(\boldsymbol{\beta}_2^T \mathbf{X}^{(2)}) + \varepsilon\right) \mid \mathbf{X}^{(1)}, \mathbf{X}^{(2)}\right\} \\
 &= E\left\{E\left\{\rho(pF_1(\boldsymbol{\beta}_{10}^T \mathbf{X}^{(1)}) + qF_2(\boldsymbol{\beta}_{20}^T \mathbf{X}^{(2)}) - pF_1(\boldsymbol{\beta}_1^T \mathbf{X}^{(1)}) - qF_2(\boldsymbol{\beta}_2^T \mathbf{X}^{(2)}) \right. \right. \\
 &\quad + F_1(\boldsymbol{\beta}_{10}^T \mathbf{X}^{(1)})(I_{\{\delta=1\}} - p) + F_2(\boldsymbol{\beta}_{20}^T \mathbf{X}^{(2)})(I_{\{\delta=2\}} - q) \\
 &\quad \left. \left. + \varepsilon^{(1)}I_{\{\delta=1\}} + \varepsilon^{(2)}I_{\{\delta=2\}}\right) \mid \mathbf{X}^{(1)}, \mathbf{X}^{(2)}\right\} \\
 &\triangleq E\left\{E\left\{\rho(A + a(I_{\{\delta=1\}} - p) + b(I_{\{\delta=2\}} - q) + \tilde{\varepsilon}) \mid \mathbf{X}^{(1)}, \mathbf{X}^{(2)}\right\}\right\},
 \end{aligned}$$

where  $A = pF_1(\boldsymbol{\beta}_{10}^T \mathbf{X}^{(1)}) + qF_2(\boldsymbol{\beta}_{20}^T \mathbf{X}^{(2)}) - pF_1(\boldsymbol{\beta}_1^T \mathbf{X}^{(1)}) - qF_2(\boldsymbol{\beta}_2^T \mathbf{X}^{(2)})$ ,  $a = F_1(\boldsymbol{\beta}_{10}^T \mathbf{X}^{(1)})$ ,  $b = F_2(\boldsymbol{\beta}_{20}^T \mathbf{X}^{(2)})$  and  $\tilde{\varepsilon} = \varepsilon^{(1)}I_{\{\delta=1\}} + \varepsilon^{(2)}I_{\{\delta=2\}}$ .

By the condition (A2), for any  $a$  and  $b$ , it follows that  $E(\rho(\tilde{\varepsilon} + a(I_{\{\delta=1\}} - p) + b(I_{\{\delta=2\}} - q) + A))$  achieves its minimum when  $A = 0$ , where  $\tilde{\varepsilon} = \varepsilon^{(1)}I_{\{\delta=1\}} + \varepsilon^{(2)}I_{\{\delta=2\}}$ . That is to say, when  $pF_1(\boldsymbol{\beta}_{10}^T \mathbf{X}^{(1)}) + qF_2(\boldsymbol{\beta}_{20}^T \mathbf{X}^{(2)}) - pF_1(\boldsymbol{\beta}_1^T \mathbf{X}^{(1)}) - qF_2(\boldsymbol{\beta}_2^T \mathbf{X}^{(2)}) = 0$  a.s.,  $S(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$  achieves its minimum.

By the condition (A3), we know that when  $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$  and  $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}$ ,  $S(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$  achieves its minimum, and it is the unique minimizer of  $S(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ . So we have

$$\begin{aligned}
 S_N(\hat{\boldsymbol{\beta}}_{1N}, \hat{\boldsymbol{\beta}}_{2N}) - S(\hat{\boldsymbol{\beta}}_{1N}, \hat{\boldsymbol{\beta}}_{2N}) &\leq S_N(\hat{\boldsymbol{\beta}}_{1N}, \hat{\boldsymbol{\beta}}_{2N}) - S(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}) \\
 &\leq S_N(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}) - S(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}).
 \end{aligned}$$

Obviously, we have

$$\lim_{N \rightarrow \infty} (S_N(\hat{\boldsymbol{\beta}}_{1N}, \hat{\boldsymbol{\beta}}_{2N}) - S(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20})) = 0 \text{ a.s.},$$

and

$$\lim_{N \rightarrow \infty} (S_N(\hat{\boldsymbol{\beta}}_{10}, \hat{\boldsymbol{\beta}}_{20}) - S(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20})) = 0 \text{ a.s.}$$

By the squeeze theorem, we obtain

$$(3.4) \quad \lim_{N \rightarrow \infty} (S_N(\hat{\boldsymbol{\beta}}_{1N}, \hat{\boldsymbol{\beta}}_{2N}) - S(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20})) = 0 \text{ a.s.}$$

As  $\Theta$  is a bounded closed set, there exist convergent subseries  $\{\hat{\boldsymbol{\beta}}_{1\hat{N}}\}$  and  $\{\hat{\boldsymbol{\beta}}_{2\hat{N}}\}$  of the series  $\{\hat{\boldsymbol{\beta}}_{1N}\}$  and  $\{\hat{\boldsymbol{\beta}}_{2N}\}$ , respectively. Then we have

$$\lim_{\hat{N} \rightarrow \infty} \hat{\boldsymbol{\beta}}_{1\hat{N}} = \boldsymbol{\beta}_1^0, \quad \boldsymbol{\beta}_1^0 \in \Theta_1 \subset \Theta \text{ a.s.},$$



and

$$\lim_{\hat{N} \rightarrow \infty} \hat{\beta}_{2\hat{N}} = \beta_2^0, \quad \beta_2^0 \in \Theta_2 \subset \Theta \text{ a.s.}$$

Because  $S(\beta_1, \beta_2)$  is continuous, we get

$$\lim_{\hat{N} \rightarrow \infty} S(\hat{\beta}_{1\hat{N}}, \hat{\beta}_{2\hat{N}}) = S(\beta_1^0, \beta_2^0).$$

According to (3.4), we have

$$\lim_{N \rightarrow \infty} S_N(\hat{\beta}_{1N}, \hat{\beta}_{2N}) = S(\beta_{10}, \beta_{20}).$$

As  $\{\hat{\beta}_{1\hat{N}}\}$  and  $\{\hat{\beta}_{2\hat{N}}\}$  are the convergent subseries of  $\{\hat{\beta}_{1N}\}$  and  $\{\hat{\beta}_{2N}\}$ , we have

$$S(\beta_1^0, \beta_2^0) = S(\beta_{10}, \beta_{20}).$$

And from the uniqueness of the minimizer, we obtain

$$\beta_1^0 = \beta_{10} \quad \text{and} \quad \beta_2^0 = \beta_{20}.$$

So we have

$$\lim_{N \rightarrow \infty} \hat{\beta}_{1N} = \beta_{10} \quad \text{and} \quad \lim_{N \rightarrow \infty} \hat{\beta}_{2N} = \beta_{20}.$$

This completes the proof of Theorem 3.1. ■

**3.2. Asymptotic normality for the M-estimator of the mixed-type GLM.** The asymptotic normality for the M-estimator of the mixed-type GLM can be proved by using Theorem 3.2.

**THEOREM 3.2.** *Under conditions (A1)–(A7), we have*

$$\sqrt{N}(\hat{\beta}_N - \beta_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, K^{-1}\Sigma K^{-1}),$$

where

$$\beta_0 = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix}, \quad \hat{\beta}_N = \begin{pmatrix} \hat{\beta}_{1N} \\ \hat{\beta}_{2N} \end{pmatrix}, \quad K = \begin{pmatrix} K^{(1)} & K^{(1,2)} \\ K^{(1,2)T} & K^{(2)} \end{pmatrix},$$

$$K^{(1,2)} \triangleq E\{ [pqF'_1(\beta_{10}^T \mathbf{X}_1^{(1)})F'_2(\beta_{20}^T \mathbf{X}_1^{(2)})\rho''(\varepsilon_1)] \mathbf{X}_1^{(1)} \mathbf{X}_1^{(2)T} \},$$

$$K^{(1)} \triangleq E\{ [p^2\rho''(\varepsilon_1)(F'_1(\beta_{10}^T \mathbf{X}_1^{(1)}))^2 - p\rho'(\varepsilon_1)(F''_1(\beta_{10}^T \mathbf{X}_1^{(1)}))] \mathbf{X}_1^{(1)} \mathbf{X}_1^{(1)T} \},$$

$$K^{(2)} \triangleq E\{ [q^2\rho''(\varepsilon_1)(F'_2(\beta_{20}^T \mathbf{X}_1^{(2)}))^2 - q\rho'(\varepsilon_1)(F''_2(\beta_{20}^T \mathbf{X}_1^{(2)}))] \mathbf{X}_1^{(2)} \mathbf{X}_1^{(2)T} \}.$$

The form of  $\Sigma$  is the same as that in the following Proposition 3.3.

In order to get the asymptotic normality, we need the following important proposition.

PROPOSITION 3.3. *Under conditions (A1)–(A7), we have*

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \rho'(\varepsilon_i) \begin{pmatrix} pF_1'(\beta_{10}^T \mathbf{X}_i^{(1)}) \mathbf{X}_i^{(1)} \\ qF_2'(\beta_{20}^T \mathbf{X}_i^{(2)}) \mathbf{X}_i^{(2)} \end{pmatrix} \xrightarrow{\mathcal{D}} N(\mathbf{0}, \Sigma),$$

where

$$\begin{aligned} \Sigma &= \begin{pmatrix} \Sigma^{(1)} & \Sigma^{(1,2)} \\ \Sigma^{(1,2)^T} & \Sigma^{(2)} \end{pmatrix}, \\ \Sigma^{(1,2)} &\triangleq E[(pqF_1'(\beta_{10}^T \mathbf{X}_1^{(1)})F_2'(\beta_{20}^T \mathbf{X}_1^{(2)}))\mathbf{X}_1^{(1)}\mathbf{X}_1^{(2)^T} L], \\ \Sigma^{(1)} &\triangleq E[(pF_1'(\beta_{10}^T \mathbf{X}_1^{(1)}))^2 \mathbf{X}_1^{(1)}\mathbf{X}_1^{(1)^T} L], \\ \Sigma^{(2)} &\triangleq E[(qF_2'(\beta_{20}^T \mathbf{X}_1^{(2)}))^2 \mathbf{X}_1^{(2)}\mathbf{X}_1^{(2)^T} L]. \end{aligned}$$

Proof.  $\Theta$  is a bounded closed set, and the true values of parameters are the inner point of  $\Theta$ .

By the condition (A2), for any  $a$  and  $b$ , it follows that  $E(\rho(\tilde{\varepsilon} + a(I_{\{\delta=1\}} - p) + b(I_{\{\delta=2\}} - q) + A))$  achieves its minimum when  $A = 0$ , where  $\tilde{\varepsilon} = \varepsilon^{(1)}I_{\{\delta=1\}} + \varepsilon^{(2)}I_{\{\delta=2\}}$ . Namely,  $E(\rho'(\tilde{\varepsilon} + a(I_{\{\delta=1\}} - p) + b(I_{\{\delta=2\}} - q))) = 0$  when  $A = 0$ , where  $\tilde{\varepsilon} = \varepsilon^{(1)}I_{\{\delta=1\}} + \varepsilon^{(2)}I_{\{\delta=2\}}$ .

Let us put

$$\boldsymbol{\tau}_i = \frac{1}{\sqrt{N}} \rho'(\varepsilon_i) \begin{pmatrix} pF_1'(\beta_{10}^T \mathbf{X}_i^{(1)}) \mathbf{X}_i^{(1)} \\ qF_2'(\beta_{20}^T \mathbf{X}_i^{(2)}) \mathbf{X}_i^{(2)} \end{pmatrix}.$$

Then we have  $E(\boldsymbol{\tau}_i) = \mathbf{0}$ . Furthermore, we obtain  $E(\sum_{i=1}^N \boldsymbol{\tau}_i) = \mathbf{0}$ .

Next, because  $\{\mathbf{X}_i^\delta\}$  and  $\{\varepsilon_i\}$  ( $i = 1, 2, \dots, N$ ;  $\delta = 1, 2$ ) are i.i.d., we consider

$$\begin{aligned} \text{Var}(\boldsymbol{\tau}_i) &= E(\boldsymbol{\tau}_i \boldsymbol{\tau}_i^T) = \frac{1}{N} E \left\{ (\rho'(\varepsilon_i))^2 \begin{pmatrix} pF_1'(\beta_{10}^T \mathbf{X}_i^{(1)}) \mathbf{X}_i^{(1)} \\ qF_2'(\beta_{20}^T \mathbf{X}_i^{(2)}) \mathbf{X}_i^{(2)} \end{pmatrix} \right. \\ &\quad \left. \times \begin{pmatrix} pF_1'(\beta_{10}^T \mathbf{X}_i^{(1)}) \mathbf{X}_i^{(1)^T} & qF_2'(\beta_{20}^T \mathbf{X}_i^{(2)}) \mathbf{X}_i^{(2)^T} \end{pmatrix} \right\} \\ &= \frac{1}{N} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \end{aligned}$$

where

$$\begin{aligned} A_{11} &= p^2 E \left( (F_1'(\beta_{10}^T \mathbf{X}_1^{(1)}))^2 \mathbf{X}_1^{(1)} \mathbf{X}_1^{(1)^T} \right) L, \\ A_{12} &= pq E (F_1'(\beta_{10}^T \mathbf{X}_1^{(1)}) F_2'(\beta_{20}^T \mathbf{X}_1^{(2)}) \mathbf{X}_1^{(1)} \mathbf{X}_1^{(2)^T}) L, \\ A_{21} &= qp E (F_2'(\beta_{20}^T \mathbf{X}_1^{(2)}) F_1'(\beta_{10}^T \mathbf{X}_1^{(1)}) \mathbf{X}_1^{(2)} \mathbf{X}_1^{(1)^T}) L, \\ A_{22} &= q^2 E \left( (F_2'(\beta_{20}^T \mathbf{X}_1^{(2)}))^2 \mathbf{X}_1^{(2)} \mathbf{X}_1^{(2)^T} \right) L. \end{aligned}$$

So, we have

$$\text{Var}\left(\sum_{i=1}^N \tau_i\right) = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

By the Lévy central limit theorem and the law of large numbers, we have

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \rho'(\varepsilon_i) \begin{pmatrix} pF'_1(\beta_{10}^T \mathbf{X}_i^{(1)}) \mathbf{X}_i^{(1)} \\ qF'_2(\beta_{20}^T \mathbf{X}_i^{(2)}) \mathbf{X}_i^{(2)} \end{pmatrix} = \sum_{i=1}^N \tau_i \xrightarrow{\mathcal{D}} N(\mathbf{0}, \Sigma),$$

where  $\Sigma$  is defined as in Proposition 3.3. This completes the proof of Proposition 3.3. ■

**Proof of Theorem 3.2.** By Taylor’s expansion, we have

$$\frac{\partial S_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{\partial S_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} + \frac{\partial^2 S_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \cdot (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o(\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|),$$

i.e.

$$\begin{aligned} \begin{pmatrix} \frac{\partial S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_1} \\ \frac{\partial S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_2} \end{pmatrix} &= \begin{pmatrix} \frac{\partial S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_1} \\ \frac{\partial S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_2} \end{pmatrix} \Big|_{\substack{\boldsymbol{\beta}_1=\boldsymbol{\beta}_{10} \\ \boldsymbol{\beta}_2=\boldsymbol{\beta}_{20}}} \\ &+ \begin{pmatrix} \frac{\partial^2 S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} & \frac{\partial^2 S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} \\ \frac{\partial^2 S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_1^T} & \frac{\partial^2 S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} \end{pmatrix} \Big|_{\substack{\boldsymbol{\beta}_1=\boldsymbol{\beta}_{10} \\ \boldsymbol{\beta}_2=\boldsymbol{\beta}_{20}}} \cdot (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o(\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|). \end{aligned}$$

Because  $\hat{\boldsymbol{\beta}}_N$  is the minimizer of  $S_N(\boldsymbol{\beta})$  in  $\Theta$ , we have

$$(3.5) \quad 0 = \begin{pmatrix} \frac{\partial S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_1} \\ \frac{\partial S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_2} \end{pmatrix} \Big|_{\substack{\boldsymbol{\beta}_1=\boldsymbol{\beta}_{10} \\ \boldsymbol{\beta}_2=\boldsymbol{\beta}_{20}}} + \begin{pmatrix} \frac{\partial^2 S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} & \frac{\partial^2 S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} \\ \frac{\partial^2 S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_1^T} & \frac{\partial^2 S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} \end{pmatrix} \Big|_{\substack{\boldsymbol{\beta}_1=\boldsymbol{\beta}_{10} \\ \boldsymbol{\beta}_2=\boldsymbol{\beta}_{20}}} \cdot (\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0) + o(\|\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0\|).$$

Indeed, this can easily be obtained as follows:

$$\begin{aligned} \frac{\partial S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_1} \Big|_{\substack{\boldsymbol{\beta}_1=\boldsymbol{\beta}_{10} \\ \boldsymbol{\beta}_2=\boldsymbol{\beta}_{20}}} &= -\frac{p}{N} \sum_{i=1}^N \rho'(\varepsilon_i) F'_1(\beta_{10}^T \mathbf{X}_i^{(1)}) \mathbf{X}_i^{(1)}, \\ \frac{\partial S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_2} \Big|_{\substack{\boldsymbol{\beta}_1=\boldsymbol{\beta}_{10} \\ \boldsymbol{\beta}_2=\boldsymbol{\beta}_{20}}} &= -\frac{q}{N} \sum_{i=1}^N \rho'(\varepsilon_i) F'_2(\beta_{20}^T \mathbf{X}_i^{(2)}) \mathbf{X}_i^{(2)}, \end{aligned}$$

$$\frac{\partial^2 S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} \Big|_{\substack{\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10} \\ \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}}} = \frac{1}{N} \sum_{i=1}^N \{ [p^2 \rho''(\varepsilon_i) (F_1'(\boldsymbol{\beta}_{10}^T \mathbf{X}_i^{(1)}))^2 - p \rho'(\varepsilon_i) F_1''(\boldsymbol{\beta}_{10}^T \mathbf{X}_i^{(1)})] \mathbf{X}_i^{(1)} \mathbf{X}_i^{(1)T} \} \triangleq K_N^{(1)},$$

$$\frac{\partial^2 S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} \Big|_{\substack{\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10} \\ \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}}} = \frac{1}{N} \sum_{i=1}^N \{ pq [\rho''(\varepsilon_i) F_1'(\boldsymbol{\beta}_{10}^T \mathbf{X}_i^{(1)}) F_2'(\boldsymbol{\beta}_{20}^T \mathbf{X}_i^{(2)})] \mathbf{X}_i^{(1)} \mathbf{X}_i^{(2)T} \} \\ \triangleq K_N^{(1,2)},$$

$$\frac{\partial^2 S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_1^T} \Big|_{\substack{\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10} \\ \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}}} = \frac{1}{N} \sum_{i=1}^N \{ pq [\rho''(\varepsilon_i) F_2'(\boldsymbol{\beta}_{20}^T \mathbf{X}_i^{(2)}) F_1'(\boldsymbol{\beta}_{10}^T \mathbf{X}_i^{(1)})] \mathbf{X}_i^{(2)} \mathbf{X}_i^{(1)T} \} \\ \triangleq K_N^{(1,2)T},$$

$$\frac{\partial^2 S_N(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} \Big|_{\substack{\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10} \\ \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20}}} = \frac{1}{N} \sum_{i=1}^N \{ [q^2 \rho''(\varepsilon_i) (F_2'(\boldsymbol{\beta}_{20}^T \mathbf{X}_i^{(2)}))^2 - q \rho'(\varepsilon_i) F_2''(\boldsymbol{\beta}_{20}^T \mathbf{X}_i^{(2)})] \mathbf{X}_i^{(2)} \mathbf{X}_i^{(2)T} \} \triangleq K_N^{(2)}.$$

Then the formula (3.5) can be rewritten as

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \rho'(\varepsilon_i) \begin{pmatrix} p F_1'(\boldsymbol{\beta}_{10}^T \mathbf{X}_i^{(1)}) \mathbf{X}_i^{(1)} \\ q F_2'(\boldsymbol{\beta}_{20}^T \mathbf{X}_i^{(2)}) \mathbf{X}_i^{(2)} \end{pmatrix} \\ = \sqrt{N} \begin{pmatrix} K_N^{(1)} & K_N^{(1,2)} \\ K_N^{(1,2)T} & K_N^{(2)} \end{pmatrix} (\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0) + o(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0) \sqrt{N}.$$

Let us define

$$K_N \triangleq \begin{pmatrix} K_N^{(1)} & K_N^{(1,2)} \\ K_N^{(1,2)T} & K_N^{(2)} \end{pmatrix}.$$

Because  $\{X_i^\delta\}$  and  $\{\varepsilon_i\}$  ( $i = 1, 2, \dots, N$ ) are i.i.d., using the law of large numbers, we have

$$\lim_{N \rightarrow \infty} K_N = K \triangleq \begin{pmatrix} K^{(1)} & K^{(1,2)} \\ K^{(1,2)T} & K^{(2)} \end{pmatrix},$$

where  $\boldsymbol{\beta}_0, \hat{\boldsymbol{\beta}}_N, K^{(1,2)}, K^{(1)}$  and  $K^{(2)}$  are defined as in Theorem 3.2. So by Proposition 3.3 and Slutsky's theorem, we obtain

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, K^{-1} \Sigma K^{-1}),$$

the form of  $\Sigma$  being the same as in Proposition 3.3. This completes the proof of Theorem 3.2. ■

#### 4. A SIMULATION STUDY

In this section, to indicate the reasonableness of the M-estimation for the mixed-type generalized linear model, some examples are presented. The response variable  $Y$  comes from the formula (2.1), i.e.,

$$E(Y|\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = pF_1(\boldsymbol{\beta}_{10}^T \mathbf{X}^{(1)}) + qF_2(\boldsymbol{\beta}_{20}^T \mathbf{X}^{(2)}).$$

Let  $\Pr\{\delta = 1\} = p = 0.3$  and  $0.8$ , respectively. Let  $F_1(\cdot)$  be the general linear model

$$F_1(\boldsymbol{\beta}_1^T \mathbf{X}) = \boldsymbol{\beta}_1^T \mathbf{X},$$

and  $F_2(\cdot)$  be the Poisson model

$$F_2(\boldsymbol{\beta}_2^T \mathbf{X}) = \exp(\boldsymbol{\beta}_2^T \mathbf{X}).$$

Assume that  $\mathbf{X}_i^{(1)}$  ( $i = 1, 2, \dots, N$ ) is a  $d_1$ -dimensional random vector, with mean zero and covariance a unit matrix, each component of  $\mathbf{X}_i^{(1)}$  ( $i = 1, 2, \dots, N$ ) being independent. Moreover, assume that  $\mathbf{X}_i^{(2)}$  ( $i = 1, 2, \dots, N$ ) is a  $d_2$ -dimensional random vector, with mean zero and covariance a unit matrix, each component of  $\mathbf{X}_i^{(2)}$  ( $i = 1, 2, \dots, N$ ) being independent.  $\boldsymbol{\beta}_1$  is a  $d_1$ -dimensional vector and its true value is  $\boldsymbol{\beta}_{10} = (-1, -2)^T$  with  $d_1 = 2$ , and  $\boldsymbol{\beta}_2$  is a  $d_2$ -dimensional vector and its true value is  $\boldsymbol{\beta}_{20} = (0.5, 0.5)^T$  with  $d_2 = 2$ . So we can get a whole data  $\{Y_i, \mathbf{X}_i^{(1)}, \mathbf{X}_i^{(2)}; i = 1, 2, \dots, N\}$  by our previously proposed model. For the general linear model, the error term is assumed to be normal distribution, but in the case of Poisson distribution, the simulated value of the response is generated from Poisson distribution. In the simulation we take  $N = 100$ .

**EXAMPLE 4.1** (Least absolute deviations). In this example, the function of M-estimation is taken as the least absolute, and denote its results as least absolute deviations (LAD).

Let us put  $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1^{(1)}, \hat{\beta}_2^{(1)}, \hat{\beta}_1^{(2)}, \hat{\beta}_2^{(2)})^T$ . To measure the performance of the estimator, we take the mean square errors (MSE) of  $\hat{\boldsymbol{\beta}}$  as

$$\text{MSE}(\hat{\boldsymbol{\beta}}) = E(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2),$$

and the bias of estimation (BE) of  $\hat{\boldsymbol{\beta}}$  as

$$\text{BE}(\hat{\boldsymbol{\beta}}) = \|E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}_0\|.$$

Summary statistics are computed based on 100, 300 and 600 repetitions, respectively. Simulation results are exhibited in Table 1.

TABLE 1. Least absolute deviations,  $\beta_0 = (-1, -2, 0.5, 0.5)^T$ .

$(N, p)$	Repetitions	Estimates ( $\hat{\beta}$ )				MSE( $\hat{\beta}$ )	BE( $\hat{\beta}$ )
		$\hat{\beta}_1^{(1)}$	$\hat{\beta}_2^{(1)}$	$\hat{\beta}_1^{(2)}$	$\hat{\beta}_2^{(2)}$		
(100, 0.3)	100	-0.9949	-2.0028	0.5322	0.5294	0.0051	0.0439
	300	-0.9975	-2.0023	0.5310	0.5314	0.0055	0.0442
	600	-0.9969	-2.0039	0.5299	0.5286	0.0053	0.0416
(100, 0.8)	100	-0.9809	-2.0155	0.5386	0.5329	0.0111	0.0563
	300	-0.9836	-2.0185	0.5366	0.5356	0.0117	0.0568
	600	-0.9814	-2.0212	0.5371	0.5321	0.0117	0.0566

EXAMPLE 4.2 (Ordinary least squares). In this example, the function of M-estimation is taken as the least square, and denote its results as ordinary least squares (OLS). The other assumptions and conditions that we considered are the same as in Example 4.1. Results of simulation are given in Table 2.

TABLE 2. Ordinary least squares,  $\beta_0 = (-1, -2, 0.5, 0.5)^T$ .

$(N, p)$	Repetitions	Estimates ( $\hat{\beta}$ )				MSE( $\hat{\beta}$ )	BE( $\hat{\beta}$ )
		$\hat{\beta}_1^{(1)}$	$\hat{\beta}_2^{(1)}$	$\hat{\beta}_1^{(2)}$	$\hat{\beta}_2^{(2)}$		
(100, 0.3)	100	-0.9652	-2.0120	0.4908	0.4687	0.6896	0.0493
	300	-1.0034	-2.0091	0.4893	0.4851	0.5393	0.0208
	600	-0.9730	-1.9942	0.4900	0.4853	0.5067	0.0329
(100, 0.8)	100	-0.9876	-2.0035	0.4430	0.4054	0.3343	0.1112
	300	-1.0009	-2.0033	0.4456	0.4255	0.3074	0.0923
	600	-0.9906	-1.9980	0.4385	0.4275	0.2783	0.0956

Tables 1 and 2 show that the method based on our proposed model performs quite well. It is also observed that the MSE of LAD performs better than OLS in the simulations. This relative large size occurs because the variance of Poisson model is large. The results indicate that the M-estimation of the mixed-type generalized linear model is a favorable method.

### 5. CONCLUSION

To investigate the features of the individual from the mixed-type model, a novel model, named the mixed-type generalized linear model, was proposed in this paper. The M-estimation of the unknown parameter of this model was analyzed. Furthermore, the consistency and asymptotic normality of the M-estimation for the mixed-type generalized linear model were obtained with some assumptions. The simulation study clearly showed that the proposed model has good performance. The mixed-type GLM could be applied in many different kinds of areas, such as biomedical statistics and social economical data statistics analysis. So the proposed model is very extensive and meaningful in real application area.

## REFERENCES

- [1] Z. D. Bai, C. R. Rao, and Y. Wu, *M-estimation of multivariate linear regression parameters under a convex discrepancy function*, *Statist. Sinica* 2 (1992), pp. 237–254.
- [2] Z. D. Bai, C. R. Rao, and Y. Wu, *M-estimation of multivariate linear regression by minimizing the difference of two convex functions*, *Handbook of Statist.*, Vol. 15, 1997.
- [3] C. C. Heyde, *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*, Springer, New York 1997.
- [4] Y. J. Huang and L. X. Song, *M-estimator of a generalized linear model with measurement errors*, *Comm. Statist. Theory Methods* 40 (2011), pp. 532–546.
- [5] P. J. Huber, *Robust estimation of a location parameter*, *Ann. Math. Statist.* 35 (1964), pp. 73–101.
- [6] P. J. Huber, *The behavior of maximum likelihood estimates under nonstandard conditions*, in: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, Berkeley 1967, pp. 221–233.
- [7] P. J. Huber, *Robust regression: Asymptotics, conjectures and Monte Carlo*, *Ann. Statist.* 5 (1973), pp. 799–821.
- [8] P. J. Huber, *Robust Statistics*, Wiley, New York 1981.
- [9] P. J. Huber, *Robust Statistical Procedures*, Society for Industrial Mathematics, Philadelphia 1989.
- [10] L. Mancini, E. Ronchetti, and F. Trojani, *Optimal conditionally unbiased bounded-influence inference in dynamic location and scale models*, *J. Amer. Statist. Assoc.* 105 (2005), pp. 628–641.
- [11] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, second edition, Chapman and Hall, London 1989.
- [12] N. Muler and V. J. Yohai, *Robust estimates for ARCH processes*, *J. Time Series Anal.* 23 (2002), pp. 79–109.
- [13] J. A. Nelder and R. W. M. Wedderburn, *Generalized linear models*, *J. Roy. Statist. Soc. Ser. B* 135 (1972), pp. 370–384.
- [14] J. S. Preisser and B. F. Qaqish, *Robust regression for clustered data with applications to binary regression*, *Biometrics* 55 (1999), pp. 574–579.
- [15] S. K. Sinha, *Robust analysis of generalized linear mixed models*, *J. Amer. Statist. Assoc.* 99 (2004), pp. 451–460.
- [16] V. J. Yohai and R. A. Maronna, *Asymptotic behavior of M-estimators for the linear model*, *Ann. Statist.* 2 (1979), pp. 258–268.

Ying Dong  
Faculty of Science  
Dalian Minzu University  
Dalian, 116600, P.R. China  
E-mail: dongying@dlnu.edu.cn

Lixin Song  
School of Mathematical Sciences  
Dalian University of Technology  
Dalian, 116023, P.R. China  
E-mail: lxsong@dlut.edu.cn

Mingqiu Wang  
School of Mathematical Sciences  
Qufu Normal University  
Shandong, Qufu, 273165, P.R. China  
E-mail: wmq0829@gmail.com

Muhammad Amin  
Nuclear Institute for Food and Agriculture  
446, Peshawar, Pakistan  
E-mail: aminkanju@gmail.com

Received on 4.9.2016;  
revised version on 4.1.2017