Quasi-likelihood Estimation for GLM with Random Scales

Jye-Chyi Lu*, Di Chen* and Weixing Zhou+

*: School of Industrial and Systems Engineering, Georgia Institute of Technology,
Atlanta, GA 30332-0205, U.S.A.

+: Department of Statistics, North Carolina State University,
Raleigh, NC 27695-8203, U.S.A.

ABSTRACT

This paper uses random scales similar to random effects used in the generalized linear mixed models to describe "inter-location" population variation in variance components for modeling complicated data obtained from applications such as antenna manufacturing. Our distribution studies lead to a complicated integrated extended quasi-likelihood (IEQL) for parameter estimations and large sample inference derivations. Laplace's expansion and several approximation methods are employed to simplify the IEQL estimation procedures. Asymptotic properties of the approximate IEQL estimates are derived for general structures of the covariance matrix of random scales. Focusing on a few special covariance structures in simpler forms, the authors further simplify IEQL estimates such that typically used software tools such as weighted regression can perform the estimates easily. Moreover, these special cases allow us to derive interesting asymptotic results in much more compact expressions. Finally, numerical simulation results show that IEQL estimates perform very well in several special cases studied.

KEY WORDS: Generalized Linear Model; Integrated Extended Quasi-Likelihood; Random Scale; Multi-Normality; Consistent Estimate; Unbiasedness.

AMS Subject Classification: Primary 62J12; Secondary 62F10, 62F12.

1. Introduction

Recently, generalized linear models (GLM) (Nelder & Wedderburn, 1972; McCullagh & Nelder, 1989) have been extended in many directions to model various data types difficult to handle with traditional models. For example, linear random coefficient models (Longford, 1993), nonlinear random coefficient models (Chen, Lu, Huo and Ming, 2001) and hierarchical nonlinear models (Davidian and Giltinan, 1995 page 98) are useful in describing time-sequence repeated measurements frequently encountered in biological or medical research. Lindsey (1993) used variance component models with certain correlation structure on errors for modeling dependent repeated measurements. See Appendix A for a brief review of the GLM and its extension to quasi-likelihood (QL; Wedderburn, 1974) and extended quasi-likelihood (EQL; Nelder & Pregibon (1987) and Davidian & Caroll (1988)) which serve as the foundation of our proposal of the integrated EQL (IEQL) approach.

Most random coefficient models assume that the mean regression coefficients are normally distributed with the same mean parameters across all subjects and constant variance. These mean parameters represent the underlying common characteristics among subjects. The general linear mixed models (GLMM; Harville, 1977) uses random effects to describe the "inter-subject" population variation in mean regression function and to handle more general error distributions.

In many situations there exists population variation in variance components. For example, in a study of monitoring spatial-temporal data, Host, Omre & Switzer (1995) stated that the random variation of responses could be partitioned into three separated fields, namely, mean field, scale field and residual error field as follows: $Y(x) = M(x) + S(x)\varepsilon(x)$, where x is the location. Our study focuses on the case where S is random. In a real-life example (see Lu, Zhou, Chen, Hughes-Oliver and Ghosh, 2002 for details) of developing process fault detection procedures to monitor spatial antenna data at Nortel, we have documented variations in the scale component $S(x)$ from 20 sets of antenna data. Figure 1 presents one set of the antenna data collected at 181 $\times$ 181 locations. Although the mean and scale components did not have "perfect" relationship, the mean parameters in the distribution of the random scale are related to the mean regression function systematically similar to the GLM model. See Figure 2 from a "de-noised" log-variance plot for an example.

(please place Figures 1 and 2 here.)

Just like many studies of spatial data, in this article the modeling of data similar to Figure 1 will be restricted to a single data set, but not their replicates (i.e., no third index in the following formula). Motivated from the physical background of the antenna behavior, patterns of the mean function in one coordinate (the elevation with the index j) should be very similar at every cut of the other coordinate (the azimuth with the index i) except a shift. See Figure 1 for an example. Thus we propose the following model with a random scale component:

$$\left\{ \begin{array}{rcl} Y_{ij} & = & \mu_j + v^{1/2}(\mu_j) w_j^{1/2} \varepsilon_{ij}, \quad i = 1, 2, \ldots, n; \quad j = 1, 2, \ldots, m, \\ \log(\boldsymbol{w}) & = & (\log(w_1), \ldots, \log(w_m))^\top \sim N_m(0, \boldsymbol{\Sigma}(\boldsymbol{\theta})), \end{array} \right. \qquad (1)$$

where Y_{ij} is response (e.g., amplitude representing the quality of the signal sent out from antenna). The mean vector $\boldsymbol{\mu}$ can also be expressed by a function $\boldsymbol{\mu}(\boldsymbol{\beta}) = \boldsymbol{f}(\boldsymbol{A}\boldsymbol{\beta})$ with a p-dimension vector of unknown parameters $\boldsymbol{\beta}$ and a $m \times p$ constant matrix $\boldsymbol{A}$. See Appendix B for details. Because of the limited resources in our data collection process, in the motivating example, the numbers n and m are both fixed at 181. Both of them could be made much larger for supporting the assumption in deriving the asymptotic results. See Theorems 2-5 for details in the cases where both n and m go to infinity or only n goes to infinity while holding m fixed.

We use the same random scale $\sigma_j = v^{1/2}(\mu_j)w_j^{1/2}$ for data at all azimuth location with $i = 1, 2, \ldots, n$. The power $1/2$ on v and w_j is for computational convenience. Adding a parameter ϕ to $\sigma_j{}^* = \phi^{1/2}v^{1/2}(\mu_j)w_j^{1/2}$ is useful to break the linkage between the mean and variance of the random scales. This implies that $\log(\boldsymbol{w})$ has a non-zero mean $\log(\phi)$ and variance-covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. Lemma 2 shows that $\tau_j = \log[\sum_{i=1}^{n}(y_{ij} - \bar{y}_{.j})^2/nv(\bar{y}_{.j})]$ converges with probability one to $\log(w_j)$, $j = 1, \ldots, m$. Thus one can easily check if the mean of τ_j's is equal to zero for deciding if ϕ should be equal to one. Our experience indicates that the IEQL estimation procedure of $\boldsymbol{\theta}$ is not affected with ϕ, and the estimation of $\log(\phi)$ can be easily obtained from $\mathbf{1}_m{}^\top\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\boldsymbol{\tau}/[\mathbf{1}_m{}^\top\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\mathbf{1}_m]$, where $\mathbf{1}_m = (1, \ldots, 1)^\top$ is an m-dimension vector, $\hat{\boldsymbol{\theta}}$ is an estimate of $\boldsymbol{\theta}$ based on the approximated IEQL estimating equations derived in Section 4.1, and $\boldsymbol{\tau}$ is a vector of τ_j's. The asymptotic properties of the estimate of ϕ can therefore be obtained via the asymptotic results of $\hat{\boldsymbol{\theta}}$. Because our focus in this article is the variance-covariance parameters $\boldsymbol{\theta}$ of the random scales, the parameter ϕ is not added to Model (1) for keeping the presentation brief.

Although more general forms of transformation of $\boldsymbol{w}$ could be used for normality, to keep our presentation brief, for a positive scale, we use the log-normal distribution of $\log(\boldsymbol{w})$. The matrix $\boldsymbol{\Sigma}$ could include certain spatial structure or be very general without any restricted structure. This article uses a k-dimension parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^\top$ to model $\boldsymbol{\Sigma}$, where k is a positive integer and $k \leq m$. Note that, if $k > m$, then there is no unique estimate of $\boldsymbol{\theta}$. The random error term $w_j{}^{1/2}\varepsilon_{ij}$ in (1) has a complicated covariance matrix. For simplicity of presentation, we assume that conditioning on $\boldsymbol{w}$, the errors ε_{ij} are independent and identically distributed (*iid*) with zero mean and unit variance.

Random scale $S(x)$ has been considered in other content such as Bayesian models (e.g., Carlin and Poison, 1991) and Robust models (e.g., Lange, Little and Taylor, 1989). In particular, Lange, *et al.* (1991) studied the t distribution in several applications with the interpretation that t distribution for observations can be derived from a conditional normal with u in the scale parameter (ξ/u) having a chi-square distribution. In our study, the unconditional distribution of the observations does not have a closed-form expression like the t-distribution and is very complicated involving integrations of m- dimensional normal density. Hence approximation methods are needed to get sensible estimates. This makes the study of asymptotic properties of the estimates worthwhile.

Section 2 presents the approach of using an IEQL for representing the data generated from the model (1). Note that our IEQL approach focuses on the random scale component and is not aimed to be robust against outliers. Section 3 provides details of IEQL estimation procedures and also discusses properties of the estimator for mean parameters. Section 4 discusses techniques to solve IEQL estimating equations for variance-covariance parameters and investigates asymptotic properties of the IEQL estimators. Section 5 reports various simulation studies conducted with different data sizes and variance-covariance structures of the random-scale vector. Section 6 concludes this article.

2. The Integrated Extended Quasi-Likelihood (IEQL)

Denoted by $\boldsymbol{y} = (y_{11}, \ldots, y_{1m}, \ldots, y_{n1}, \ldots, y_{nm})^\top$, the observations of repeat measurements. To simplify the presentation, let $z_j = \log(w_j)$ and $\boldsymbol{z} = (z_1, \ldots, z_m)^\top$. Let us generalize the extended quasi-likelihood (EQL) proposed by Nelder & Pregibon (1987) to handle the random scale components. Note that for given $\boldsymbol{z}$, the conditional EQL of data $\boldsymbol{y}$ is (ignoring the constant 2π)

$$q^*(\boldsymbol{\mu}|\boldsymbol{z}) = \exp\left\{-\frac{1}{2}\sum_{j=1}^{m}\sum_{i=1}^{n}\left(\log[v(y_{ij})e^{z_j}] - 2e^{-z_j}\int_{y_{ij}}^{\mu_j}\frac{y_{ij}-x}{v(x)}dx\right)\right\}.$$

Integrating $q^*(\boldsymbol{\mu}|\boldsymbol{z})$ over the normal distribution of $\boldsymbol{z}$ leads to the following integrated EQL (IEQL):

$$Q(\boldsymbol{\mu}, \boldsymbol{\theta}) \quad \propto \quad |\boldsymbol{\Sigma}(\boldsymbol{\theta})|^{-1/2}\int_{R^m} q^*(\boldsymbol{\mu}|\boldsymbol{z}) \cdot \exp\left(-\frac{1}{2}\boldsymbol{z}^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{z}\right) d\boldsymbol{z}. \qquad (2)$$

To derive the IEQL estimating equations for the unknown parameters, and discuss the asymptotic properties of the IEQL estimate, the following regularity conditions are necessary.

(R1) Set $\{\mu : v(\mu) = 0\}$ is zero-Lebesgue measure.

(R2) $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is full rank for every $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^k$. The parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^k$ is compact without isolated points and contains a neighborhood of true parameter, say $\boldsymbol{\theta}_0$.

(R3) Let $\sigma_{ij}(\boldsymbol{\theta})$ be the ijth element of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. Assume that $\partial\sigma_{ij}(\boldsymbol{\theta})/\partial\theta_l$ exists for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $i, j = 1, \ldots, m$, $l = 1, \ldots, k$, and there exists $N_\sigma > 0$ such that, for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$|\sigma_{ij}(\boldsymbol{\theta})| < N_\sigma, \quad \text{and} \quad |\frac{\partial\sigma_{ij}(\boldsymbol{\theta})}{\partial\theta_l}| < N_\sigma, \quad i, j = 1, \ldots, m, \quad l = 1, \ldots, k.$$

(R4) For given $\boldsymbol{z}$, conditionally, $\mathrm{E}(\varepsilon_j^4|\boldsymbol{z}) < \infty$, $j = 1, \ldots, m$.

Regularity conditions (R1) and (R2) ensure that the IEQL estimating equations are always well defined. (R3) and (R4) ensure the basic convergence of random variables and the exchangeability between limit and expectation.

To simplify the notations, denoted by

$$\begin{aligned} d_j(\mu_j) &= -2\sum_{i=1}^{n}\int_{y_{ij}}^{\mu_j} \frac{y_{ij}-x}{v(x)}dx, \quad j = 1, \ldots, m, \quad \text{and} & (3) \\ k(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta}) &= \sum_{j=1}^{m} d_j(\mu_j)e^{-z_j} + \sum_{j=1}^{m}\sum_{i=1}^{n}\log[v(y_{ij})] + n\sum_{j=1}^{m} z_j + \boldsymbol{z}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{z}, & (4) \end{aligned}$$

where $\boldsymbol{d}(\boldsymbol{\mu}) = (d_1(\mu_1), \ldots, d_m(\mu_m))^\top$. Then, the IEQL (2) is written as:

$$Q(\boldsymbol{\mu}, \boldsymbol{\theta}) \propto |\boldsymbol{\Sigma}(\boldsymbol{\theta})|^{-1/2} \int_{I\!R^m} \exp\left(-\frac{k(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})}{2}\right) d\boldsymbol{z}.$$

The above integration is difficult and time-consuming during the estimation process. We apply the Laplace's approach to approximate the IEQL (2). This technique was used by Luke & Joseph (1990), Wolfinger (1993), Breslow & Clayton (1993), Edward (1996) and Shun (1997). This approach allows us to approximate the integral $\int_{R^m} \exp(-k(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})/2)d\boldsymbol{z}$ by expanding $k(\boldsymbol{Y}, \boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})$ in quadratic terms around its maximum point $\boldsymbol{z}^*$ before the integration.

The following present the first and second derivatives, respectively, of $k(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})$ with respect to the elements in the random scales $\boldsymbol{z} = (z_1, z_2, \ldots, z_m)^\top$:

$$\begin{aligned} \boldsymbol{k}'_z(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta}) &= -\boldsymbol{d}(\boldsymbol{\mu})e^{-\boldsymbol{z}} + n\mathbf{1}_m + 2\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{z} \quad \text{and} & (5) \\ \boldsymbol{k}''_z(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta}) &= diag[\boldsymbol{d}(\boldsymbol{\mu})e^{-\boldsymbol{z}}] + 2\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}, & (6) \end{aligned}$$

where $diag[\boldsymbol{d}(\boldsymbol{\mu})e^{-\boldsymbol{z}}]$ is the diagonal matrix with the following elements:

$$\boldsymbol{d}(\boldsymbol{\mu})e^{-\boldsymbol{z}} = (d_1(\mu_1)e^{-z_1}, \ldots, d_m(\mu_m)e^{-z_m})^\top. \tag{7}$$

Let $\boldsymbol{z}^*$ be the point maximizing the value of $k(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})$ given by $k'_z(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}^*, \boldsymbol{\theta}) = 0$. That is, $\boldsymbol{z}^* = \boldsymbol{\tau}(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{\theta})$ is an implicit function determined by the equation

$$-\boldsymbol{d}(\boldsymbol{\mu})e^{-\boldsymbol{z}} + n\mathbf{1}_m + 2\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{z} = 0. \tag{8}$$

Thus, the second-order Taylor expansion of $k(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})$ around $\boldsymbol{z}^*$ leads to

$$k(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta}) \approx k(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}^*, \boldsymbol{\theta}) + \frac{1}{2}(\boldsymbol{z} - \boldsymbol{z}^*)^\top k''_z(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}^*, \boldsymbol{\theta})(\boldsymbol{z} - \boldsymbol{z}^*).$$

Then, the Laplace's approach gives

$$\int_{I\!R^m} \exp\left(-\frac{k(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})}{2}\right) d\boldsymbol{z}$$

$$\approx \int_{I\!R^m} \exp\left(-\frac{1}{2}k(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}^*, \boldsymbol{\theta}) - \frac{1}{2}(\boldsymbol{z}-\boldsymbol{z}^*)^\top k_z''(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}^*, \boldsymbol{\theta})(\boldsymbol{z}-\boldsymbol{z}^*)\right) d\boldsymbol{z}$$

$$= \exp\left(-\frac{1}{2}k(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}^*, \boldsymbol{\theta})\right) \cdot (2\pi)^{m/2} \cdot \frac{1}{2}|k_z''(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}^*, \boldsymbol{\theta})|^{-1/2}.$$

Thus, an approximate logarithm IEQL function $L(\boldsymbol{\mu}, \boldsymbol{\theta}) = \log[Q(\boldsymbol{\mu}, \boldsymbol{\theta})]$ can be obtained as

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\theta}) &\propto -\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - k(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}^*, \boldsymbol{\theta}) - \log|k_z''(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}^*, \boldsymbol{\theta})| \\ &= -\log|\boldsymbol{\Sigma}(\boldsymbol{\theta}) diag[\boldsymbol{d}(\boldsymbol{\mu})e^{-\boldsymbol{z}^*}] + 2\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}| - k(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}^*, \boldsymbol{\theta}) \\ &= -\log|2\boldsymbol{I}_m + diag[\boldsymbol{d}(\boldsymbol{\mu})e^{-\boldsymbol{z}^*}]\boldsymbol{\Sigma}(\boldsymbol{\theta})| - k(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}^*, \boldsymbol{\theta}), \end{aligned} \tag{9}$$

where $\boldsymbol{I}_m$ is an $m \times m$ unit matrix, and the second equality is from (6).

In the next section, we derive the estimates of parameters $\boldsymbol{\mu}$ and $\boldsymbol{\theta}$ based on the IEQL (9). For convenience, without any confusion, we will omit the star symbol from $\boldsymbol{z}^*$ in the rest of the paper when we mention $\boldsymbol{z}$ being the point maximizing the value of $k(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})$.

3. The Maximum IEQL Estimation of Model Parameters

3.1 Mean Parameters

Denoted by $l(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})$ the right side of (9). The IEQL estimator of parameter vector $\boldsymbol{\mu}$ is obtained by solving equations $\partial l(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})/\partial\boldsymbol{\mu} = 0$. Note that $l(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})$ and $\boldsymbol{z}(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{\theta})$ depend on $\boldsymbol{\mu}$ through the $\boldsymbol{d}$ function only. Thus,

$$\begin{aligned} \frac{\partial l(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})}{\partial\boldsymbol{\mu}} &= \frac{\partial l(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})}{\partial\boldsymbol{d}} \cdot \frac{\partial\boldsymbol{d}}{\partial\boldsymbol{\mu}} + \frac{\partial l(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})}{\partial\boldsymbol{z}} \cdot \frac{\partial\boldsymbol{z}}{\partial\boldsymbol{d}} \cdot \frac{\partial\boldsymbol{d}}{\partial\boldsymbol{\mu}} \\ &= \left(\frac{\partial l(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})}{\partial\boldsymbol{d}} + \frac{\partial l(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})}{\partial\boldsymbol{z}} \cdot \frac{\partial\boldsymbol{z}}{\partial\boldsymbol{d}}\right) \frac{\partial\boldsymbol{d}}{\partial\boldsymbol{\mu}}. \end{aligned}$$

Under the regularity conditions (R2) and (R3), using Lemma 1 in the next subsection, we can show that the matrix

$$\boldsymbol{M} = \frac{\partial l(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})}{\partial\boldsymbol{d}} + \frac{\partial l(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})}{\partial\boldsymbol{z}} \cdot \frac{\partial\boldsymbol{z}}{\partial\boldsymbol{d}}$$

is of full-rank. Therefore, the IEQL estimating equation $\partial l(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})/\partial\boldsymbol{\mu} = 0$ can be equivalently written as $\partial\boldsymbol{d}/\partial\boldsymbol{\mu} = 0$. It follows from (3) that the estimating equations for mean parameter μ_j

are:

$$\frac{\partial d_j}{\partial \mu_j} = -2\sum_{i=1}^{n} \frac{y_{ij} - \mu_j}{v(\mu_j)} = 0, \quad j = 1, \ldots, m. \tag{10}$$

Thus, the IEQL estimates $\hat{\mu}_j = \sum_{i=1}^{n} y_{ij}/n = \bar{y}_{.j}$, $j = 1, \ldots, m$, are the regular averages as expected.

According to the model assumption (1), we have

$$\mathrm{E}(\hat{\mu}_j|\boldsymbol{z}) = \mu_j + v^{1/2}(\mu_j)e^{z_j/2}\mathrm{E}(\bar{\varepsilon}_{.j}|\boldsymbol{z}) = \mu_j.$$

Because $\mathrm{E}(\hat{\mu}_j) = \mathrm{E}[\mathrm{E}(\hat{\mu}_j|\boldsymbol{z})] = \mu_j$, $j = 1, \ldots, m$, we know that $\hat{\mu}_j$ is unbiased.

Remark: Because we will need to calculate $\partial l(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})/\partial \boldsymbol{\theta}$, $\partial l(\boldsymbol{d}(\boldsymbol{\mu}), \boldsymbol{z}, \boldsymbol{\theta})/\partial \boldsymbol{z}$, and $\partial \boldsymbol{z}/\partial \boldsymbol{\theta}$ in details when we derive the estimating equations for the scale-variance parameter $\boldsymbol{\theta}$ in the next subsection for the random scales, and their algebraic manipulations are almost the same as in the calculation of $\boldsymbol{M}$, we skip the details of their calculation here. □

For Model (1) using the IEQL method, the mean parameter $\boldsymbol{\mu}$ can be estimated well without any information about variance-covariance parameter $\boldsymbol{\theta}$. The following theorem presents properties of the IEQL estimate $\hat{\boldsymbol{\mu}}$. Note that this theorem works for every $j = 1, 2, \ldots, m$.

Theorem 1. *Under the regularity conditions (R2) and (R3), for the random-scale GLM (1), the IEQL estimates of model mean parameters are:* $\hat{\mu}_j = \sum_{i=1}^{n} y_{ij}/n = \bar{y}_{.j}$, $j = 1, \ldots, m$, and,

(i) $\hat{\mu}_j$ *is unbiased,* $j = 1, 2, \ldots, m$;

(ii) $\hat{\mu}_j$ *is a consistent estimate, as* $n \to \infty$, *for every* $j = 1, 2, \ldots, m$ *with any size of* m;

(iii) *let* $\boldsymbol{\mu}_l = (\mu_{j1}, \mu_{j2}, \ldots, \mu_{jl})$ *be a* l*-dimension sub-vector of* $\boldsymbol{\mu}$. *As* $n \to \infty$, *the asymptotic density of* $\sqrt{n}(\hat{\boldsymbol{\mu}}_l - \boldsymbol{\mu}_l)$ *is given by*

$$f(\boldsymbol{x}_l) = (2\pi)^{-l} \int_{R^l} |\boldsymbol{\Sigma}_l(\boldsymbol{\theta})\boldsymbol{\Gamma}(\boldsymbol{z}_l)|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}_l^\top \boldsymbol{\Gamma}^{-1}(\boldsymbol{z}_l)\boldsymbol{x}_l - \frac{1}{2}\boldsymbol{z}_l^\top \boldsymbol{\Sigma}_l(\boldsymbol{\theta})^{-1}\boldsymbol{z}_l\right) d\boldsymbol{z}_l, \tag{11}$$

where $\boldsymbol{x}_l = (x_{j1}, x_{j2}, \ldots, x_{jl})$, $\boldsymbol{z}_l = (z_{j1}, z_{j2}, \ldots, z_{jl})$, $\boldsymbol{\Gamma}(\boldsymbol{z}_l) = diag\,(v(\mu_{j1})e^{z_{j1}}, \ldots, v(\mu_{jl})e^{z_{jl}})$, *and* $\boldsymbol{\Sigma}_l(\boldsymbol{\theta})$ *consists of the components of* $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ *corresponding to* $(z_{j1}, z_{j2}, \ldots, z_{jl})$.

Proof: We only need to show (ii) and (iii).

From model assumption (1), we have $\hat{\mu}_j = \bar{y}_{.j} = \mu_j + v^{1/2}(\mu_j)e^{z_j/2}\bar{\varepsilon}_{.j}$, $j = 1, \ldots, m$. It follows that $(\hat{\mu}_j - \mu_j) = v^{1/2}(\mu_j)e^{z_j/2}\bar{\varepsilon}_{.j}$. Assumption (1) implies that, conditioning on z_j, ε_{ij}, for $i = 1, \ldots, n$, are *iid* with $\mathrm{E}(\varepsilon_{ij}|z_j) = 0$ and $\mathrm{var}(\varepsilon_{ij}|z_j) = 1$. Applying Lindeberg-Levy central limit theorem, we know that, conditioning on $\boldsymbol{z}$ and any size of m, $\sqrt{n}\bar{\varepsilon}_{.j} \xrightarrow{d} N(0,1)$ as $n \to \infty$. Therefore, conditionally, $\sqrt{n}(\hat{\mu}_j - \mu_j) \xrightarrow{d} N[0, v(\mu_j)e^{z_j}]$ as $n \to \infty$.

For arbitrary $\epsilon > 0$,

$$\Pr(|\hat{\mu}_j - \mu_j| \geq \epsilon) = \Pr(\sqrt{n}|\hat{\mu}_j - \mu_j| \geq \sqrt{n}\epsilon) = \mathrm{E}\left[\Pr\left(v^{1/2}(\mu_j)e^{z_j/2}\sqrt{n}\cdot|\bar{\varepsilon}_{.j}| \geq \sqrt{n}\cdot\epsilon|z_j\right)\right] \to 0,$$

as $n \to \infty$. Thus, $\hat{\mu}_j$ is consistent.

Now, let $F(\boldsymbol{x}) = \Pr(\sqrt{n}(\hat{\boldsymbol{\mu}}_l - \boldsymbol{\mu}_l) \leq \boldsymbol{x}_l)$ be the unconditional distribution function, where $\boldsymbol{x}_l = (x_{j1}, \ldots, x_{jl})^\top \in \mathbb{R}^l$, and $\{\sqrt{n}(\hat{\boldsymbol{\mu}}_l - \boldsymbol{\mu}_l) \leq \boldsymbol{x}_l\} = \{\sqrt{n}(\hat{\mu}_{js} - \mu_{js}) \leq x_{js}, \quad s = 1, \ldots, l\} \subset \mathbb{R}^l$. We have

$$\begin{aligned}
F(\boldsymbol{x}_l) &= \mathrm{E}[I(\sqrt{n}(\hat{\boldsymbol{\mu}}_l - \boldsymbol{\mu}_l) \leq \boldsymbol{x}_l)] = \mathrm{E}\{\mathrm{E}[I(\sqrt{n}(\hat{\boldsymbol{\mu}}_l - \boldsymbol{\mu}_l) \leq \boldsymbol{x}_l)]|\boldsymbol{z}_l]\} \\
&= \mathrm{E}\left\{(2\pi)^{-l/2}|\boldsymbol{\Gamma}(\boldsymbol{z}_l)|^{-1/2}\left[\int_{-\infty}^{x_{j1}}\cdots\int_{-\infty}^{x_{jl}}\exp\left(-\frac{1}{2}\boldsymbol{s}^\top\boldsymbol{\Gamma}(\boldsymbol{z}_l)^{-1}\boldsymbol{s}\right)d\boldsymbol{s}\right]\right\} \\
&= (2\pi)^{-l}\int_{\mathbb{R}^l}|\boldsymbol{\Sigma}_l(\boldsymbol{\theta})\boldsymbol{\Gamma}(\boldsymbol{z}_l)|^{-1/2}\left[\int_{-\infty}^{x_{j1}}\cdots\int_{-\infty}^{x_{jl}}\exp\left(-\frac{1}{2}\boldsymbol{s}^\top\boldsymbol{\Gamma}(\boldsymbol{z}_l)^{-1}\boldsymbol{s}\right)d\boldsymbol{s}\right]\cdot \\
&\quad \exp\left(-\frac{1}{2}\boldsymbol{z}_l^\top\boldsymbol{\Sigma}_l(\boldsymbol{\theta})^{-1}\boldsymbol{z}_l\right)d\boldsymbol{z}_l.
\end{aligned}$$

Taking the derivative of $F(\boldsymbol{x}_l)$ leads the asymptotic density (11). □

Let us explore the variance and covariance of the above (unconditional) asymptotic distribution in details. It follows from Theorem 1(iii) that the asymptotic density function of $\sqrt{n}(\hat{\mu}_j - \mu_j)$ is

$$f(x_j) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\sigma_{jj}(\boldsymbol{\theta})^{-1/2}v(\mu_j)^{-1/2}e^{-z_j/2}\exp\left(-\frac{x_j^2}{2v(\mu_j)}e^{-z_l/2} - \frac{z_j^2}{2\sigma_{jj}(\boldsymbol{\theta})}\right)dz_j,$$

For each j, the asymptotic variance of $\sqrt{n}(\hat{\mu}_j - \mu_j)$ is calculated by

$$\int_R x_j^2 f(x_j)dx_j = v(\mu_j)\exp\left(\frac{\sigma_{jj}(\boldsymbol{\theta})}{8}\right),$$

where $\sigma_{jj}(\boldsymbol{\theta})$ is the jth diagonal element of variance-covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. Use the results of Theorem 1(iii) and go through some algebra. We can show that the asymptotic covariance of $\hat{\mu}_i$ and $\hat{\mu}_l$ is zero.

The above result implies that $\hat{\mu}_i = \bar{y}_{(.i)}$ and $\hat{\mu}_l = \bar{y}_{(.l)}$ are asymptotically uncorrelated. Therefore, for a fixed size m as $n \to \infty$, we have that $\mathrm{Var}[\sqrt{n}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})] \to$

$$\Sigma_{\hat{\boldsymbol{\mu}}} = diag\left[v(\mu_1)\exp\left(\frac{\sigma_{11}(\boldsymbol{\theta})}{8}\right), \ldots, v(\mu_m)\exp\left(\frac{\sigma_{mm}(\boldsymbol{\theta})}{8}\right)\right]. \quad (12)$$

In practice, the mean vector $\boldsymbol{\mu}$ is often associated with an unkown parameter vector, say $\boldsymbol{\beta}$, by a function $f(\cdot)$. See Appendix B for the discussion of the estimate of the parameter $\boldsymbol{\beta}$ in mean functions, where (12) plays an important role in calculating the asymptotic covariance matrix for $\hat{\beta}$. Next section will discuss estimates of the parameters $\boldsymbol{\theta}$ in the variance-covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$.

3.2 Parameters in the Variance-Covariance Matrix

For estimating $\boldsymbol{\theta}$ in the matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, we will work on the profile function of the approximated IEQL (9) with mean parameters estimated from the method given in Section 3.1. To simplify the expressions, the following notations are used.

Let $\boldsymbol{\tau}(\boldsymbol{d}(\hat{\boldsymbol{\mu}}), \boldsymbol{\theta})$ be the solution of the equation $\boldsymbol{k}'_z(\boldsymbol{d}(\hat{\boldsymbol{\mu}}), \boldsymbol{z}, \boldsymbol{\theta}) = 0$ with respect to $\boldsymbol{z}$, where $\boldsymbol{k}'_z(\cdot)$ is defined by (5). Use $\boldsymbol{z}^* = \boldsymbol{\tau}(\boldsymbol{d}(\hat{\mu}), \boldsymbol{\theta})$ and $\boldsymbol{d}(\hat{\boldsymbol{\mu}})$ in (9), and ignore the double summation term in $k(\boldsymbol{d}, \boldsymbol{z}, \boldsymbol{\theta})$ of (4), which is nothing to do with the unknown parameters. This results in the following profile function of $\boldsymbol{\theta}$:

$$\begin{aligned} l_0(\boldsymbol{\theta}) &= l(\boldsymbol{\tau}(\hat{\mu}), \boldsymbol{\theta}), \boldsymbol{\theta}) \\ &= \log |2\boldsymbol{I}_m + diag(\boldsymbol{d}(\hat{\boldsymbol{\mu}})e^{\boldsymbol{\tau}})\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \sum_{j=1}^{m} d_j(\hat{\mu}_j)e^{\tau_j} + n\sum_{j=1}^{m} \tau_j + \boldsymbol{\tau}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{\tau}. \end{aligned} \quad (13)$$

The IEQL estimator of parameter vector $\boldsymbol{\theta}$ is obtained by solving equations $\partial l_0(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = 0$. Note that $\boldsymbol{\tau}$ involves $\boldsymbol{\theta}$, thus

$$\frac{\partial l_0(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial l(\boldsymbol{\tau}, \boldsymbol{\theta})}{\partial \boldsymbol{\tau}} \cdot \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}} + \frac{\partial l(\boldsymbol{\tau}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \quad (14)$$

Because the implicit function $\boldsymbol{z} = \boldsymbol{\tau}(\boldsymbol{d}(\hat{\boldsymbol{\mu}}), \boldsymbol{\theta})$ are defined by $\boldsymbol{k}'_z(\boldsymbol{d}(\hat{\boldsymbol{\mu}}), \boldsymbol{z}, \boldsymbol{\theta}) = 0$, it follows that

$$\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\theta}} = -\left(\frac{\partial \boldsymbol{k}'_z(\boldsymbol{d}(\hat{\boldsymbol{\mu}}), \boldsymbol{\tau}, \boldsymbol{\theta})}{\partial \boldsymbol{\tau}}\right)^{-1} \frac{\partial \boldsymbol{k}'_z(\boldsymbol{d}(\hat{\boldsymbol{\mu}}), \boldsymbol{\tau}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\left(\boldsymbol{k}''_z(\boldsymbol{d}(\hat{\boldsymbol{\mu}}), \boldsymbol{\tau}, \boldsymbol{\theta})\right)^{-1} \frac{\partial \boldsymbol{k}'_z(\boldsymbol{d}(\hat{\boldsymbol{\mu}}), \boldsymbol{\tau}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \quad (15)$$

Plugging (5) and (6) into (15), under the regularity conditions (R2) and (R3), we have

$$\frac{\partial \boldsymbol{\tau}}{\partial \theta_i} = -2[diag(\boldsymbol{d}(\hat{\boldsymbol{\mu}})e^{\boldsymbol{\tau}} + 2\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}]^{-1} \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}}{\partial \theta_i}\boldsymbol{\tau}. \quad (16)$$

The following Lemma, derived from Nering (1983), is used to derive the differentiation of (14) and (16).

Lemma 1. *Let $\boldsymbol{X}$ be an $m \times m$ matrix such that each element of $\boldsymbol{X}$ is a real function of t; that is $x_{ij} = x_{ij}(t)$. Then, we have*

$$\frac{\partial \log |\boldsymbol{X}|}{\partial t} = tr\left(\boldsymbol{X}^{-1}\frac{\partial \boldsymbol{X}}{\partial t}\right), \quad \text{and} \quad \frac{\partial \boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}}{\partial t} = -\boldsymbol{A}\boldsymbol{X}^{-1}\frac{\partial \boldsymbol{X}}{\partial t}\boldsymbol{X}^{-1}\boldsymbol{B},$$

where $\boldsymbol{A}$ is a $p \times m$ matrix, $\boldsymbol{B}$ is a $m \times q$ matrix, and both $\boldsymbol{A}$ and $\boldsymbol{B}$ are free from t.

Note that, when $\boldsymbol{A} = \boldsymbol{B} = \boldsymbol{I}_m$, it follows from the Lemma 1 that

$$\frac{\partial \boldsymbol{X}^{-1}}{\partial t} = -\boldsymbol{X}^{-1}\frac{\partial \boldsymbol{X}}{\partial t}\boldsymbol{X}^{-1}. \tag{17}$$

Denoted by

$$\boldsymbol{W}(\boldsymbol{\theta}) = 2\boldsymbol{I}_m + diag(\boldsymbol{d}(\hat{\boldsymbol{\mu}})e^{-\boldsymbol{\tau}})\boldsymbol{\Sigma}(\boldsymbol{\theta}), \quad \text{and} \tag{18}$$

$$K(\boldsymbol{\theta}) = \boldsymbol{k}(\boldsymbol{d}(\hat{\boldsymbol{\mu}}), \boldsymbol{\tau}, \boldsymbol{\theta}) = \sum_{j=1}^{m} d_j(\hat{\mu}_j)e^{-\tau_j} + n\sum_{j=1}^{m}\tau_j + \boldsymbol{\tau}^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{\tau}. \tag{19}$$

Then, $l_0(\boldsymbol{\theta}) = \log|\boldsymbol{W}(\boldsymbol{\theta})| + K(\boldsymbol{\theta})$. Following Lemma 1 and recalling that $\boldsymbol{\tau}$ satisfies $k'_z(\boldsymbol{d}(\hat{\boldsymbol{\mu}}), \boldsymbol{\tau}, \boldsymbol{\theta}) = 0$, we have

$$\frac{\partial \log|\boldsymbol{W}(\boldsymbol{\theta})|}{\partial \theta_i} = tr\left(\boldsymbol{W}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{W}(\boldsymbol{\theta})}{\partial \theta_i}\right) \quad \text{and} \tag{20}$$

$$\frac{\partial K(\boldsymbol{\theta})}{\partial \theta_i} = \boldsymbol{k}'_z(\boldsymbol{d}(\hat{\boldsymbol{\mu}}), \boldsymbol{\tau}, \boldsymbol{\theta})\frac{\partial \boldsymbol{\tau}}{\partial \theta_i} - \boldsymbol{\tau}^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\hat{\boldsymbol{z}} \tag{21}$$

$$= -\boldsymbol{\tau}^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{\tau}. \tag{22}$$

Finally, the estimation equations, $\partial l_0/\partial \boldsymbol{\theta} = 0$, can be written as

$$tr\left(\boldsymbol{W}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{W}(\boldsymbol{\theta})}{\partial \theta_i}\right) - \boldsymbol{\tau}(\boldsymbol{d}(\hat{\boldsymbol{\mu}}), \boldsymbol{\theta})^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{\tau}(\boldsymbol{d}(\hat{\boldsymbol{\mu}}), \boldsymbol{\theta}) = 0, \quad i = 1, \ldots, k. \tag{23}$$

Therefore, the IEQL estimate of variance-covariance parameter $\boldsymbol{\theta}$ is determined by equation (23).

In general, there is no explicit solution from the estimating equations (23) for the parameters $\boldsymbol{\theta}$. Moreover, the complicated nonlinear form of the estimating equations makes it difficult to use the traditional Taylor expansion method for deriving the asymptotic distribution of $\hat{\boldsymbol{\theta}}$. However, in the case of large n we can simplify the estimating equations and obtain nice asymptotic properties of the resulted estimates. See Sections 4.1 and 4.2 for details. Moreover, in some special cases, such as $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = diag(\boldsymbol{X}\boldsymbol{\theta})$, where $\boldsymbol{X}$ is a $m \times k$ known matrix, and $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \theta\boldsymbol{\Sigma}_0$, where θ is a scalar parameter and $\boldsymbol{\Sigma}_0$ is an $m \times m$ known positive matrix, we can obtain closed form estimates of $\boldsymbol{\theta}$ and more compact asymptotics. See Section 4.3 for details.

4. Approximate Solutions of IEQL Estimating Equations

4.1 Approximate Solutions of Equation $\boldsymbol{k}'_z = 0$

To get IEQL estimate $\hat{\boldsymbol{\theta}}$ by solving estimating equations (23), we have to solve equation (8), i.e., $\boldsymbol{k}'_z(\boldsymbol{d}, \boldsymbol{z}, \boldsymbol{\theta}) = 0$. Generally, for finite n, there is no explicit solution of $\boldsymbol{z}$ from (8), and the solution

$\boldsymbol{z} = \boldsymbol{\tau}(\boldsymbol{d}, \boldsymbol{\theta})$ as an implicit function of $\boldsymbol{\theta}$ essentially depends on the structure of the variance-covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. However, when n is large (m fixed), we can get an approximate solution of (8) in a closed form. In fact, equation (8) can be written as

$$-\frac{1}{n}\boldsymbol{d}(\boldsymbol{\mu})e^{-\boldsymbol{z}} + \mathbf{1}_m + \frac{2}{n}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{z} = 0. \tag{24}$$

Under the regularity conditions (R1) and (R4), there exists $N_e > 0$ such that $\mathrm{E}(\varepsilon_j^4|\boldsymbol{z}) < N_e$. Then,

$$\begin{aligned}
\mathrm{E}\left[\left(-2\int_{Y_j}^{\mu_j}\frac{Y_j - x}{v(x)}dx\right)^2\right] &< \frac{1}{\delta^2}\mathrm{E}\left[\left(-2\int_{Y_j}^{\mu_j}(Y_j - x)dx\right)^2\right] = \frac{1}{\delta^2}\mathrm{E}[(Y_j-\mu_j)^4] \\
&= \frac{1}{\delta^2}\mathrm{E}\left([\mathrm{E}(Y_j-\mu_j)^4|\boldsymbol{z}]\right) = \frac{1}{\delta^2}\mathrm{E}\left([v^2(\mu_j)e^{2z_j}\mathrm{E}(\varepsilon_j^4|\boldsymbol{z}]\right) \\
&< \frac{N_e v^2(\mu_j)}{\delta^2}\mathrm{E}\left(e^{2z_j}\right) < \infty,
\end{aligned}$$

if $w_j = e^{z_j}$, the random scale defined in (1), has a second moment. Thus, we know that

$$\mathrm{E}\left(\int_{Y_j}^{\mu_j}\frac{Y_j - x}{v(x)}dx\right) \quad \text{and} \quad \mathrm{E}\left[\left(\int_{Y_j}^{\mu_j}\frac{Y_j - x}{v(x)}dx\right)^2\right] \text{exist for} \quad j = 1, \ldots, m.$$

It follows from (3) and Kolmogorov Large Number theory, that with probability one (w.p.1),

$$\frac{1}{n}d_j(\mu_j) = -\frac{2}{n}\sum_{i=1}^{n}\int_{y_{ij}}^{\mu_j}\frac{y_{ij} - x}{v(x)}dx \quad \longrightarrow \quad -2\mathrm{E}\left(\int_{Y_j}^{\mu_j}\frac{Y_j - x}{v(x)}dx\right) > 0 \ \text{ as } \ n \to \infty,$$

which implies that, with probability one, $d_j(\mu_j) = O(n)$.

Under the regularity condition (R3) that the component values of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ are bounded uniformly for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. When n is large (m fixed), the term $2n^{-1}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{z}$ in equation (24) can be ignored, and the equation can be solved approximately by $n^{-1}\boldsymbol{d}(\boldsymbol{\mu})e^{-\boldsymbol{z}} = \mathbf{1}_m$, or equivalently, $d_j(\mu_j)e^{-z_j} = n$, $j = 1, \ldots, m$, and thus, the approximate solution of z_j is given by

$$z_j = \log\left(\frac{d_j(\mu_j)}{n}\right), \quad j = 1, ..., m. \tag{25}$$

Then, approximate solution $z_j = \tau_j(d_j(\hat{\mu}_j), \boldsymbol{\theta})$ is given by

$$\tau_j = \log\left(\frac{d_j(\hat{\mu}_j)}{n}\right) = \log\left(-\frac{2}{n}\sum_{i=1}^{n}\int_{y_{ij}}^{\hat{\mu}_j}\frac{y_{ij} - x}{v(x)}dx\right), \quad j = 1, ..., m, \tag{26}$$

in which we replace μ_j by $\hat{\mu}_j = \bar{y}_{\cdot j}$.

To avoid the integration in the expression of $d_j(\mu_j)$, which is hard to deal with when we discuss the asymptotic properties of IEQL estimate $\hat{\boldsymbol{\theta}}$, we show that d_j can be approximated well by the Pearson χ^2-statistic, i.e.,

$$d_j(\mu_j) \approx \frac{1}{v(\mu_j)} \sum_{i=1}^{n} (y_{ij} - \mu_j)^2, \quad j = 1, 2, \ldots, m. \tag{27}$$

Denoted by $d(\mu) = -2 \int_y^\mu (y-t)/v(t)dt$. We have

$$d(y) = 0, \quad d'(y) = \frac{\partial d(\mu)}{\partial \mu}|_{\mu=y} = 0, \quad \text{and} \quad d''(y) = \frac{\partial^2 d(\mu)}{\partial \mu^2}|_{\mu=y} = \frac{2}{v(\mu)}.$$

Hence, the second-order Taylor expansion of $d(\mu)$ around $\mu = y$ gives

$$d(\mu) \quad = \quad d(y) + d'(y)(y-\mu) + \frac{1}{2} d''(\mu^*)(y-\mu)^2 = \frac{(y-\mu)^2}{v(\mu^*)} \approx \frac{(y-\mu)^2}{v(\mu)},$$

where μ^* lies between y and μ. Then, (27) is obtained by summing over i when substituting μ with μ_j and y with y_{ij}.

It follows from (27) that, for large n, the solution (26) can be further approximated by:

$$\tau_j = \log \left(\frac{1}{nv(\hat{\mu}_j)} \sum_{i=1}^{n} (y_{ij} - \hat{\mu}_j)^2 \right), \quad j = 1, \ldots, m. \tag{28}$$

The next lemma states that τ_j obtained from (28) is asymptotically equivalent to z_j as $n \to \infty$.

Lemma 2. *Under the model assumption (1) and a fixed size m, denoted by*

$$\tau_j = \log \left(\frac{1}{nv(\bar{y}_{\cdot j})} \sum_{i=1}^{n} (y_{ij} - \bar{y}_{\cdot j})^2 \right), \quad j = 1, \ldots, m. \tag{29}$$

Then, for given $\boldsymbol{z}$, as $n \to \infty$, conditionally there are $\tau_j \overset{w.p.1}{\longrightarrow} z_j$, $j = 1, \ldots, m$.

Proof: It follows from model assumption (1) that we have:

$$\begin{aligned} e^{\tau_j} &= \frac{d_j(\hat{\mu}_j)}{n} = \frac{1}{nv(\bar{y}_{\cdot j})} \sum_{i=1}^{n} (y_{ij} - \bar{y}_{\cdot j})^2 = \frac{1}{nv(\bar{y}_{\cdot j})} \sum_{i=1}^{n} v(\bar{y}_{\cdot j}) e^{z_j} (\varepsilon_{ij} - \varepsilon_{\cdot j})^2 \\ &= \frac{e^{z_j}}{n} \sum_{i=1}^{n} (\varepsilon_{ij} - \bar{\varepsilon}_{\cdot j})^2. \end{aligned}$$

Then, from the assumption (1) and Kolmogorov Large Number theory, we know that as $n \to \infty$, conditioning on $\boldsymbol{z}$, we have $e^{\tau_j} = d_j(\hat{\mu}_j)/n \overset{w.p.1}{\longrightarrow} e^{z_j} \mathrm{Var}(\varepsilon_{ij}|\boldsymbol{z}) = e^{z_j}$ or equivalently, $\tau_j = \log(d_j(\hat{\mu}_j)/n) \overset{w.p.1}{\longrightarrow} z_j$, $j = 1, \ldots, m$. □

If function $v(\mu)$ is constant, then τ_j's given by (26) and (29) are exactly the same. When function $v(\mu)$ varies slowly in μ, those τ_j's are very close. Note that, the approximate solution of τ_j does not depend on $\boldsymbol{\theta}$. The most important step is to recognize the property of $d_j(\hat{\mu}_j)e^{-\tau_j} = n$. This leads to $diag(\boldsymbol{d}(\hat{\boldsymbol{\mu}})e^{-\boldsymbol{\tau}}) = n\boldsymbol{I}_m$, where $\boldsymbol{\tau} = (\tau_1, \dots, \tau_m)^\top$. Thus, the IEQL equation (23) becomes much simpler.

It follows from (18) that, with the τ_j given by (26) or (29), $\boldsymbol{W}(\boldsymbol{\theta}) = 2\boldsymbol{I}_m + diag(\boldsymbol{d}(\hat{\boldsymbol{\mu}})e^{-\boldsymbol{\tau}})\boldsymbol{\Sigma}(\boldsymbol{\theta}) = 2\boldsymbol{I}_m + n\boldsymbol{\Sigma}(\boldsymbol{\theta})$. Moreover, we have

$$\frac{\partial \boldsymbol{W}(\boldsymbol{\theta})}{\partial \theta_i} = n\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}.$$

Then,

$$tr\left(W(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{W}(\boldsymbol{\theta})}{\partial \theta_i}\right) = tr\left((2\boldsymbol{I}_m + n\boldsymbol{\Sigma}(\boldsymbol{\theta}))^{-1}\; n\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\right).$$

Bring n into the reciprocal function and shorten the notation $\boldsymbol{\tau}(\boldsymbol{d}(\hat{\boldsymbol{\mu}}), \boldsymbol{\theta})$ to $\boldsymbol{\tau}$. The IEQL estimating equation (23) thus becomes

$$tr\left((2\boldsymbol{I}_m/n + \boldsymbol{\Sigma}(\boldsymbol{\theta}))^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\right) - \boldsymbol{\tau}^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{\tau} = 0, \quad i = 1, \dots, k. \tag{30}$$

Next, we discuss the cases that the above IEQL equations of $\boldsymbol{\theta}$ could be simplified.

Case I. For large n and finite m, $2\boldsymbol{I}_m/n$ in the first term of equation (30) could be ignored. Thus, equation (30) can be approximated very well by

$$\boldsymbol{\tau}^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{\tau} - tr\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\right) = 0, \quad i = 1, \dots, k. \tag{31}$$

Follow from the properties of quadratic form of multivariate normal distribution. We know

$$\lim_{n\to\infty} \mathrm{E}(\boldsymbol{\tau}^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{\tau}) = tr\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\right). \tag{32}$$

Thus, define

$$H_i(\boldsymbol{\theta}) = \frac{1}{m}\boldsymbol{\tau}^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{\tau} - \frac{1}{m}tr\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\right), \quad i = 1, \dots, k. \tag{33}$$

The approximated IEQL estimate $\hat{\boldsymbol{\theta}}$ can be obtained by solving $H_i(\boldsymbol{\theta}) = 0$, $i = 1, \dots, k$.

Case II. For finite n and m, when the structure of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is specified, the equations may be simplified further. For example, If $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = diag(\sigma_1(\boldsymbol{\theta}), \dots, \sigma_m(\boldsymbol{\theta}))$, the estimating equations can be written as

$$\sum_{j=1}^{m}\frac{\partial \sigma_j(\boldsymbol{\theta})}{\partial \theta_i}\left(\frac{n}{2 + n\sigma_j(\boldsymbol{\theta})} - \frac{\tau_j^2}{\sigma_j^2(\boldsymbol{\theta})}\right) = 0, \quad i = 1, \dots, k, \tag{34}$$

which can be easily solved as the weighted regression equations. Furthermore, for large n (fixed size m), the equations (34) can be approximated very well by

$$\sum_{j=1}^{m} \frac{1}{\sigma_j^2(\boldsymbol{\theta})} \frac{\partial \sigma_j(\boldsymbol{\theta})}{\partial \theta_i} (\tau_j^2 - \sigma_j(\boldsymbol{\theta})) = 0, \quad i = 1, \ldots, k, \tag{35}$$

which are very similar to the generalized estimating equations (Liang & Zeger, 1986). In some cases discussed in Section 4.3, equation (35) leads to a consistent estimate, when both n and $m \to \infty$.

4.2 Asymptotic properties of the IEQL estimate $\hat{\boldsymbol{\theta}}$ - General Case

In the situation that m is finite, even if n is sufficiently larger, we are not able to get enough information about the random scales to get a consistent estimate for variance-covariance parameters $\boldsymbol{\theta}$. However, if we allow m to be sufficiently large comparing with k the number of the unknown parameters, we can get a "consistent" estimate of $\boldsymbol{\theta}$ in the sense of that the asymptotic variance of $\hat{\boldsymbol{\theta}}$ is small enough, say $O(m^{-1})$. Nevertheless, in some special cases, with finite m we can show that IEQL estimates $\hat{\boldsymbol{\theta}}$ is asymptotically unbiased, i.e., $\lim_{n \to \infty} \mathrm{E}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$, and we can get the asymptotic distribution of $\hat{\boldsymbol{\theta}}$. See Theorem 3(i) and 5(ii) in the next section for details.

Since this section discusses the asymptotic properties of $\hat{\boldsymbol{\theta}}$ in the general case, we let both n and m go to infinity. We shall discuss the asymptotic properties of $H_i(\boldsymbol{\theta})$ first. In the following discussion, the technique of matrix characteristic decomposition will be used without trivial details.

Let $\delta_1(\boldsymbol{\theta}) \geq \delta_2(\boldsymbol{\theta}) \geq \cdots \geq \delta_m(\boldsymbol{\theta}) > 0$ be the characteristic roots of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, and $\boldsymbol{q}_j(\boldsymbol{\theta})$ satisfied $||\boldsymbol{q}_j(\boldsymbol{\theta})|| = 1$ be the standardized characteristic vectors with respect to $\delta_j(\boldsymbol{\theta})$, $j = 1, \ldots, m$, respectively. Then we have

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{Q}(\boldsymbol{\theta}) \boldsymbol{\Delta}(\boldsymbol{\theta}) \boldsymbol{Q}^\top(\boldsymbol{\theta}) = \sum_{j=1}^{m} \delta_j(\boldsymbol{\theta}) \boldsymbol{q}_j(\boldsymbol{\theta}) \boldsymbol{q}_j^\top(\boldsymbol{\theta}),$$

where $\boldsymbol{Q}(\boldsymbol{\theta}) = (\boldsymbol{q}_1(\boldsymbol{\theta}), \ldots, \boldsymbol{q}_m(\boldsymbol{\theta}))$, an $m \times m$ orthogonal matrix, and $\boldsymbol{\Delta}(\boldsymbol{\theta}) = diag(\delta_1(\boldsymbol{\theta}), \ldots, \delta_m(\boldsymbol{\theta}))$.

According to Lemma 2, as $n \to \infty$, we know that $\boldsymbol{\zeta} = \boldsymbol{Q}^\top(\boldsymbol{\theta}) \boldsymbol{\tau} \xrightarrow{d} N_m(0, \boldsymbol{\Delta}(\boldsymbol{\theta}))$. Note that $\boldsymbol{Q}^\top(\boldsymbol{\theta}) \boldsymbol{Q}(\boldsymbol{\theta}) = \boldsymbol{I}$. Thus, $\boldsymbol{\tau} = \boldsymbol{Q}(\boldsymbol{\theta}) \boldsymbol{\zeta}$. Therefore, we have

$$H_i(\boldsymbol{\theta}) = \frac{1}{m} \boldsymbol{\zeta}^\top \boldsymbol{Q}^\top(\boldsymbol{\theta}) \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i} \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \boldsymbol{Q}(\boldsymbol{\theta}) \boldsymbol{\zeta} - \frac{1}{m} tr \left(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i} \right), \quad i = 1, \ldots, k. \tag{36}$$

Denoted by

$$\boldsymbol{A}_i(\boldsymbol{\theta}) = \boldsymbol{Q}^\top(\boldsymbol{\theta}) \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i} \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \boldsymbol{Q}(\boldsymbol{\theta}) \quad \text{and} \quad \boldsymbol{B}_i(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}, i = 1, \ldots, k.$$

Let $\lambda_{i1}(\boldsymbol{\theta}) \geq \lambda_{i2}(\boldsymbol{\theta}) \geq \cdots \geq \lambda_{im}(\boldsymbol{\theta})$ be the characteristic roots of $\boldsymbol{A}_i(\boldsymbol{\theta})$, and $\boldsymbol{r}_{ij}(\boldsymbol{\theta})$ satisfied $||\boldsymbol{r}_{ij}(\boldsymbol{\theta})|| = 1$ be the standardized characteristic vectors with respect to $\lambda_{ij}(\boldsymbol{\theta})$, for $j = 1, \ldots, m$, and $i = 1, \ldots, k$, respectively. Then, we know that

$$\boldsymbol{\tau}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i} \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \boldsymbol{\tau} = \boldsymbol{\zeta}^\top \boldsymbol{Q}^\top(\boldsymbol{\theta}) \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i} \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \boldsymbol{Q}(\boldsymbol{\theta}) \boldsymbol{\zeta} = \boldsymbol{\zeta}^\top \boldsymbol{A}_i(\boldsymbol{\theta}) \boldsymbol{\zeta}.$$

Therefore, we have that, as $n \to \infty$,

$$\begin{aligned} & \mathrm{E}\left(\boldsymbol{\tau}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i} \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \boldsymbol{\tau}\right) = \mathrm{E}(\boldsymbol{\zeta}^\top \boldsymbol{A}_i(\boldsymbol{\theta}) \boldsymbol{\zeta}) = tr(\boldsymbol{A}_i(\boldsymbol{\theta}) \boldsymbol{\Delta}(\boldsymbol{\theta})) \\ = \ & tr(\sum_{j=1}^m \lambda_{ij}(\boldsymbol{\theta}) \boldsymbol{r}_{ij}(\boldsymbol{\theta}) \boldsymbol{r}_{ij}(\boldsymbol{\theta})^\top \boldsymbol{\Delta}(\boldsymbol{\theta})) = \sum_{j=1}^m \lambda_{ij}(\boldsymbol{\theta}) v_{ij}^2(\boldsymbol{\theta}), \end{aligned} \tag{37}$$

where $v_{ij}^2(\boldsymbol{\theta}) = \boldsymbol{r}_{ij}^\top(\boldsymbol{\theta}) \boldsymbol{\Delta}(\boldsymbol{\theta}) \boldsymbol{r}_{ij}(\boldsymbol{\theta})$. Note that the charactersitc vectors $\boldsymbol{r}_{ij}(\boldsymbol{\theta})$, $j = 1, 2, \ldots, m$, are orthogonal each other, it follows that

$$\begin{aligned} & \mathrm{Var}\left(\boldsymbol{\tau}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i} \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \boldsymbol{\tau}\right) = \mathrm{Var}(\boldsymbol{\zeta}^\top \boldsymbol{A}_i(\boldsymbol{\theta}) \boldsymbol{\zeta}) = 2tr(\boldsymbol{A}_i(\boldsymbol{\theta}) \boldsymbol{\Delta}(\boldsymbol{\theta}) \boldsymbol{A}_i(\boldsymbol{\theta}) \boldsymbol{\Delta}(\boldsymbol{\theta})) \\ = \ & 2 \sum_{j=1}^m \lambda_{ij}^2(\boldsymbol{\theta}) v_{ij}^4(\boldsymbol{\theta}) \end{aligned}$$

Comparing (32) to (37), for large n $H_i(\boldsymbol{\theta})$ can be equivalently written as

$$H_i(\boldsymbol{\theta}) = \frac{1}{m} \sum_{j=1}^m \lambda_{ij}(\boldsymbol{\theta})(\xi_{ij}^2 - v_{ij}^2(\boldsymbol{\theta})), i = 1, \ldots, k.$$

Let $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ be the ture parameter. Under the following regularity conditions that, for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

(R1*)

$$\lim_{n,m \to \infty} \mathrm{Var}[\sqrt{m} H_i(\boldsymbol{\theta})] = \lim_{m \to \infty} \frac{2}{m} \sum_{j=1}^m \lambda_{ij}^2(\boldsymbol{\theta}) v_{ij}^4(\boldsymbol{\theta}) = \boldsymbol{U}(\boldsymbol{\theta}) \text{ exists, and}$$

$$\lim_{n,m \to \infty} \mathrm{Var}[H_i(\boldsymbol{\theta})] = \lim_{m \to \infty} \frac{2}{m^2} \sum_{j=1}^m \lambda_{ij}^2(\boldsymbol{\theta}) v_{ij}^4(\boldsymbol{\theta}) \to 0;$$

(R2*) $\partial H_i(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ and $\partial^2 H_i(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$ exist, and the all elements in the first two derivates are bounded uniformly, $i = 1, \ldots, k$;

(R3*) denoting $\boldsymbol{H}(\boldsymbol{\theta}) = (H_1(\boldsymbol{\theta}), \ldots, H_k(\boldsymbol{\theta}))^\top$, assume that, $k \times k$ matrix $\partial \boldsymbol{H}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ is fullly ranked with probability one for sufficient large n and m.

Now, under the regularity conditions (R1*)-(R3*), we are ready to show that $\hat{\boldsymbol{\theta}}$, as the solution of (33), is strongly consistent and asymptotic normal.

Denoted by $\boldsymbol{H}(\boldsymbol{\theta}) = (H_1(\boldsymbol{\theta}), \ldots, H_k(\boldsymbol{\theta}))^\top$. According to (R1*), we know that $\boldsymbol{H}(\boldsymbol{\theta}_0) \xrightarrow{wp1} 0$, $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, and based on (R2*), Taylor expansion of $\boldsymbol{H}(\hat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}_0$ leads

$$\boldsymbol{H}(\boldsymbol{\theta}_0) = \frac{\partial \boldsymbol{H}}{\partial \boldsymbol{\theta}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + O((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^2) = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \times \left(-\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{\theta}} + O(|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|) \right) \xrightarrow{wp1} 0, \tag{38}$$

which implies that $\hat{\boldsymbol{\theta}} \xrightarrow{wp1} \boldsymbol{\theta}_0$, i.e., IEQL estimate of $\boldsymbol{\theta}$ is strongly consistent.

Note that condition $\lim_{n,m\to\infty} \mathrm{Var}[H_i(\boldsymbol{\theta})] = 0$ in (R1*) implies the Lindeberg-Feller condition, as $m \to \infty$, for each $\epsilon > 0$,

$$\frac{1}{m}\sum_{j=1}^{m} \mathrm{E}[(\xi_{ij}^2 - v_{ij}^2)^2 I(|\xi_{ij}^2 - v_{ij}^2| > \epsilon\sqrt{m})] \longrightarrow 0,$$

is held in our case (Rao (1973), page 147; Serfling (1980), page 31), and it follows that $\boldsymbol{H}(\boldsymbol{\theta})$ is asymptotic normal, i.e., $\sqrt{m}\boldsymbol{H}(\boldsymbol{\theta}) \xrightarrow{d} N(0, \boldsymbol{U}(\boldsymbol{\theta}))$, as $n, m \to \infty$.

With condition (R3*), the asymptotic normality of $\hat{\boldsymbol{\theta}}$ is obtained immediately from (38) as $\sqrt{m}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{V}(\boldsymbol{\theta}_0))$ when $n, m \to \infty$, where

$$\boldsymbol{V}(\boldsymbol{\theta}_0) = \left(\frac{\partial \boldsymbol{H}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)^{-1} \boldsymbol{U}(\boldsymbol{\theta}_0) \left(\left(\frac{\partial \boldsymbol{H}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)^{-1}\right)^\top . \tag{39}$$

We state the asymptotic results formally in the following theorem.

Theorem 2. *Denote $\hat{\boldsymbol{\theta}}$ be a solution of equation (31), and $\boldsymbol{\theta}_0$ be the true parameter. Under the regularity conditions (R1*)-(R3*), we have that, as $n, m \to \infty$,*

(1) $\hat{\boldsymbol{\theta}} \xrightarrow{wp1} \boldsymbol{\theta}_0$, and

(2) $\sqrt{m}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{V}(\boldsymbol{\theta}_0))$, where $\boldsymbol{V}(\boldsymbol{\theta}_0)$ is gaven in (39).

In general cases, regularity conditions (R1*) - (R3*) are necessary. However, when the structure of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is specified, the equations (31) and (33) and regularity conditions (R1*) - (R3*) may be simplified further in the derivation of the simplied version of the IEQL estimate $\hat{\boldsymbol{\theta}}$ given in (34) and (35). Details are presented in the next section.

4.3 Asymptotic properties of the estimate $\hat{\boldsymbol{\theta}}$ - Special Cases

Let us first consider the case that $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = diag(\boldsymbol{\sigma}(\boldsymbol{\theta}))$, where $\boldsymbol{\sigma}(\boldsymbol{\theta}) = (\sigma_1(\boldsymbol{\theta}), \ldots, \sigma_m(\boldsymbol{\theta}))^\top = \boldsymbol{X}\boldsymbol{\theta}$, and $\boldsymbol{X}$ is a $m \times k$ known matrix with a full column-rank. In this case, the IEQL estimating equation (35) can be written in the following matrix format:

$$(\boldsymbol{\tau}^2 - \boldsymbol{X}\boldsymbol{\theta})^\top [\boldsymbol{\Sigma}(\boldsymbol{\theta})]^{-2} \boldsymbol{X} = 0, \tag{40}$$

where $\boldsymbol{\tau}^2 = (\tau_1^2, \ldots, \tau_m^2)^\top$.

The following present an iterative algorithm similar to the generalized least squares (GLS) method for getting the solution of $\boldsymbol{\theta}$. Consider the initial estimate,

$$\hat{\boldsymbol{\theta}}^{(1)} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\tau}^2. \tag{41}$$

Use the following formula to update the estimate,

$$\hat{\boldsymbol{\theta}}^{(p+1)} = [\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-2}(\hat{\boldsymbol{\theta}}^{(p)}) \boldsymbol{X}]^{-1} [\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-2}(\hat{\boldsymbol{\theta}}^{(p)}) \boldsymbol{\tau}^2]. \tag{42}$$

Stop the iteration if $|\hat{\boldsymbol{\theta}}^{(p)} - \hat{\boldsymbol{\theta}}^{(p+1)}| < \epsilon$.

It can be shown that when $\boldsymbol{X}$ is an orthogonal column-matrix, the iteration process will stop at the second step, i.e., $\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(2)}$. Other cases will converge with more iterations. The following theorem show that the estimates obtained from this iterative algorithm have excellent asymptotic properties such as asymptotic unbiasness, consistency and normal distribution.

Theorem 3. *Under the model assumption (1) and regularity conditions (R1) and (R4), if* $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = diag(\boldsymbol{\sigma}(\boldsymbol{\theta}))$, $\boldsymbol{\sigma}(\boldsymbol{\theta}) = \boldsymbol{X}\boldsymbol{\theta}$, *and* $\boldsymbol{X}$ *is an* $m \times k$ *matrix with a full column-rank, then we have the following results:*

(i) *For finite* m *consider*

$$\tau_j = \log\left(\frac{1}{nv(\bar{y}_{.j})} \sum_{i=1}^{n} (y_{ij} - \bar{y}_{.j})^2\right), \quad j = 1, \ldots, m,$$

in Eq. (41), then $\hat{\boldsymbol{\theta}}^{(1)}$ *is asymptotically unbiased as* $n \to \infty$, *i.e.,* $\lim_{n\to\infty} \mathrm{E}(\hat{\boldsymbol{\theta}}^{(1)}) = \boldsymbol{\theta}$. *Moreover, when* $n \to \infty$ *the asymptotic distribution of* $\hat{\boldsymbol{\theta}}^{(1)}$ *is a linear sum of independent chi-square distribution with one degree of freedom.*

(ii) *Under the regularity condition that*

(R.6) $\boldsymbol{X}$ *and* $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ *satisfy*

$$\lim_{m\to\infty} \frac{1}{m} \boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{A} \quad and \quad \lim_{m\to\infty} \frac{1}{m} \boldsymbol{X}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^2 \boldsymbol{X} = \boldsymbol{B} \quad exists,$$

then $\hat{\boldsymbol{\theta}}^{(1)} \xrightarrow{p} \boldsymbol{\theta}$ *as* $n, m \to \infty$.

Proof: (i) Follows from Lemma 2 that, conditioning on given $\boldsymbol{z}$, $\boldsymbol{\tau}^2 \xrightarrow{w.p.1} \boldsymbol{z}^2$, as $n \to \infty$. Furthermore, from the multivariate normal assumption on the logarithm of the random scale vector

$\boldsymbol{z}$ and the fact that $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = diag(\boldsymbol{X}\boldsymbol{\theta})$, the asymptotic unbiasedness of $\hat{\boldsymbol{\theta}}^{(1)}$ can be shown as follows:

$$\begin{aligned}
\lim_{n\to\infty} \mathrm{E}(\hat{\boldsymbol{\theta}}^{(1)}) &= \mathrm{E}[\lim_{n\to\infty} \mathrm{E}(\hat{\boldsymbol{\theta}}^{(1)}|\boldsymbol{z})] = \mathrm{E}[\mathrm{E}(\lim_{n\to\infty} \hat{\boldsymbol{\theta}}^{(1)}|\boldsymbol{z})] \\
&= \mathrm{E}[\mathrm{E}(\lim_{n\to\infty} (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\tau}^2|\boldsymbol{z}] = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\mathrm{E}(\boldsymbol{z}^2) \\
&= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\theta} = \boldsymbol{\theta},
\end{aligned}$$

where $\boldsymbol{z}^2 = (z_1^2, \ldots, z_m^2)^\top$, and the exchangeability of the limit and expectation is provided by the model assumption (1) and the regularity conditions.

Similarly, when n is large, with probability one, we have

$$\lim_{n\to\infty} \Pr(\hat{\boldsymbol{\theta}}^{(1)} \le t) = \Pr[\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{z}^2 \le t].$$

According to the normal distribution of $\boldsymbol{z}$ in (1) and the indpendence assumption of ε_{ij} (conditioning on $\boldsymbol{z}$), in this special case the asymptotic distribution of the IEQL estimate $\hat{\boldsymbol{\theta}}^{(1)} = \boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\tau}^2$ has a linear sum of chi-square distribution with one degree of freedom.

(ii) To show that $\hat{\boldsymbol{\theta}}^{(1)}$ is consistent, we only need to show that $\mathrm{Var}(\hat{\boldsymbol{\theta}}^{(1)}) \to 0$, as $n, m \to \infty$. Note that, when n is large, with probability one, we have

$$\begin{aligned}
\lim_{n\to\infty} \mathrm{Var}(\hat{\boldsymbol{\theta}}^{(1)}) &= \lim_{n\to\infty} \mathrm{Var}[\mathrm{E}(\hat{\boldsymbol{\theta}}^{(1)}|\boldsymbol{z})] = \mathrm{Var}[\lim_{n\to\infty} \mathrm{E}(\hat{\boldsymbol{\theta}}^{(1)}|\boldsymbol{z})] \\
&= \mathrm{Var}((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{z}^2) = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(2\boldsymbol{\Sigma}(\boldsymbol{\theta})^2)\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}.
\end{aligned}$$

Regularity condition (R.6) implies that $(\boldsymbol{X}^\top\boldsymbol{X})^{-1} = O(m^{-1}\boldsymbol{A}^{-1})$ and $(\boldsymbol{X}^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^2\boldsymbol{X}) = O(m\boldsymbol{B})$, it follows that, as $n, m \to \infty$, we have

$$\mathrm{Var}(\hat{\boldsymbol{\theta}}^{(1)}) = O(m^{-1}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}) \to 0.$$

The proof of Theorem 3 is thus completed. □

Theorem 4. *Under the conditions stated in Theorem 3, we have the following results.*

(i) *Let $\hat{\boldsymbol{\theta}}^{(p+1)}$ be the estimate from (42) with any p iteration steps. Then, $\hat{\boldsymbol{\theta}}^{(p+1)}$ is asymptotically unbiased and consistent as $n \to \infty$, i.e., $\lim_{n\to\infty} \mathrm{E}(\hat{\boldsymbol{\theta}}^{(p+1)}) = \boldsymbol{\theta}$, and as $n, m \to \infty$, $\hat{\boldsymbol{\theta}}^{(p+1)} \xrightarrow{p} \boldsymbol{\theta}$.*

(ii) *Let $\hat{\boldsymbol{\theta}}$ be the sulotion of (40), then $\sqrt{m}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N_k(0, 2\boldsymbol{B}^{-1})$, where*

$$\lim_{m\to\infty} \frac{1}{m}\boldsymbol{X}^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-2}\boldsymbol{X} = \boldsymbol{B}$$

Proof: For (i) and (ii), it is sufficient to show the case when $p = 1$ only. The proofs follow from a similar procedure given in Theorem 2 (i) and (ii), respectively. In particular, as $m \to \infty$,

$$\mathrm{Var}(\hat{\boldsymbol{\theta}}^{(p+1)}) = O((\boldsymbol{X}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-2}\boldsymbol{X})^{-1}) = O(m^{-1}\boldsymbol{B}^{-1}) \to 0.$$

(iii) The asymptotic normality can be obtained from (35). Note that the equation (35) is asymptotically equal to

$$\sum_{j=1}^{m} \frac{1}{\sigma_j^2(\boldsymbol{\theta})} \frac{\partial \sigma_j(\boldsymbol{\theta})}{\partial \theta_i}(z_j^2 - \sigma_j(\boldsymbol{\theta})) = 0, \quad i = 1, \ldots, k.$$

According to the normal assumption about z_j that $z_j^2/\sigma_j(\boldsymbol{\theta})$, $j = 1, \ldots, m$, are *iid* with $\chi^2(1)$ distribution. Thus, $U_j(\boldsymbol{\theta}) = (z_j^2 - \sigma_j(\boldsymbol{\theta}))/\sqrt{2}\sigma_j(\boldsymbol{\theta})$, $j = 1, \ldots, m$, are *iid* with random variables having zero-mean and unit-variance. Then, the function on the left side of (35) can be written as a linear combination of $U_j(\boldsymbol{\theta})$, $j = 1, \ldots, m$, such as $\boldsymbol{H}(\boldsymbol{\theta}) = \sum_{j=1}^m \boldsymbol{a}_j U_j(\boldsymbol{\theta})$, where $\boldsymbol{a}_j = \partial \log \sigma_j(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$. Applying Lindeberg-Levy central limit theorem, we can obtain the asymptotic normality of $m^{1/2}H(\boldsymbol{\theta})$. Then, by using the first-order Taylor expansion $m^{-1/2}H(\hat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}$, we obtain that

$$\sqrt{m}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = (\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{\theta}})^{-1} H(\boldsymbol{\theta})$$

is asymptotic normal with mean zero, as $n, m \to \infty$.

Similar to the proof of Theorem 2 (ii), from (42), we know that, when n is large, with probability one, we have

$$\begin{aligned}
\lim_{n\to\infty} \mathrm{Var}(\hat{\boldsymbol{\theta}}) &= \lim_{n\to\infty} \mathrm{Var}[\mathrm{E}(\hat{\boldsymbol{\theta}}|\boldsymbol{z})] = \mathrm{Var}[\lim_{n\to\infty} \mathrm{E}(\hat{\boldsymbol{\theta}}|\boldsymbol{z})] \\
&= \mathrm{Var}((\boldsymbol{X}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-2}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-2}\boldsymbol{z}^2) \\
&= (\boldsymbol{X}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-2}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-2}(2\boldsymbol{\Sigma}(\boldsymbol{\theta})^2)\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-2}\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-2}\boldsymbol{X})^{-1} \\
&= 2(\boldsymbol{X}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-2}\boldsymbol{X})^{-1}.
\end{aligned}$$

The assumption (R.6) that $\lim_{m\to\infty} \boldsymbol{X}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-2}\boldsymbol{X}/m = \boldsymbol{B}$ exists implies that $(\boldsymbol{X}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-2}\boldsymbol{X})^{-1} = O(m^{-1}\boldsymbol{B}^{-1})$. It follows that, as $n, m \to \infty$, we have

$$\mathrm{Var}(\sqrt{m}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})) \to 2\boldsymbol{B}^{-1}.$$

and the proof is completed. □

Consider another case of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. The next theorem presents an explicit form of the asymptotic unbiased IEQL estimator $\hat{\theta}$ and its asymptotic distribution.

Theorem 5. *Under the model assumption (1) and regularity conditions (R1) and (R4), when $\mathbf{\Sigma}(\boldsymbol{\theta}) = \theta\mathbf{\Sigma}_0$, where $\mathbf{\Sigma}_0$ is a known, positive definite $m \times m$ matrix and $\theta > 0$ is a scalar parameter, then*

(i) *The IEQL estimate for θ is given by $\hat{\theta} = \boldsymbol{\tau}^\top \mathbf{\Sigma}_0^{-1}\boldsymbol{\tau}/m$, where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_m)^\top$, and*

$$\tau_j = \log\left(\frac{1}{nv(\bar{y}_{.j})}\sum_{i=1}^{n}(y_{ij} - \bar{y}_{.j})^2\right), \quad j = 1, ..., m.$$

(ii) *As $n \to \infty$, for finite m $\hat{\theta}$ is asymptotically unbiased, and $m\hat{\theta}/\theta \xrightarrow{d} \chi^2(m)$.*

(iii) *As $n, m \to \infty$, $\hat{\theta}$ is consistent, and $\sqrt{m}(\hat{\theta} - \theta) \xrightarrow{d} N(0, 2\theta^2)$.*

Proof: (i) For the model that $\mathbf{\Sigma}(\theta) = \theta\mathbf{\Sigma}_0$, we have $\partial\mathbf{\Sigma}/\partial\theta = \mathbf{\Sigma}_0$ and $\partial\mathbf{\Sigma}^{-1}/\partial\theta = -\mathbf{\Sigma}_0^{-1}/\theta^2$. Using notations $\boldsymbol{W}(\boldsymbol{\theta})$ and $K(\theta)$ given by (18) and (19), respectively, the profile IEQL of θ is $l_0(\theta) = \log|\boldsymbol{W}(\theta)| + K(\theta)$. From Lemma 1, we have

$$\begin{aligned}\frac{\partial \log|\boldsymbol{W}|}{\partial\theta} &= tr\left(\mathbf{\Sigma}(\theta)^{-1}\frac{\partial\mathbf{\Sigma}(\theta)}{\partial\theta}\right) = tr(\mathbf{\Sigma}_0^{-1}\theta^{-1}\mathbf{\Sigma}_0) = tr(\boldsymbol{I}_m\theta^{-1}) = \frac{m}{\theta}, \quad \text{and} \\ \frac{\partial K}{\partial\theta} &= -\boldsymbol{\tau}^\top\mathbf{\Sigma}(\theta)^{-1}\frac{\partial\mathbf{\Sigma}(\theta)}{\partial\theta}\mathbf{\Sigma}(\theta)^{-1}\boldsymbol{\tau} = -\boldsymbol{\tau}^\top\frac{\mathbf{\Sigma}_0^{-1}}{\theta}\mathbf{\Sigma}_0\frac{\mathbf{\Sigma}_0^{-1}}{\theta}\boldsymbol{\tau} = -\frac{1}{\theta^2}\boldsymbol{\tau}^\top\mathbf{\Sigma}_0^{-1}\boldsymbol{\tau}.\end{aligned}$$

Then the IEQL equation $\partial l_0(\theta)/\partial\theta = 0$ becomes

$$\frac{m}{\theta} - \frac{\boldsymbol{\tau}^\top\mathbf{\Sigma}_0^{-1}\boldsymbol{\tau}}{\theta^2} = 0.$$

Solve this equation. The IEQL estimator of θ is given by $\hat{\theta} = \boldsymbol{\tau}^\top\mathbf{\Sigma}_0^{-1}\boldsymbol{\tau}/m$.

(ii) From Lemma 2, we know that, conditioning on $\boldsymbol{z}$,

$$\hat{\theta} = \frac{1}{m}(\boldsymbol{\tau}^\top\mathbf{\Sigma}_0^{-1}\boldsymbol{\tau}) \xrightarrow{w.p.1} \frac{1}{m}(\boldsymbol{z}^\top\mathbf{\Sigma}_0^{-1}\boldsymbol{z}).$$

Recall that $\boldsymbol{z} = (z_1, \ldots, z_m)^\top$ has $MN(0, \theta\mathbf{\Sigma}_0)$ distribution. Thus, $\boldsymbol{z}^\top\mathbf{\Sigma}_0^{-1}\boldsymbol{z}/\theta$ has $\chi^2(m)$ distribution, and $\mathrm{E}(\boldsymbol{z}^\top\mathbf{\Sigma}_0^{-1}\boldsymbol{z}) = m\theta$. Then,

$$\lim_{n\to\infty}\mathrm{E}(\hat{\theta}) = \lim_{n\to\infty}\mathrm{E}[\mathrm{E}(\hat{\theta}|\boldsymbol{z})] = \lim_{n\to\infty}\mathrm{E}\left[\mathrm{E}\left(\frac{1}{m}\boldsymbol{\tau}^\top\mathbf{\Sigma}_0^{-1}\boldsymbol{\tau}|\boldsymbol{z}\right)\right] = \mathrm{E}\left(\frac{1}{m}\boldsymbol{z}^\top\mathbf{\Sigma}_0^{-1}\boldsymbol{z}\right) = \theta.$$

This shows that $\hat{\theta}$ is asymptotic unbiased. The model assumption (1) and the regularity conditions provides the exchangeability of limit and expectation in the above derivation.

(iii) Because $m\hat{\theta}/\theta$ is asymptotically $\chi^2(m)$ distributed, we know that, as $n \to \infty$, $\mathrm{E}(m\hat{\theta}/\theta) \to m$ and $\mathrm{Var}(m\hat{\theta}/\theta) \to 2m$. It follows from the properties of $\chi^2(m)$ distribution that, as $n, m \to \infty$,

$(m\hat{\theta}/\theta - m)/\sqrt{2m} \xrightarrow{d} N(0,1)$, or equivalently, $\sqrt{m}(\hat{\theta} - \theta) \xrightarrow{d} N(0, 2\theta^2)$. The proof is thus completed. □

5 Simulation Studies and Numerical Results

Simulation studies are conducted to evaluate the performance of estimates for the mean and variance-covariance parameters of model (1). To conduct a smaller scale estimation for understanding simulation results without running into messy computational problems, we fix the true values of model mean as $\boldsymbol{\mu} = \boldsymbol{A}\boldsymbol{\beta}$ where $\boldsymbol{A}$ is a known $m \times 2$ matrix,

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & \ldots & 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 1 & \ldots & 1 \end{pmatrix}^\top,$$

in which each row has $m/2$ 1's and $m/2$ 0's. The mean parameter $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$ and $\beta_1 = 10$ and $\beta_2 = 20$.

In the simulation study, random scale $\boldsymbol{z}$ is distributed as a m-dimension normal with zero mean and some specific variance-covariance structures described below. The cases with different sizes of m and n were also investigated. The sizes considered in the simulation studies are $n = 10$, 20, 30, 40, 50, 75, 100, 150 and 200, and $m = 0.6n$.

The following three cases of the variance-covariance matrices are considered in our simulation study.

Case I: $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = diag(\boldsymbol{X}\boldsymbol{\theta})$, where

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & \ldots & 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 1 & \ldots & 1 \end{pmatrix}^\top,$$

in which each row has $m/2$ 1's and $m/2$ 0's, and $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$, and $\theta_1 = 4$ and $\theta_2 = 1$.

Case II: $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \theta \boldsymbol{\Sigma}_0$, where $\theta = 1$ and

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} 4 & 0.5 & 0 & \ldots & 0 & 0 \\ 0.5 & 4 & 0.5 & \ldots & 0 & 0 \\ & \ldots & \ldots & \ldots & & \\ 0 & \ldots & 0 & 0.5 & 1 & 0.5 \\ 0 & \ldots & 0 & 0 & 0.5 & 1 \end{pmatrix},$$

where the first $m/2$ diagonal elements are 4, and the last $m/2$ diagonal elements are 1.

Case III:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \theta_1 & \theta_2 & 0 & \dots & 0 \\ \theta_2 & \theta_1 & \theta_2 & \dots & 0 \\ & \dots & & \dots & \\ 0 & \dots & \theta_2 & \theta_1 & \theta_2 \\ 0 & \dots & 0 & \theta_2 & \theta_1 \end{pmatrix},$$

where $\theta_1 = 4$ and $\theta_2 = 1$.

In all three cases, the mean parameter $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$ is estimated by

$$\hat{\beta}_1 = \frac{2}{m}\sum_{j=1}^{m/2} \bar{y}_{\cdot j}, \quad \text{and} \quad \hat{\beta}_2 = \frac{2}{m}\sum_{j=m/2+1}^{m} \bar{y}_{\cdot j},$$

but the variance-covariance parameters θ_1, θ_2 or θ are estimated differently case by case.

In Case I, it follows from (41) and (42) that the IEQL estimate $(\hat{\theta}_1, \hat{\theta}_2)$ is given by

$$\hat{\theta}_1 = \frac{2}{m}\sum_{j=1}^{m/2} \tau_j^2 \quad \text{and} \quad \hat{\theta}_2 = \frac{2}{m}\sum_{j=m/2+1}^{m} \tau_j^2,$$

where

$$\tau_j = \log\left(\frac{d_j(\hat{\mu}_j)}{n}\right) = \log\left(\frac{1}{nv(\bar{y}_{\cdot j})}\sum_{i=1}^{n}(y_{ij} - \bar{y}_{\cdot j})^2\right), \quad j = 1, \dots, m. \tag{43}$$

In Case II, it follows from Theorem 4 that the IEQL estimate $\hat{\theta}$ is given by $\hat{\theta} = \frac{1}{m}\hat{\boldsymbol{\tau}}^\top \boldsymbol{\Sigma}_0^{-1}\hat{\boldsymbol{\tau}}$, where $\boldsymbol{\tau} = (\tau_1, \dots, \tau_m)^\top$, and τ_j, $j = 1, \dots, m$, are given by (43).

In Case III, the IEQL estimate of $(\hat{\theta}_1, \hat{\theta}_2)$ is obtained by solving the IEQL estimating equations (23), i.e.,

$$tr\left(\boldsymbol{W}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{W}(\boldsymbol{\theta})}{\partial \theta_j}\right) - \hat{\boldsymbol{\tau}}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_j}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\hat{\boldsymbol{\tau}} = 0, \quad j = 1, 2.$$

The simulation procedure consists of the following steps.

Step 1. Generating a m-dimensional normal random variables $\boldsymbol{z} = (z_1, \dots, z_m)^\top$ from $MN(0, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$.

Step 2. Conditioning on $\boldsymbol{z}$, generating n *iid* normal random variables $\{y_{ij}\}$ from $N(\mu_j, e^{z_j})$, where $j = 1, \dots, m$, and $i = 1, \dots, n$, and then obtaining the IQEL estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$.

Step 3. Repeating 1000 times of Step 1 - Step 2, and taking the average of the 1000 estimates as the $\mathrm{E}(\hat{\boldsymbol{\beta}})$ and $\mathrm{E}(\hat{\boldsymbol{\theta}})$. Then, the bias of the IEQL estimates are evaluated by $\boldsymbol{\mu} - \sum_{l=1}^{1000}\hat{\boldsymbol{\beta}}_l/1000$ and $\boldsymbol{\theta} - \sum_{l=1}^{1000}\hat{\boldsymbol{\theta}}_l/1000$.

Step 4. Repeating Step 1 - Step 3 for the sizes $n = 10, 20, 30, 40, 50, 75, 100, 150$ and 200. The position number $m = 0.6n$ are 6, 12, 18, 24, 30, 45, 60, 90, and 120, respectively.

Step 5. Repeating Step 1 - Step 4 for Case I, II and III.

To investigate the simulation error, for each case, at $n = 20$ and $m = 12$, we repeat the 1000 simulations 10 times, and the simulation error is evaluated by the standard deviation from 10 replications of the estimates. The simulation errors are also reported for each estimate in each case. Tables 5.1, 5.2 and 5.3 present the numerical results of simulations of IEQL estimates for Case I, II and III, respectively, and $serr$ is "simulation error" from 10 replications at $n = 20$ and $m = 12$. In the tables, the biases ($bias$) are calculated in Step 3, and the standard deviation (sdv) is obtained from the 1000 simulation replicates.

In all three cases, the mean parameters of the IEQL estimates of β_1 and β_2 are very close to the true parameter values even with small sizes of n. Their bias values are less than the simulation errors when $n > 10$. The standard deviations are monotone decreasing as n (and m) increasing. These results agree with the consistency of the $\hat{\boldsymbol{\mu}} = \boldsymbol{A}\hat{\boldsymbol{\beta}}$ as stated in Theorem 1.

For Case I, Table 1 shows that The variance parameters θ_1 and θ_2 are over-estimated slightly in small sizes of n and m. In the cases of small size n, the problem of the slightly over-estimation is basically due to the approximate solution $\boldsymbol{\tau} = \log(\boldsymbol{d}(\hat{\boldsymbol{\mu}})/n)$, which is derived from a large size n.

However, the biases become smaller as n and m increasing. When $n \geq 40$ and $m \geq 24$, the biases of IEQL estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ are less than the simulation error, and the estimate can be considered as unbiased. Note that the standard deviations of $\hat{\theta}_1$ and $\hat{\theta}_2$ go to zero while sizes of n and m are increasing, i.e., $\hat{\boldsymbol{\theta}}$ is consistent as we stated in Theorems 2 and 3.

(Please put Table 1 here.)

Table 2 presents the simulation results for Case II. The results demonstrate very similar performance as that in Case I. Both mean parameters β_1 and β_2 are very well estimated even with small sizes of n. The scale parameter θ is over-estimated slightly with smaller n and m, but the mean value of the IEQL estimate is very close to the true parameter value when the sizes of n and m become large. The simulation results agree with the theoretical results of Theorem 4 that IEQL estimator $\hat{\boldsymbol{\theta}}$ is consistent.

(Please put Table 2 here.)

For Case III, Table 3 shows that, for estimating θ_1 and θ_2, the results are similar to the corresponding results obtained for Cases I and II. The variance parameter θ_1 is over-estimated slightly same as in Cases I and II with small n, but covariance parameter θ_2 is not over-estimated. In this case, θ_2 is well estimated even with small size of n. Although we do not have the theoretical result for IEQL estimate in this cases, both $\hat{\theta}_1$ and $\hat{\theta}_2$ appear to be consistent.

(Please put Table 3 here.)

6. Conclusion and Discussion

This article uses a random-scale based generalized linear model to describe possible randomness in variance components. The quasi-likelihood technique is extended to this model for allowing flexibility in distribution of modeling errors. Laplace's expansion method is a good tool to deal with the complicated integrated quasi-likelihood for getting estimates with expected asymptotic properties.

We obtained the estimating equations for the variance-covariance parameter $\boldsymbol{\theta}$. Approximations are performed to simplify the IEQL estimating equations. The results obtained from the approximated equations and the original estimating equations are the same in the large sample situation. Simulation results show that our estimating methods performance well. Under the general variance-covariance structure of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, the IEQL estimate $\hat{\boldsymbol{\mu}}$ of mean parameter is unbiased and consistent, and the asymptotic distribution of $\hat{\boldsymbol{\mu}}$ is obtained. Under several specific variance-covariance structures of $\boldsymbol{\theta}$, this paper shows that the IEQL estimators of $\boldsymbol{\theta}$ have the unbiasedness property asymptotically, and the asymptotic distributions of $\hat{\boldsymbol{\theta}}$ are obtained.

Appendix A - Review of the GLM, QL and EQL

Consider a GLM that $\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}$, where $\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)^\top$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$. The GLM framework assumes that the observations y_i, $i = 1, 2, \ldots, n$, are independent and have discrete or absolutely continuous distribution given by the density (McCullagh & Nelder, 1989)

$$f(y_i, \theta_i, \phi) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{a(\varphi)} + \tau(y_i, \varphi)\right),$$

where $a(\varphi)$, $b(\theta_i)$, and $\tau(y_i, \varphi)$ are some differentiable functions, and $b(\theta)$ is twice differentiable with $b''(\theta) > 0$. This distribution family includes some frequently used models, which have the common properties

$$\mathrm{E}(y_i) = b'(\theta_i), \quad \text{and} \quad \mathrm{Var}(y_i) = b''(\theta_i)a(\varphi).$$

When φ is known, the above density belongs to the exponential family. For example, let $a(\varphi) = 1$,

$b(\theta) = -\log(1-p) = \log[1+\exp(\theta)]$ and $\tau(y_i, \varphi) = 1$. This family becomes a binary distribution. Poisson distribution is another example, which is the case with $a(\varphi) = 1$, $b(\theta) = \exp(\theta)$, and $\tau(y_i, \varphi) = -\log(y_i)$. This family can also cover the normal distribution with $a(\varphi) = \sigma^2$, $b(\theta) = \theta^2/2$, and $\tau(y, \varphi) = -[\log(2\pi\sigma^2) + y^2/\sigma^2]/2$.

Wedderburn (1974) studied GLM through the following quasi-likelihood (QL) method based on the link between the mean and variance components. Consider the GLM, $\boldsymbol{Y} = \boldsymbol{\mu}(\boldsymbol{\beta}) + \boldsymbol{\varepsilon}$, where $\boldsymbol{\beta}$ is the vector of regression parameters. Define the quasi-likelihood $q(.)$ by the differential equation:

$$\frac{\partial q}{\partial \boldsymbol{\beta}} = \boldsymbol{Q}(\boldsymbol{\beta}) = \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}}\right)^{\top} \boldsymbol{V}^{-1}[\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta})],$$

where $\boldsymbol{V} = \mathrm{Var}(\boldsymbol{Y})$, which might also involve $\boldsymbol{\beta}$. This quasi-likelihood has similar properties as the regular log-likelihood such as

$$\mathrm{E}[\boldsymbol{Q}(\boldsymbol{\beta})] = 0, \ \mathrm{E}\left(\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{\beta}}\right) = \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}}\right)^{\top} \boldsymbol{V}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} = \mathrm{Cov}[\boldsymbol{Q}(\boldsymbol{\beta})].$$

Quasi-likelihood approach widened the scope of the GLM by replacing the distribution assumption of response variables with weaker assumptions, where only the first and second moments of the responses are needed. Nelder & Pregibon (1987) and Davidian & Carroll (1988) extended the quasi-likelihood approach to handle more flexible variance components.

Nelder & Pregibon (1987) considered a distribution family with mean and variance functions given by the following model:

$$\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{\mu}(\boldsymbol{\beta}) \ \text{ and } \ \mathrm{Var}(\boldsymbol{Y}) = \sigma^2 v^2(\boldsymbol{\mu}, \boldsymbol{z}, \boldsymbol{\theta}), \tag{44}$$

where $\boldsymbol{z}$ is a vector of some explanatory variables and $\boldsymbol{\theta}$ is a vector of other parameters. They proposed the following extended quasi-likelihood (EQL) as a generalization of the quasi-likelihood:

$$q^*(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma) = \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\log[2\pi\sigma^2 v^2(y_i, z_i, \boldsymbol{\theta})] - \frac{2}{\sigma^2}\int_{y_i}^{\mu_i(\boldsymbol{\beta})} \frac{y_i - x}{v^2(x, z_i, \boldsymbol{\theta})} dx\right)\right\}. \tag{45}$$

If $v(x, z, \boldsymbol{\theta}) = 1$, $q^*(.)$ is the regular normal likelihood. The EQL (45) is close related to the following pseudo-likelihood (Carroll & Ruppert, 1982):

$$l_{PL}(\beta, \theta, \sigma) = -n\log\sigma - \sum_{i=1}^{n}\left(\log[v(\mu_i(\boldsymbol{\beta}), z_i, \boldsymbol{\theta})] - \frac{(y_i - \mu_i(\boldsymbol{\beta}))^2}{2\sigma^2 v^2(\mu_i(\boldsymbol{\beta}), z_i, \boldsymbol{\theta})}\right),$$

which is asymptotically equivalent to weighted regression on squared residuals with an estimated weight. See Davidian & Carroll (1987, 1988) for more discussions about the relationship between EQL and pseudo-likelihood, and the motivations for the form of the EQL (45).

Appendix B - Estimates of Parameters in the Mean Function

Suppose that the mean vector $\boldsymbol{\mu}$ can be expressed by a p-dimension vector of unknown parameters $\boldsymbol{\beta}$ with a link function $f(\cdot)$ and a $m \times p$ constant matrix $\boldsymbol{A}$, i.e.,

$$\boldsymbol{\mu}(\beta) = \boldsymbol{f}(\boldsymbol{A}\beta) = (f(\boldsymbol{a}_1^\top \beta), \ldots, f(\boldsymbol{a}_m^\top \beta))^\top,$$

where $\boldsymbol{a}_i$ is the ith row vector of $\boldsymbol{A}$, i.e., $\boldsymbol{A}^\top = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m)$, $\boldsymbol{a}_i = (a_{i1}, \ldots, a_{ip})^\top$, $i = 1, 2, \ldots, m$, $m \geq p$ and $\boldsymbol{A}$ has column rank p.

In this case, note that

$$\frac{\partial \boldsymbol{d}}{\partial \beta} = \frac{\partial \boldsymbol{d}}{\partial \boldsymbol{\mu}} \cdot \frac{\partial \boldsymbol{\mu}}{\partial \beta}$$

and

$$\begin{aligned}
\frac{\partial \boldsymbol{\mu}}{\partial \beta} &= \left(\frac{\partial \mu_i}{\partial \beta_j}, i = 1, \ldots, m, j = 1, \ldots, p \right)_{m \times p} \\
&= (f'(\boldsymbol{a}_i^\top \boldsymbol{\beta}) a_{ij}, i = 1, \ldots, m, j = 1, \ldots, p)_{m \times p} \\
&= diag(\boldsymbol{f}'(\boldsymbol{A}\beta))\boldsymbol{A}.
\end{aligned}$$

The IEQL estimating equation $\partial l(\boldsymbol{d}(\boldsymbol{\mu}(\beta)), \boldsymbol{z}, \boldsymbol{\theta})/\partial \boldsymbol{\beta} = 0$ can be written as

$$\partial \boldsymbol{d}/\partial \boldsymbol{\beta} = (\partial \boldsymbol{d}/\partial \boldsymbol{\mu})(\partial \boldsymbol{\mu}/\partial \beta) = 0.$$

It follows from (10) that the equation $\partial \boldsymbol{d}/\partial \boldsymbol{\beta} = 0$ can be expressed in matrix form as

$$\boldsymbol{A}^\top \boldsymbol{V}(\boldsymbol{\mu}(\beta))(\bar{\boldsymbol{y}}_{(\cdot)} - \boldsymbol{f}(\boldsymbol{A}\beta)) = 0, \tag{46}$$

where

$$\begin{aligned}
\boldsymbol{V}(\boldsymbol{\mu}(\beta)) &= diag\left(\frac{\mu_1'(\boldsymbol{\beta})}{v(\mu_1(\boldsymbol{\beta}))}, \frac{\mu_2'(\boldsymbol{\beta})}{v(\mu_2(\boldsymbol{\beta}))}, \ldots, \frac{\mu_m'(\boldsymbol{\beta})}{v(\mu_m(\boldsymbol{\beta}))} \right) \\
&= diag\left(\frac{f'(\boldsymbol{a}_1^\top \boldsymbol{\beta})}{v(f(\boldsymbol{a}_1^\top \boldsymbol{\beta}))}, \frac{f'(\boldsymbol{a}_2^\top \boldsymbol{\beta})}{v(f(\boldsymbol{a}_2^\top \boldsymbol{\beta}))}, \ldots, \frac{f'(\boldsymbol{a}_m^\top \beta)}{v(f(\boldsymbol{a}_m^\top \beta))} \right),
\end{aligned}$$

$\bar{\boldsymbol{y}}_{(\cdot)} = (\bar{y}_{(.1)}, \bar{y}_{(.2)}, \ldots, \bar{y}_{(.m)})^\top$, and $\bar{y}_{(.j)} = \sum_{i=1}^n y_{ij}/n$, $j = 1, 2 \ldots, m$.

The equation (46) can be solved by the following extension of the GLS algorithm:

(1) Estimate $\boldsymbol{\mu}$ by the preliminary estimator $\hat{\boldsymbol{\mu}}^{(0)} = \bar{\boldsymbol{y}}_{(\cdot)}$.

(2) Obtain $\hat{\boldsymbol{\beta}}^{(q)}$ by solving equation

$$\boldsymbol{A}^\top \boldsymbol{V}(\hat{\boldsymbol{\mu}}^{(q)})(\bar{\boldsymbol{y}}_{(\cdot)} - \boldsymbol{f}(\boldsymbol{A}\beta)) = 0,$$

(3) Set $\hat{\boldsymbol{\mu}}^{(q+1)} = \boldsymbol{f}(\boldsymbol{A}\hat{\boldsymbol{\beta}}^{(q)})$ as the new preliminary estimator, return to step (2).

This process may be iterated a fixed number of times for convergence. At least one iteration is recommended.

Denoted the estimator for $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$. Because $\hat{\boldsymbol{\mu}}^{(0)} = \bar{\boldsymbol{y}}_{(\cdot)}$ is consistent as $n \to \infty$, we know that $\hat{\boldsymbol{\beta}}$ is also consistent as $n \to \infty$. The asymptotic distribution of $\hat{\boldsymbol{\beta}}$ depends on the link function $f(\cdot)$. But, in the case of the identical link function, i.e., $\boldsymbol{\mu} = \boldsymbol{A}\boldsymbol{\beta}$, then $\hat{\boldsymbol{\beta}} = (\boldsymbol{A}^\top \boldsymbol{V}\boldsymbol{A})^{-1}\boldsymbol{A}^\top \boldsymbol{V}\bar{\boldsymbol{y}}_{(\cdot)}$, where

$$\boldsymbol{V} = diag\left(\frac{1}{v(\bar{y}_{(.1)})}, \ldots, \frac{1}{v(\bar{y}_{(.m)})}\right).$$

It follows from (12) that the asymptotic corvariance matrix of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ can be calculated by

$$\Sigma_{\hat{\boldsymbol{\beta}}} = (\boldsymbol{A}^\top \boldsymbol{V}\boldsymbol{A})^{-1}\boldsymbol{A}^\top \boldsymbol{V}\Sigma_{\hat{\boldsymbol{\mu}}}\boldsymbol{V}^\top \boldsymbol{A}(\boldsymbol{A}^\top \boldsymbol{V}\boldsymbol{A})^{-1}.$$

Note that, in the above discussion of the asymptotic properties for mean parameters, we only assume that $n \to \infty$ and let m be fixed. If we also let $m \to \infty$, the regularty condition that

(R5.) Both

$$\lim_{m\to\infty} \frac{1}{m}(\boldsymbol{A}^\top \boldsymbol{V}\boldsymbol{A}) = \boldsymbol{B}(\boldsymbol{\beta}, \boldsymbol{\theta}) \quad \text{and} \quad \lim_{m\to\infty} \frac{1}{m}(\boldsymbol{A}^\top \boldsymbol{V}\Sigma_{\hat{\boldsymbol{\mu}}}\boldsymbol{V}^\top \boldsymbol{A}) = \boldsymbol{U}(\boldsymbol{\beta}, \boldsymbol{\theta})$$

exist

is necessary. In this case, as $n, m \to \infty$, $\hat{\boldsymbol{\beta}}$ obtained from (46) is also consistent, and the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ becomes

$$\Sigma_{\hat{\boldsymbol{\beta}}} = \boldsymbol{B}(\boldsymbol{\beta}, \boldsymbol{\theta})^{-1}\boldsymbol{U}(\boldsymbol{\beta}, \boldsymbol{\theta})\boldsymbol{B}(\boldsymbol{\beta}, \boldsymbol{\theta})^{-1}.$$

References

Breslow, N. E. and Clayton, D. G. (1993). Approximation Inference in Generalized Linear Mixed Models, *Journal of the American Statistical Association*, 88, 9-25.

Carlin, B. P. and Polson, N. G. (1991). Inference for non-conjugate Bayesian models using Gibbs Sampler. *Canadian Journal of Statistics*, 19, 399-405.

Carroll, R. J. and Ruppert, D. (1982). A Comparison between Maximum Likelihood and Gerenalized Least Squares in a heteroscedastic Linear Model, *Journal of the American Statistical Association*, 77, 878-892.

Chen, D., Lu, J. C., X. Huo, and Ming, Y. (2001). Robust Estimation with Estimating Equations for Nonlinear Random Coefficients Model, *Journal of Statistical Planning and Inference*, 37, 275-292.

Davidian, M. and Corral, R. J. (1988). A Note on Extended Quasi-likelihood, *Journal of Royal Statistical Society*, 50, 74-82.

Davidian, M. and Corral, R. J. (1987). Variance Function Estimation, *Journal of the American Statistical Association*, 82, 1079-1091.

Davidian, M. and Giltinan, D. M.(1995). Nonlinear Models for Repeated Measurement Data, *Chapman & Hall.*

Edward, F. V. (1996). A Note on the Use of Laplace's Approximation for Nonlinear Mixed effects Models. *Biometrika*, 83, 447-52.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of American Statistical Association*, 72, 320-340.

Host, G., Omre, H. and Switzer P. (1995). Spatial Interpolation Error for Monitoring Data. *Journal of the American Statistical Association*, 90, 853-860.

Johnson, N. L. and Kotz, S. (1969). Distribution in Statistics: Discrete Distributions, *Boston: Houghton Mifflin.*

Lambert, D. (1992). Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing. *Technometrics*, 34, 1-14.

Lange, K. L., Little, R. J. A. and Taylor, J. M. G. (1989). Robust statistical modeling using the t-distribution. *Journal of the American Statistical Association*, 84, 881-896.

Liang, k. and Zeger, S. L. (1986). Longitudinal Data Analysis Using Generalized Linear Models, *Biometrika*, 73, 13-22.

Lindsey, J. K. (1993). Models of Repeated Measurements. *Oxford, New York.*

Longford, N. T. (1993). Random Coefficient Models. *Oxford University Press, New York.*

Luke, T. and Joseph, B. K. (1990). Accurate Approximation for Posterior Moments and Marginal Densities. *Journal of the American Statistical Association*, 85, 652-63.

McCullagh, P. (1983). Quasi-likelihood Functions. *Annals of Statistics*, 11, 59-67.

McCullagh, P. and Nelder, J. A. (1989). Generalized linear models (2nd ed.). *Chapman and Hall, London.*

Nelder, J. A. and Pregibon, D. (1987). An Extended Quasi-likelihood Function. *Biometrika*, 74, 221-232.

Nelder, J. A. and Wedderburn, R. W. M. (1972) Generalized linear model. *J. R. Statistist. Soc.*, 135, 370-384.

Nering, E. D. (1970). Linear Algebra and Matrix Theory. *Wiley, New York*

Rao, C. R. (1973). Linear Statistical Inference and Its Applications, 2nd ed., *Wiley, New York.*

Serfling, R. J. (1980). Approximation Theorem of mathematical Statistical. *John Wiley & Sons, New York,*

Shun, Z. (1997). Another Look at the Salamander Mating Data: A Modified Laplace Approximation Approach. *Journal of the American Statistical Association*, 92, 341-9

Wedderburn, R. W. M. (1974). Quasi-likelihood Functions, generalized linear models and the Gauss-Newton method. *Biometrika*, 61, 439-447

Wolfinger, R. (1993). Laplace's Approximation for Nonlinear Mixed Models. *Biometrika*, 80, 791-795.

Table 1 Bias and Standard Deviation of the Estimates in Case I

($\beta_1 = 10$, $\beta_2 = 20$, $\theta_1 = 4$ and $\theta_2 = 1$)

	$\hat{\beta}_1$ $serr = 0.0027$		$\hat{\beta}_2$ $serr = 0.0011$		$\hat{\theta}_1$ $serr = 0.0233$		$\hat{\theta}_2$ $serr = 0.0051$	
$n(m)$	*bias*	*sdv*	*bias*	*sdv*	*bias*	*sdv*	*bias*	*sdv*
10 (6)	-0.0002	0.7961	0.0012	0.3962	0.1533	2.5893	0.1314	0.7762
20 (12)	0.0016	0.4087	0.0003	0.1782	0.0412	2.0030	0.0284	0.4889
30 (18)	0.0005	0.2841	-0.0003	0.1071	0.0187	1.4207	0.0074	0.4039
40 (24)	-0.0001	0.1803	0.0005	0.0908	0.0053	1.3187	0.0048	0.3464
50 (30)	0.0006	0.1643	-0.0001	0.0611	0.0030	1.1931	0.0028	0.2936
75 (45)	0.0008	0.1017	0.0001	0.0380	0.0022	0.9374	0.0019	0.2383
100(60)	0.0006	0.0687	0.0002	0.0314	0.0025	0.7898	0.0012	0.2027
150(90)	-0.0003	0.0451	-0.0002	0.0197	0.0021	0.6374	0.0009	0.1621
200(120)	-0.0004	0.0371	-0.0001	0.0167	0.0015	0.5499	0.0008	0.1381

Table 2 Bias and Standard Deviation of the Estimates in Case II

($\beta_1 = 10$, $\beta_2 = 20$, and $\theta = 1$)

	$\hat{\beta}_1$ $serr = 0.0011$		$\hat{\beta}_2$ $serr = 0.0027$		$\hat{\theta}$ $serr = 0.0046$	
$n(m)$	$bias$	sdv	$bias$	sdv	$bias$	sdv
10 (6)	-0.0001	0.8120	0.0061	0.8788	0.1156	1.0753
20 (12)	0.0004	0.3842	-0.0001	0.4142	0.0167	0.6511
30 (18)	-0.0001	0.2484	-0.0008	0.2287	0.0048	0.5736
40 (24)	0.0002	0.1943	-0.0005	0.2263	0.0030	0.4576
50 (30)	-0.0003	0.1651	0.0004	0.1965	0.0025	0.3905
75 (45)	0.0004	0.0874	-0.0002	0.1208	0.0024	0.3238
100(60)	-0.0001	0.0838	-0.0001	0.0676	0.0017	0.2869
150(90)	0.0004	0.0532	-0.0001	0.0460	0.0011	0.2248
200(120)	0.0001	0.0394	0.0002	0.0379	0.0007	0.1900

Table 3 Bias and Standard Deviation of the Estimates in Case III

($\beta_1 = 10$, $\beta_2 = 20$, $\theta_1 = 4$ and $\theta_2 = 1$)

	$\hat{\beta}_1$ $serr = 0.0014$		$\hat{\beta}_2$ $serr = 0.0021$		$\hat{\theta}_1$ $serr = 0.0162$		$\hat{\theta}_2$ $serr = 0.0131$	
$n(m)$	$bias$	sdv	$bias$	sdv	$bias$	sdv	$bias$	sdv
10 (6)	-0.0021	0.7809	-0.0003	0.5623	0.1368	2.0260	-0.0098	2.3129
20 (12)	0.0008	0.3483	-0.0009	0.3193	0.0933	1.4115	-0.0120	1.4131
30 (18)	0.0014	0.2363	0.0001	0.1371	0.0124	1.1200	0.0065	1.1373
40 (24)	-0.0013	0.1738	0.0014	0.1316	0.0046	0.8905	-0.0014	0.9694
50 (30)	-0.0009	0.1319	-0.0011	0.1268	0.0078	0.7972	-0.0087	0.7505
75 (45)	0.0012	0.1001	0.0008	0.0946	0.0052	0.6532	-0.0067	0.6562
100(60)	-0.0005	0.0674	0.0010	0.0551	0.0100	0.5524	0.0051	0.5526
150(90)	0.0002	0.0472	0.0000	0.0347	0.0022	0.4434	-0.0028	0.4418
200(120)	-0.0005	0.0445	0.0002	0.0176	0.0017	0.3819	0.0041	0.3730

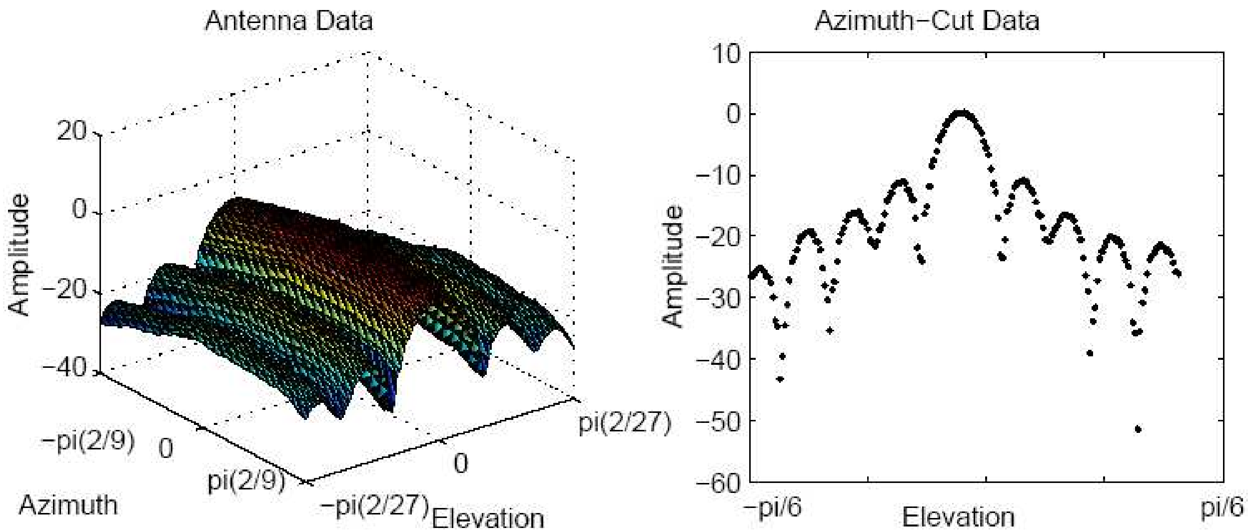

Figure 1: Antenna Signal Patterns

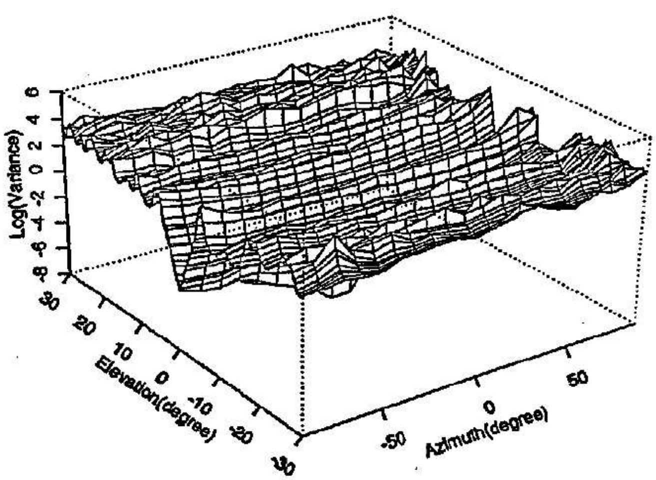

Figure 2: Log-variance Patterns from 20 Antennas