Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

OPTIMAL ESTIMATING EQUATIONS FOR MIXED EFFECTS MODELS WITH DEPENDENT OBSERVATIONS

Jeong-gun Park and I.V. Basawa
University of Georgia

Abstract

Optimal joint estimating equations for fixed and random parameters are derived via an extension of the Godambe criterion. Applications to autoregressive processes and generalized linear mixed models for Markov processes are discussed. Marginal optimal estimating functions for fixed parameters are also discussed.
Key Words: Optimal Estimating Functions; Generalized Linear Mixed Models; Autoregressive Processes; Markov Processes.

1 Introduction

Mixed effects models containing both fixed and random parameters are used extensively in both applied and methodological literature. A review of linear mixed models and their applications is given by Robinson(1991). Generalized linear mixed models are discussed by Breslow and Clayton(1993) among others. Work on the extension of mixed effects models to dependent observations appears relatively scarce. Our main goal in this paper is to develop optimal estimating equations for mixed effects models with dependent observations. See Basawa et al.(1997), Heyde(1997), and Godambe(1991) for recent literature on optimal estimating functions. Desmond(1997) gives an overview of estimating functions.

Suppose Y_t is a vector of observations on n individuals at time t, and $Y(t-1) = (Y_1, \dots, Y_{t-1})$. Conditional on $Y(t-1)$ and a random parameter γ, the density of Y_t is denoted by $p(y_t|y(t-1), \beta, \gamma)$ where β is a fixed parameter. Let $\pi(\gamma|\alpha)$ denote the (prior) density of γ which may depend on a parameter α. Suppose, for simplicity, α is known, and we wish to estimate β and γ from a sample $Y(T) = (Y_1, \dots, Y_T)$.

The likelihood function, conditional on γ, is given by

$$L(\beta, \gamma) = p(y_0)\Pi_{t=1}^{T} p(y_t|y(t-1), \beta, \gamma). \tag{1.1}$$

The joint density of γ and $Y(T)$ is

$$p(\gamma, y(T)|\alpha, \beta) = L(\beta, \gamma)\pi(\gamma|\alpha). \tag{1.2}$$

An intuitive way, often used in practice, to estimate the "mixed effects" β and γ, is to maximize $p(\gamma, y(T)|\alpha, \beta)$ with respect to β and γ and obtain formally the estimating equations :

$$\frac{\partial logL(\beta, \gamma)}{\partial \beta} = 0, \tag{1.3}$$

$$\frac{\partial logL(\beta, \gamma)}{\partial \gamma} + \frac{\partial log\pi(\gamma|\alpha)}{\partial \gamma} = 0. \tag{1.4}$$

Example 1. Linear mixed models

Take $T = 1$, and $Y_0 = y_0$ fixed(given). Suppose, conditional on γ, Y_1 is normal with mean vector $X'\beta + Z'\gamma$ and covariance matrix Σ, where X and Z are known covariate matrices. Further, assume that γ is a normal vector with mean zero and covariance matrix Γ. It is assumed, for simplicity, that Σ and Γ are known. Equations (1.3) and (1.4) then lead to the well known mixed linear model equations, see, for instance, Robinson(1991) for a review.

Example 2. Generalized Linear Mixed Models

Again, set $T = 1$, and suppose

$$E[Y_1|\beta, \gamma] = \mu(\beta, \gamma), \tag{1.5}$$

and

$$Var[Y_1|\beta, \gamma] = V(\beta, \gamma) = U(\mu(\beta, \gamma)). \tag{1.6}$$

Without further assumptions regarding the conditional density of Y_1 given γ, one may be interested in estimating β and γ. If we choose

$$h(\mu(\beta, \gamma)) = X'\beta + Z'\gamma, \tag{1.7}$$

for an appropriate link function $h(\cdot)$, and retain the normality assumption regarding γ (*i.e.* $\gamma \sim N\big(0, \Gamma\big)$), a penalized quasi-likelihood approach (see Breslow and Clayton(1993)) yields the estimating equations :

$$\left(\frac{\partial \mu}{\partial \beta'}\right)' V(\beta, \gamma)^{-1}(Y_1 - \mu) = 0, \tag{1.8}$$

$$\left(\frac{\partial \mu}{\partial \gamma'}\right)' V(\beta, \gamma)^{-1}(Y_1 - \mu) - \Gamma^{-1}\gamma = 0, \tag{1.9}$$

where $\mu = \mu(\beta, \gamma) = h^{-1}(X'\beta + Z'\gamma) = g(X'\beta + Z'\gamma)$, say. Equations (1.8) and (1.9) correspond to (1.3) and (1.4) if the conditional density of

Y_1 given γ is a member of an exponential family. See also Sutradhar and Godambe(1997), and Waclawiw and Liang(1993).

Analogous to (1.8) and (1.9), we now propose a general method based on the penalized quasi-likelihood equations for dependent data. Let

$$E[Y_t|Y(t-1),\beta,\gamma] = \mu_t(\beta,\gamma), \tag{1.10}$$

and

$$Var[Y_t|Y(t-1),\beta,\gamma] = V_t(\beta,\gamma). \tag{1.11}$$

Suppose further that the prior density of γ is known to be $\pi(\gamma|\alpha)$. Consider the estimating equations:

$$\sum_{t=1}^{T}\left(\frac{\partial\mu_t}{\partial\beta'}\right)' V_t^{-1}(\beta,\gamma)(Y_t-\mu_t) = 0, \tag{1.12}$$

$$\sum_{t=1}^{T}\left(\frac{\partial\mu_t}{\partial\gamma'}\right)' V_t^{-1}(\beta,\gamma)(Y_t-\mu_t) + \frac{\partial log\pi(\gamma|\alpha)}{\partial\gamma} = 0. \tag{1.13}$$

If the conditional density of Y_t given $Y(t-1)$ and γ belongs to an exponential family, the equations in (1.12) and (1.13) correpond to (1.3) and (1.4).

Suppose that γ has a prior density $\pi(\gamma|\alpha)$ with mean 0 and variance $\Gamma(\alpha)$. Consider a modified equation for γ :

$$\sum_{t=1}^{T}\left(\frac{\partial\mu_t}{\partial\gamma'}\right)' V_t^{-1}(\beta,\gamma)(Y_t-\mu_t) + E\left[\gamma\frac{\partial log\pi(\gamma|\alpha)}{\partial\gamma'}\right]'\Gamma(\alpha)^{-1}\gamma = 0. \tag{1.14}$$

If $\gamma \sim N\Big(0,\Gamma(\alpha)\Big)$, we have

$$\begin{aligned} E\left[\gamma\frac{\partial log\pi(\gamma|\alpha)}{\partial\gamma'}\right]'\Gamma(\alpha)^{-1}\gamma &= -\Gamma(\alpha)^{-1}\gamma \\ &= \frac{\partial log\pi(\gamma|\alpha)}{\partial\gamma}. \end{aligned}$$

Hence, in this special case, (1.14) reduces to (1.13).

We show in this paper that the equations in (1.12) and (1.14) are optimal estimating equations in the sense of a generalized Godambe criterion. The question of the performance (sampling properties) of the estimates obtained from (1.12) and (1.14) needs to be addressed via asymptotics. In the special case of linear mixed models (example 1), the estimates are known to be best linear unbiased predictors(BLUP), see Robinson(1991). In the general case, we may use an extension of the asymptotic optimality criterion discussed by Wefelmeyer(1996), at least for the estimation of the fixed effects parameter β. See also Heyde(1997). This topic will not be considered in this paper.

In general, the information matrix(or its estimate) corresponding to the optimal estimating functions can be used to compute the standard errors of the estimates.

The paper is organized as follows. The general optimality criterion for the estimation of mixed effects is introduced in Section 2. The optimal estimating equations are derived in Section 3. Section 4 discusses applications to autoregressive processes and generalized linear mixed models for Markov processes. Section 5 considers an alternative method based on a marginal model specification. Finally, Section 6 contains some concluding remarks on work in progress.

2 Optimality Criterion

Let $\mathcal{Y}$ be the sample space, $(\Theta_1 \times \Theta_2) \subset R^k$ the parameter space, where $\Theta_1 \subset R^m, m \leq k$. Assume that $\pi_\beta(\gamma)$ is a prior density of γ for a fixed β, where $\gamma \in \Theta_1$ and $\beta \in \Theta_2$. Let $\mathcal{L}$ be the set of all functions $g : \mathcal{Y} \times (\Theta_1 \times \Theta_2) \to R^k$ such that $E[g(Y, \gamma, \beta)|\beta] = 0, \ \forall \beta \in \Theta_2$, where the expectation is with respect to the joint distribution of Y and γ, for a fixed β. Assume that $E[g(Y, \gamma, \beta)g(Y, \gamma, \beta)']$ is finite for any $g \in \mathcal{L}$.

We shall use the notation

$$< g_1, g_2 >_\beta = E[g_1(Y, \gamma, \beta)g_2(Y, \gamma, \beta)'|\beta], \ \forall g_1, g_2 \in \mathcal{L}, \qquad (2.1)$$

where the expectation is with respect to the joint distribution of Y and γ, for a fixed β.

Let $p(Y, \gamma, \beta)$ denote the joint density of Y and γ. We assume throughout that the following conditions are satisfied :

$C.1$. For any $g \in \mathcal{L}$, g is differentiable with respect to both γ and β; Both $E[\frac{\partial g}{\partial \gamma}]$ and $E[\frac{\partial g}{\partial \beta}]$ exist.

$C.2$. The joint density $p(y, \gamma, \beta)$ is differentiable with respect to both γ and β. The support of $p(y, \gamma, \beta)$ does not depend on the parameters.

$C.3$. For any $g \in \mathcal{L}$, $E[g]$ is differentiable with respect to β under the integral sign.

$C.4$. The support of conditional density $p_\beta(y|\gamma)$ does not depend on the parameters γ and β. For any $g \in \mathcal{L}$, $E[g|\gamma]$ is differentiable with respect to γ under the integral sign; $E[\frac{\partial}{\partial \gamma} E(g|\gamma)]$ and $E[E(g|\gamma)\frac{\partial log \pi_\beta(\gamma)}{\partial \gamma}]$ exist.

The above conditions ensure the existence of various quantities in (2.4) and the validity of the derivation of the information function defined in (2.6) below.

Let $s = (s_1, s_2, \ldots, s_k)$, where

$$s_j = \frac{\partial log p(Y,\gamma,\beta)}{\partial\gamma_j},\ j = 1,\ldots,m,\ s_{m+l} = \frac{\partial log p(Y,\gamma,\beta))}{\partial\beta_l},\ l = 1,\ldots,k-m.$$

For any $g \in \mathcal{L}_0$, $\mathcal{L}_0 \subset \mathcal{L}$,

$$E[gs_j] = E\left[g\frac{\partial log p(Y,\gamma,\beta)}{\partial\gamma_j}\right],\ j = 1,\ldots,m.$$

Note that $p(Y,\gamma,\beta) = p_\beta(Y|\gamma)\pi_\beta(\gamma)$, where $p_\beta(Y|\gamma)$ is the conditional density of Y given γ with respect to an appropriate measure $\mu(y)$ and $\pi_\beta(\gamma)$ is the density of γ with respect to a measure $v(\gamma)$. We then have

$$\begin{aligned} E[gs_j] &= E\left[g\frac{\partial log p_\beta(Y|\gamma)}{\partial\gamma_j} + g\frac{\partial log \pi_\beta(\gamma)}{\partial\gamma_j}\right] \\ &= E\left[E\left\{g\frac{\partial log p_\beta(Y|\gamma)}{\partial\gamma_j}|\gamma\right\} + E\left\{g\frac{\partial log \pi_\beta(\gamma)}{\partial\gamma_j}|\gamma\right\}\right] \\ &= E\left[\frac{\partial E[g|\gamma]}{\partial\gamma_j}\right] - E\left[\frac{\partial g}{\partial\gamma_j}\right] + E\left[\frac{\partial log \pi_\beta(\gamma)}{\partial\gamma_j}E[g|\gamma]\right]. \quad (2.2) \end{aligned}$$

Now,

$$\begin{aligned} E[gs_{m+l}] &= E\left[g\frac{\partial log p(Y,\gamma,\beta)}{\partial\beta_l}\right] \\ &= \int g\frac{\partial p(y,\gamma,\beta)}{\partial\beta_l}d\mu(y)dv(\gamma),\ l = 1,\ldots,k-m. \\ &= -E\left[\frac{\partial g}{\partial\beta_l}\right],\ l = 1,\ldots,k-m. \quad (2.3) \end{aligned}$$

Combining (2.2) and (2.3), we obtain

$$E[gs'] = \left(-E\left[\frac{\partial g}{\partial\gamma'}\right] + E\left[\frac{\partial E[g|\gamma]}{\partial\gamma'} + E[g|\gamma]\frac{\partial log \pi_\beta(\gamma)}{\partial\gamma'}\right], -E\left[\frac{\partial g}{\partial\beta'}\right]\right). \quad (2.4)$$

Note that if $E[g|\gamma] = 0$, (2.4) reduces to

$$E[gs'] = \left(-E\left[\frac{\partial g}{\partial\gamma'}\right], -E\left[\frac{\partial g}{\partial\beta'}\right]\right). \quad (2.5)$$

We define the information function by

$$I_g(\beta) = E[gs']'E[gg']^{-1}E[gs'], \quad (2.6)$$

for any $g \in \mathcal{L}_0$, where $E[gs']$ is given by (2.4). The criterion of optimality is to maximize the information function $I_g(\beta)$ over $g \in \mathcal{L}_0$.

Definition 2.1 *A function g^* is an optimal estimating function in $\mathcal{L}_0$ if*

$$I_{g^*}(\beta) - I_g(\beta)$$

is nonnegative definite, for all $g \in \mathcal{L}_0$ and all β.

Case I (Fixed Effects) : $m = 0$, i.e. the model involves only the parameters of fixed effects.
When $m = 0$, $E[gs'] = -E[\frac{\partial g}{\partial \beta}]$. So the information function is given by

$$I_g(\beta) = E\left[\frac{\partial g}{\partial \beta}\right]' E[gg']^{-1} E\left[\frac{\partial g}{\partial \beta}\right].$$

Let g^* be the optimal estimating function for β in $\mathcal{L}_0$. Then

$$\begin{aligned} I_{g^*}(\beta) - I_g(\beta) &= E\left[\frac{\partial g^*}{\partial \beta}\right]' E[g^* g^{*'}]^{-1} E\left[\frac{\partial g^*}{\partial \beta}\right] \\ &\quad - E\left[\frac{\partial g}{\partial \beta}\right]' E[gg']^{-1} E\left[\frac{\partial g}{\partial \beta}\right]. \end{aligned}$$

It is easily verified that the nonnegative definiteness of $I_{g^*}(\beta) - I_g(\beta)$ is equivalent to that of

$$E(gg') - H(H^*)^{-1} E[g^* g^{*'}](H^{*'})^{-1} H'.$$

The latter turns out to be the optimality criterion given by Godambe(1985) when only fixed parameters are present with $H = E\left[\frac{\partial g}{\partial \beta}\right]$ and $H^* = E\left[\frac{\partial g^*}{\partial \beta}\right]$. See also Godambe(1960) and Durbin(1960).

Case II (Random Effects) : $m = k$, i.e. the model involves only the parameters of random effects.
When $m = k$, $E[gs'] = -E\left[\frac{\partial g}{\partial \gamma}\right] + E\left[\frac{\partial E[g|\gamma]}{\partial \gamma} + E[g|\gamma]\frac{\partial log \pi_\beta(\gamma)}{\partial \gamma}\right]$. Thus the information function is given by

$$I_g(\beta) = J' E[gg'] J,$$

where $J = -E\left[\frac{\partial g}{\partial \gamma}\right] + E\left[\frac{\partial E[g|\gamma]}{\partial \gamma} + E[g|\gamma]\frac{\partial log \pi_\beta(\gamma)}{\partial \gamma}\right]$. Let g^* be the optimal estimating function for γ in $\mathcal{L}_0$. Then we require

$$I_{g^*}(\beta) - I_g(\beta) = J^{*'} E[g^* g^{*'}] J^* - J' E[gg'] J$$

to be nonnegative definite. This turns out to be Chan and Ghosh(1998)'s optimality criterion when only the parameters of random effects are present and J^* is J with g replaced by g^*. See also Ferreira(1981, 1982) and Godambe(1998).

3 Optimal Estimating Equations

Here we derive optimal estimating functions for parameters of mixed effects.

Let $\mathcal{Y}$ be the sample space and F_i^Y the σ-field generated by a specified partition of $\mathcal{Y}$. Let h_i be a real-valued function of Y, γ, and β such that

$$E\left[h_i(Y,\gamma,\beta)|F_i^Y,\gamma,\beta\right] = 0, \ i = 1,2,\ldots,n.$$

Let $u(\cdot,\beta) : \Theta_1 \to R^m$ with $E[u(\cdot,\beta)] = 0$, for fixed β.

We consider the linear estimating space given by

$$\mathcal{L}_0 = \left\{ g : g = \sum_{i=1}^{n} a_i(\gamma,\beta)h_i + B(\beta)u(\gamma,\beta) \right\},$$

where $a_i(\gamma,\beta)$ is a $k \times 1$ vector which is measurable with respect to the σ-field F_i^Y and $B(\beta)$ is a $k \times m$ non-random matrix.

Let

$$a_i^*(\gamma,\beta) = -E\left[\left(\frac{\partial h_i}{\partial \gamma'}, \frac{\partial h_i}{\partial \beta'}\right)|F_i^Y,\gamma,\beta\right]' [Var(h_i|F_i^Y,\gamma,\beta)]^{-1} \tag{3.1}$$

and

$$B^*(\beta) = E\left[\left(u(\gamma,\beta)\frac{\partial log \pi_\beta(\gamma)}{\partial \gamma'}, -\frac{\partial u(\gamma,\beta)}{\partial \beta'}\right)\right]' [E(u(\gamma,\beta)u'(\gamma,\beta))]^{-1}. \tag{3.2}$$

Theorem 3.1 *Let*

$$g^* = \sum_{i=1}^{n} a_i^*(\gamma,\beta)h_i + B^*(\beta)u(\gamma,\beta).$$

Suppose that h_i's are mutually orthogonal in the sense that $E[h_i^* h_j^{*'}|F_i^Y,\gamma,\beta] = 0$ for all $i \neq j$, $i,j = 1,2,\ldots,n$, where $h_i^* = a_i^*(\gamma,\beta)h_i$.

Then g^* is an optimal estimating function in $\mathcal{L}_0$.

Proof: From Chan and Ghosh(1998), it is enough to show that g^* is an orthogonal projection of s into $\mathcal{L}_0$, i.e., $< g, s - g^* >= 0$ for any $g \in \mathcal{L}_0$. We have

$$< g, s - g^* > = < g, s > - < g, g^* > . \tag{3.3}$$

Note that, for the simplicity of notation, we use a_i and B for $a_i(\gamma,\beta)$ and $B(\beta)$, respectively in this proof. Now the first term on the righthand side of (3.3) can be expressed as

$$\begin{aligned} < g, s > &= \sum_{i=1}^{n} E[a_i h_i s'] + BE[us'] \\ &= \sum_{i=1}^{n} E[a_i E[h_i s'|F_i^Y,\gamma,\beta]] + BE[us']. \end{aligned}$$

From (2.4), since $E[h_i|F_i^Y,\gamma,\beta]=0$ for all i and any β,

$$E[h_i s'|F_i^Y,\gamma,\beta]=\left(-E\left[\frac{\partial h_i}{\partial\gamma'}|F_i^Y,\gamma,\beta\right],-E\left[\frac{\partial h_i}{\partial\beta'}|F_i^Y,\gamma,\beta\right]\right).$$

Also, from (2.4), we have

$$E[us'|F_i^Y,\gamma,\beta]=\left(u\frac{\partial log\pi_\beta(\gamma)}{\partial\gamma'},-\frac{\partial u}{\partial\beta'}\right).$$

Since the function u is only a function of γ and β, the conditional expectation of u, conditional on γ and β, is u itself. Thus we have

$$\begin{aligned}<g,s> &= -\sum_{i=1}^{n}E\left[a_iE\left\{\left(\frac{\partial h_i}{\partial\gamma'},\frac{\partial h_i}{\partial\beta'}\right)|F_i^Y,\gamma,\beta\right\}\right]\\ &\quad+BE\left[\left(u\frac{\partial log\pi_\beta(\gamma)}{\partial\gamma'},-\frac{\partial u}{\partial\beta'}\right)\right].\end{aligned}\tag{3.4}$$

Next the second term on the righthand side of (3.3)can be found as

$$\begin{aligned}<g,g^*> &= \sum_{i=1}^{n}\sum_{j=1}^{n}E[a_ih_i(a_j^*h_j)']+E[Buu'B^{*'}]\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}E\Big[E[a_ih_i(a_j^*h_j)'|F_i^Y,\gamma,\beta]\Big]+BE[uu']B^{*'}.\end{aligned}$$

Since the functions a_ih_i and $a_j^*h_j$ are orthogonal for all $i\neq j$,

$$\sum_{i=1}^{n}\sum_{j=1}^{n}E\Big[E[a_ih_i(a_j^*h_j)'|F_i^Y,\gamma,\beta]\Big]=\sum_{i=1}^{n}E\Big[E[a_ih_i(a_i^*h_i)'|F_i^Y,\gamma,\beta]\Big].$$

Substituting a_i^* and B^*, we get

$$\begin{aligned}<g,g^*> &= -\sum_{i=1}^{n}E\left[a_iE\left\{\left(\frac{\partial h_i}{\partial\gamma'},\frac{\partial h_i}{\partial\beta'}\right)|F_i^Y,\gamma,\beta\right\}\right]\\ &\quad+E\left[B\left(u\frac{\partial log\pi_\beta(\gamma)}{\partial\gamma'},-\frac{\partial u}{\partial\beta'}\right)\right].\end{aligned}\tag{3.5}$$

Hence, from 3.4 and 3.5, we have $<g,s>-<g,g^*>=0$. This complete the proof.

Independent Observations

If the elementary estimating functions $h_i, i=1,\dots,n$ are independent with $E[h_i(Y,\gamma,\beta)|\gamma,\beta]=0$, $\forall i=1,2,\dots,n$, it's not necessary to partition $\mathcal{Y}$ i.e. F_i^Y is $\mathcal{Y}$ itself. For an example, if $Y=(y_1,y_2,\dots,y_n)$ and y_i's are independent, we take h_i to be a function of y_i, γ and β such that $E[h_i|\gamma,\beta] = 0,\ \forall i$. Here h_i's are independent. Hence, the optimal estimating function in the space $\mathcal{L}_0$ is given by g^* in Theorem 3.1 with $a_i^*(\gamma,\beta)$ replaced by

$$a_i^{**}(\gamma,\beta) = -E\left[\left(\frac{\partial h_i}{\partial \gamma'}, \frac{\partial h_i}{\partial \beta'}\right)|\gamma,\beta\right]' [Var(h_i|\gamma,\beta)]^{-1}. \qquad (3.6)$$

Application to discrete time stochastic processes

As an example of dependent data we develop an optimal estimating function for a discrete time stochastic process. Let $\{Y_1, Y_2, \ldots, Y_T\}$ be a discrete time stochastic process. Let h_t be a real-valued function of $Y_1, \ldots, Y_t$, γ, and β such that

$$E\left[h_t(Y_1,\ldots,Y_t,\gamma,\beta)|F_{t-1}^Y,\gamma,\beta\right] = 0,\ t = 1,2,\ldots,T,$$

where F_{t-1}^Y is the σ-field generated by the past observations $Y_1, Y_2, \ldots, Y_{t-1}$. Let $u(\cdot,\beta) : \Theta_1 \to R^m$ with $E[u(\cdot,\beta)] = 0$, for fixed β.

We consider the estimating space

$$\mathcal{L}_0 = \left\{g : g = \sum_{t=1}^{T} a_{t-1}(\gamma,\beta)h_t + B(\beta)u(\gamma,\beta)\right\},$$

where $a_{t-1}(\gamma,\beta)$ is a $k \times 1$ vector measurable with respect to F_{t-1}^Y and $B(\beta)$ is a $k \times m$ non-random matrix.

Let

$$a_{t-1}^*(\gamma,\beta) = -E\left[\left(\frac{\partial h_t}{\partial \gamma'}, \frac{\partial h_t}{\partial \beta'}\right)|F_{t-1}^Y,\gamma,\beta\right]' [Var(h_t|F_{t-1}^Y,\gamma,\beta)]^{-1} \qquad (3.7)$$

and

$$B^*(\beta) = E\left[\left(u(\gamma,\beta)\frac{\partial log\pi_\beta(\gamma)}{\partial \gamma'}, -\frac{\partial u(\gamma,\beta)}{\partial \beta'}\right)\right]' [E(u(\gamma,\beta)u'(\gamma,\beta))]^{-1}. \qquad (3.8)$$

Theorem 3.2 *Let*

$$g^* = \sum_{t=1}^{T} a_{t-1}^*(\gamma,\beta)h_t + B^*(\beta)u(\gamma,\beta).$$

Then g^* is an optimal estimating function in $\mathcal{L}_0$.

The proof follows from Theorem 3.1 with $F_i^Y = F_{t-1}^Y$.

4 Applications

In this section, we will discuss two applications and demonstrate the derivation of optimal estimating functions.

4.1 Autoregressive Processes

Suppose that $y_t(j), t = 1, 2, \ldots, T,\ j = 1, 2, \ldots, n$ are observed at time t on the j^{th} subject from the first order autoregressive process(AR(1) process),

$$y_t(j) = \phi(j) y_{t-1}(j) + \varepsilon_t(j), \tag{4.1}$$

and

$$\phi(j) = x_j^{'}\beta + z_j^{'}\gamma, \tag{4.2}$$

where $\gamma \sim N\Big(0, \Gamma\Big)$ and $\varepsilon_t = (\varepsilon_t(1), \ldots, \varepsilon_t(n))', t = 1, 2, \ldots, T$, are *iid* random vectors with $E[\varepsilon_t] = 0$ and $Var[\varepsilon_t] = V_n(\alpha)$. Assume that γ and ε_t are independent for any t. Here β is a $(m \times 1)$ vector of fixed parameters and γ is a $(q \times 1)$ vector of random parameters, and x_j and z_j are vectors of known covariates. Note that we are not making distributional assumptions regarding ε_t other than the mean and variance assumptions.

Let

$$y_t = \Big(y_t(1), \ldots, y_t(n)\Big)^{'},\quad \phi = \Big(\phi(1), \phi(2), \ldots, \phi(n)\Big)^{'},$$

and

$$Y_{t-1} = \begin{pmatrix} y_{t-1}(1) & 0 & \cdots & 0 \\ 0 & y_{t-1}(2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & y_{t-1}(n) \end{pmatrix}.$$

Then (4.1) can be written in a vector form

$$y_t = Y_{t-1}\phi + \varepsilon_t,\ \ t = 1, 2, \ldots, T.$$

We consider optimal estimation for (γ, β) assuming α known. We choose elementary estimating functions h_t and u such that $E(h_t | F_{t-1}^y, \gamma, \beta) = 0$ and $E(u|\beta) = 0$. Let

$$h_t = y_t - Y_{t-1}(X^{'}\beta + Z^{'}\gamma) \ \text{ and } \ u = \gamma,$$

where $X = (x_1, \ldots, x_n)$ and $Z = (z_1, \ldots, z_n)$.

Consider the estimating space $\mathcal{L}_0 = \{g : g = \sum_{t=1}^{T} A_{t-1} h_t + Bu\}$. Now we let

$$A_{t-1}^{*} = -E\left\{\left(\frac{\partial h_h}{\partial \gamma^{'}}, \frac{\partial h_t}{\partial \beta^{'}}\right) | F_{t-1}^y, \gamma, \beta\right\}^{'} E[h_t h_t^{'} | F_{t-1}^y, \gamma, \beta]^{-1}$$

and

$$B^* = E\left[\left(u\frac{\partial log\pi_\beta(\gamma)}{\partial\gamma'}, -\frac{\partial u}{\partial\beta'}\right)\right]' E(uu')^{-1},$$

where $\pi(\gamma)$ is the prior distribution of γ. Then A^*_{t-1} and B^* are computed as

$$A^*_{t-1} = \begin{pmatrix} Z \\ X \end{pmatrix} Y_{t-1}V_n(\alpha)^{-1} \text{ and } B^* = \begin{pmatrix} -I_{q\times q} \\ 0 \end{pmatrix} \Gamma^{-1}.$$

Thus, the optimal estimating function for (γ, β) in the space $\mathcal{L}_0$ is given by $g^* = (g_1^*, g_2^*)$ where

$$g_1^* = \sum_{t=1}^{T} ZY_{t-1}V_n(\alpha)^{-1}\Big(y_t - Y_{t-1}(X'\beta + Z'\gamma)\Big) - \Gamma^{-1}\gamma, \tag{4.3}$$

and

$$g_2^* = \sum_{t=1}^{T} XY_{t-1}V_n(\alpha)^{-1}\Big(y_t - Y_{t-1}(X'\beta + Z'\gamma)\Big). \tag{4.4}$$

From (4.3) and (4.4), the optimal estimates $(\hat\gamma^*, \hat\beta^*)$ of (γ, β) are given by

$$\hat\gamma^* = \Delta^{-1}Z\left[\sum_{t=1}^{T} Y_{t-1}V_n(\alpha)^{-1}(y_t - Y_{t-1}X'\hat\beta^*)\right]$$

and

$$\begin{aligned}\hat\beta^* = & \left[X\sum_{t=1}^{T} Y_{t-1}V_n(\alpha)^{-1}Y_{t-1}\left\{X' - Z'\Delta^{-1}Z\Big(\sum_{t=1}^{T} Y_{t-1}V_n(\alpha)^{-1}Y_{t-1}\Big)X'\right\}\right]^{-1} \\ & \left[X\sum_{t-1}^{T} Y_{t-1}V_n(\alpha)^{-1}\left\{y_t - Y_{t-1}Z'\Delta^{-1}Z\Big(\sum_{t-1}^{T} Y_{t-1}V_n(\alpha)^{-1}y_t\Big)\right\}\right],\end{aligned}$$

where $\Delta = Z\left(\sum_{t=1}^{T} Y_{t-1}V_n(\alpha)^{-1}Y_{t-1}\right)Z' + \Gamma^{-1}$.

From (2.4), the information function corresponding to the optimal estimating function g^* is obtained as

$$\begin{aligned} I_{g^*}(\beta) &= E[g^*s']'E[g^*g^*]^{-1}E[g^*s'] \\ &= \begin{pmatrix} ZH_nZ' + \Gamma^{-1} & ZH_nX' \\ XH_nZ' & XH_nX' \end{pmatrix}, \end{aligned}$$

where $H_n = \sum_{t=1}^{T} E[Y_{t-1}V_n(\alpha)^{-1}Y_{t-1}]$.

As a special case, if the subjects are uncorrelated, i.e. $V_n(\alpha) = diag(\alpha_j, j = 1, \ldots, n)$, then the optimal estimating equations from (4.3) and (4.4) turn out to be

$$\sum_{j=1}^{n}\sum_{t=1}^{T} \alpha_j^{-1}y_{t-1}(j)z_j\Big[y_t(j) - y_{t-1}(j)(x_j'\beta + z_j'\gamma)\Big] - \Gamma^{-1}\gamma = 0,$$

and

$$\sum_{j=1}^{n}\sum_{t=1}^{T}\alpha_j^{-1}y_{t-1}(j)x_j\Big[y_t(j)-y_{t-1}(j)(x_j'\beta+z_j'\gamma)\Big]=0.$$

Thus the optimal estimates $(\hat{\gamma_I}^*, \hat{\beta_I}^*)$ of (γ, β) are obtained as

$$\hat{\gamma_I}^* = \Delta^{-1}\sum_{j=1}^{n}\sum_{t=1}^{T}\alpha_j^{-1}y_{t-1}(j)z_j\Big[y_t(j)-y_{t-1}(j)x_j'\hat{\beta_I}^*\Big],$$

and

$$\begin{aligned}\hat{\beta_I}^* &= \left[\sum_{j=1}^{n}\sum_{t=1}^{T}\alpha_j^{-1}y_{t-1}^2(j)x_j\left\{x_j'-z_j'\Delta^{-1}\Big(\sum_{j=1}^{n}\sum_{t=1}^{T}\alpha_j^{-1}y_{t-1}^2(j)z_jx_j'\Big)\right\}\right]^{-1}\\ &\quad \left[\sum_{j=1}^{n}\sum_{t=1}^{T}\alpha_j^{-1}y_{t-1}(j)x_j\left\{y_t(j)-y_{t-1}(j)z_j'\Delta^{-1}\Big(\sum_{j=1}^{n}\sum_{t=1}^{T}\alpha_j^{-1}y_{t-1}(j)y_t(j)z\right.\right.\end{aligned}$$

where $\Delta = \sum_{j=1}^{n}\sum_{t=1}^{T}\alpha_j^{-1}y_{t-1}^2(j)z_jz_j' + \Gamma^{-1}$.

We extend the method of obtaining optimal estimating functions to the p^{th} order autoregressive processes. Suppose that $y_t(j), t = 1, 2, \ldots, T, j = 1, 2, \ldots, n$, are observed at time t on the j^{th} subject from the p^{th} order autoregressive process(AR(p) process),

$$y_t(j) = \sum_{k=1}^{p}\phi_k(j)y_{t-k}(j) + \varepsilon_t(j), \tag{4.5}$$

and

$$\phi(j) = (\phi_1(j), \ldots, \phi_p(j))' = X_j'\beta + Z_j'\gamma, \tag{4.6}$$

where $\gamma \sim N_q\Big(0, \Gamma\Big)$ and $\varepsilon_t = (\varepsilon_t(1), \ldots, \varepsilon_t(n))', t = 1, \ldots, T$, are *iid* random vectors with $E[\varepsilon_t] = 0$ and $Var[\varepsilon_t] = V_n(\alpha)$. Assume that γ and ε_t are independent for any t. Here β is a $(m \times 1)$ vector of fixed parameters, γ is a $(q \times 1)$ vector of random parameters, and X_j and Z_j are matrices of known covariates.

Let

$$Y_{tp} = \begin{pmatrix} y_{pt}'(1) & 0 & \cdots & 0 \\ 0 & y_{pt}'(2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & y_{pt}'(n), \end{pmatrix}$$

and $y_{pt}(j) = (y_{t-1}(j), y_{t-2}(j), \ldots, y_{t-p}(j))'$. Elementary estimating functions h_t and u for γ and β are chosen to be

$$h_t = y_t - Y_{pt}(X'\beta + Z'\gamma),$$

and

$$u = \gamma,$$

where $X = (X(1), \ldots, X(n))$ and $Z = (Z(1), \ldots, Z(n))$. These elementary estimating functions satisfy the required condition of unbiasedness, i.e, $E[h_t|F_{t-1}^y, \gamma, \beta] = 0$ and $E[u|\beta] = 0$.

In the estimating space $\mathcal{L}_0 = \{g : g = \sum_{t=1}^T A_{t-1}h_t + Bu\}$, the optimal estimating function for (γ, β) is given by $g^{**} = (g_1^{**}, g_2^{**})$ where

$$g_1^{**} = \sum_{t=1}^{T} ZY_{pt}^{'}V_n(\alpha)^{-1}\Big(y_t - Y_{pt}(X^{'}\beta + Z^{'}\gamma)\Big) - \Gamma^{-1}\gamma,$$

and

$$g_2^{**} = \sum_{t=1}^{T} XY_{pt}^{'}V_n(\alpha)^{-1}\Big(y_t - Y_{pt}(X^{'}\beta + Z^{'}\gamma)\Big).$$

From g_1^{**} and g_2^{**}, the optimal estimates of γ and β are found as

$$\hat{\gamma}^{**} = \Delta^{-1}Z\left\{Y_{pt}^{'}V_n(\alpha)^{-1}(y_t - Y_{pt}X^{'}\hat{\beta}^{**})\right\}$$

and

$$\begin{aligned}\hat{\beta}^{**} \;=\; & \left[X\sum_{t=1}^{T} Y_{pt}^{'}V_n(\alpha)^{-1}Y_{pt}\left\{X^{'} - Z^{'}\Delta^{-1}Z\Big(\sum_{t=1}^{T} Y_{pt}^{'}V_n(\alpha)^{-1}Y_{pt}\Big)X^{'}\right\}\right]^{-1} \\ & \left[X\sum_{t=1}^{T} Y_{pt}^{'}V_n(\alpha)^{-1}\left\{y_t - Y_{pt}Z^{'}\Delta^{-1}Z\Big(\sum_{t=1}^{T} Y_{pt}^{'}V_n(\alpha)^{-1}y_t\Big)\right\}\right],\end{aligned}$$

where $\Delta = Z\Big(\sum_{t=1}^T Y_{pt}^{'}V_n^{-1}(\alpha)Y_{pt}\Big)Z^{'} + \Gamma^{-1}$.

4.2 Generalized Linear Mixed Models for Markov Processes

Suppose that $\{y_t(j), t = 0, 1, \ldots, T, j = 1, 2, \ldots, n\}$ is a markov process with a transition density

$$f(y_t(j)|y_{t-1}(j), \phi(j)) = c \cdot exp\Big\{\phi(j)m_t(y_t(j), y_{t-1}(j)) - q_t(y_{t-1}(j), \phi(j))\Big\} \tag{4.7}$$

and

$$\phi(j) = x_j^{'}\beta + z_j^{'}\gamma, \tag{4.8}$$

where $\gamma \sim N(0, \Gamma)$ and c is a function of $y_t(j)$ and $y_{t-1}(j)$. Here β is a $(m \times 1)$ vector of fixed parameters, γ is a $(r \times 1)$ vector of random parameters, and x_j and z_j are known vectors of covariates. It is assumed for simplicity that conditional on γ, the processes $\{y_t(j)\}$ are independent for different $j = 1, 2, \ldots, n$.

Let $X = (x_1, x_2, \ldots, x_n)$ and $Z = (z_1, z_2, \ldots, z_n)$. Then (4.8) can be written as

$$\phi = \begin{pmatrix} \phi(1) \\ \phi(2) \\ \vdots \\ \phi(n) \end{pmatrix} = X'\beta + Z'\gamma.$$

We consider joint optimal estimation for (γ, β), assuming that Γ is known. We choose elementary estimating functions h_t and u for γ and β such that $E[h_t|F^y_{t-1}, \gamma, \beta] = 0$ and $E[u|\beta] = 0$.

Let

$$h_t = m_t - \dot{q}_t \text{ and } u = \gamma, \text{ where}$$

$$m_t = \begin{pmatrix} m_t(y_t(1), y_{t-1}(1)) \\ m_t(y_t(2), y_{t-1}(2)) \\ \vdots \\ m_t(y_t(n), y_{t-1}(n)) \end{pmatrix} \text{ and } \dot{q}_t = \begin{pmatrix} \dot{q}_t(1) \\ \dot{q}_t(2) \\ \vdots \\ \dot{q}_t(n) \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial\phi(1)} q_t(y_{t-1}(1), \phi(1)) \\ \frac{\partial}{\partial\phi(2)} q_t(y_{t-1}(2), \phi(2)) \\ \vdots \\ \frac{\partial}{\partial\phi(n)} q_t(y_{t-1}(n), \phi(n)) \end{pmatrix}$$

Consider the estimating space $\mathcal{L}_0 = \{g : g = \sum_{t=1}^T A_t h_t + Bu\}$.

Let

$$\begin{aligned} A^*_t &= -E\left[\left(\frac{\partial h_t}{\partial\gamma'}, \frac{\partial h_t}{\partial\beta'}\right)|F^y_{t-1}, \gamma, \beta\right]' E[h_t h'_t|F^y_{t-1}, \gamma, \beta]^{-1}, \\ B^* &= E\left[\left(u\frac{\partial log\pi_\beta(\gamma)}{\partial\gamma'}, \frac{\partial u}{\partial\beta'}\right)\right]' E(uu')^{-1}. \end{aligned}$$

Then we have

$$A^*_t = (\ddot{q}_t(Z', X'))'\ddot{q}_t^{-1} = \begin{pmatrix} Z \\ X \end{pmatrix}$$

and

$$B^* = \Big(-E[\gamma\gamma']\Gamma^{-1}, 0\Big)'\Gamma^{-1} = \begin{pmatrix} -I \\ 0 \end{pmatrix}\Gamma^{-1},$$

where $\ddot{q}_t = diag\left\{\frac{\partial^2}{\partial\phi^2(j)} q_t(y_{t-1}(j), \phi(j)), j = 1, \ldots, n\right\}$. Thus, in the space $\mathcal{L}_0$ the joint optimal estimating function for (γ, β) is given by $g^* = (g^*_1, g^*_2)$, where

$$g^*_1 = \sum_{j=1}^n \sum_{t=1}^T z_j\Big(m_t(y_t(j), y_{t-1}(j)) - \dot{q}_t(j)\Big) - \Gamma^{-1}\gamma, \tag{4.9}$$

and

$$g^*_2 = \sum_{j=1}^n \sum_{t=1}^T x_j\Big(m_t(y_t(j), y_{t-1}(j)) - \dot{q}_t(j)\Big). \tag{4.10}$$

The optimal estimates $(\hat{\gamma}^*, \hat{\beta}^*)$ for (γ, β) can be obtained by solving the equation $g^* = 0$ for γ and β simultaneously.

From (2.4), the information function corresponding to the optimal estimating function g^* is obtained as

$$\begin{aligned} I_{g^*}(\beta) &= E[g^* s']' E[g^* g^*]^{-1} E[g^* s'] \\ &= \begin{pmatrix} ZH_nZ' + \Gamma^{-1} & ZH_nX' \\ XH_nZ' & XH_nX' \end{pmatrix}, \end{aligned}$$

Where $H_n = \sum_{t=1}^{T} E[\ddot{q}_t]$

We now illustrate the model by two examples.

Example 1. Consider an AR(1) process with normal errors, i.e.

$$y_t(j) = \phi(j) y_{t-1}(j) + \varepsilon_t(j) \tag{4.11}$$

and

$$\phi(j) = x_j'\beta + z_j'\gamma, \tag{4.12}$$

where $\varepsilon_t(j) \sim$ *indep* $N\Big(0, \sigma_j^2\Big)$ and $\gamma \sim N\Big(0, \Gamma\Big)$. The conditional density is given by

$$f(y_t(j)|y_{t-1}(j), \phi(j)) = c \cdot exp\left\{-\frac{1}{2\sigma_j^2}\Big(y_t(j) - \phi(j)y_{t-1}(j)\Big)^2\right\}, \tag{4.13}$$

where $c = \frac{1}{\sqrt{2\pi}\sigma_j}$.

Since (4.13) can be written as

$$f(y_t(j)|y_{t-1}(j), \phi(j)) = c^* \cdot exp\left\{\phi(j)\sigma_j^{-2} y_t(j) y_{t-1}(j) - \frac{1}{2}\sigma_j^{-2}\phi^2(j) y_{t-1}^2(j)\right\},$$

we have

$$m_t(y_t(j), y_{t-1}(j)) = \sigma_j^{-2} y_t(j) y_{t-1}(j)$$

and

$$q_t(y_t, y_{t-1}(j)) = \frac{1}{2}\sigma_j^{-2} y_{t-1}^2(j)\phi^2(j).$$

This gives $\mu_t(j) = \sigma_j^{-2} y_{t-1}^2(j)\phi(j)$ and $V_t(j) = \sigma_j^{-2} y_{t-1}^2(j)$. Hence the optimal estimating function for (γ, β) is given by $g^* = (g_1^*, g_2^*)$, where

$$g_1^* = \sum_{j=1}^{n}\sum_{t=1}^{T} \sigma_j^{-2} y_{t-1}(j) z_j \Big\{y_t(j) - y_{t-1}(j)(x_j'\beta + z_j'\gamma)\Big\} - \Gamma^{-1}\gamma$$

and

$$g_2^* = \sum_{j=1}^{n}\sum_{t=1}^{T} \sigma_j^{-2} y_{t-1}(j) x_j \Big\{y_t(j) - y_{t-1}(j)(x_j'\beta + z_j'\gamma)\Big\}.$$

The optimal estimates of γ and β are obtained by solving the equations $g_1^* = 0$ and $g_2^* = 0$ which yields the same estimates as $\hat{\gamma}_I^*$ and $\hat{\beta}_I^*$ in section 4.1.

Example 2. Let $\{y_t(j)\}, t = 0, 1, 2, \ldots,$ be a Markov chain, for each $j = 1, 2, \ldots, n,$ defined on the binary state space $\{0, 1\}$.

Denote

$$\pi_{tj} = P\Big(y_t(j) = 1 | y_{t-1}(j)\Big), \quad \theta_{tj} = logit(\pi_{tj}) = log\left(\frac{\pi_{tj}}{1-\pi_{tj}}\right).$$

We then have

$$\pi_{tj} = \frac{exp\{\theta_{tj}\}}{1+exp\{\theta_{tj}\}}, \quad 1 - \pi_{tj} = \frac{1}{1+exp\{\theta_{tj}\}}.$$

Consider the model

$$\theta_{tj} = \beta_0 + \phi_j y_{t-1}(j), \tag{4.14}$$

where $\phi_j = x_j'\beta + z_j'\gamma$. Conditionally on γ, the Markov chains $\{y_t(j)\}, j = 1, \ldots, n,$ are assumed to be independent. Suppose $\gamma \sim N\Big(0, \Gamma\Big)$. Conditional on γ, the transition densities are given by

$$\begin{aligned} p\Big(y_t(j)|y_{t-1}(j), \gamma\Big) &= \pi_{tj}^{y_t(j)}(1-\pi_{tj})^{(1-y_t(j))} \\ &= exp\Big\{\theta_{tj}y_t(j) - log\Big(1 + exp\{\theta_{tj}\}\Big)\Big\} \\ &= exp\Big\{\beta_0 y_t(j) + \phi_j y_{t-1}(j) y_t(j) \\ &\qquad - log\Big(1 + exp\{\beta_0 + \phi_j y_{t-1}(j)\}\Big)\Big\}. \end{aligned}$$

The optimal estimating functions for (γ, β, β_0) then reduce to

$$g_1^* = \sum_{j=1}^{n}\sum_{t=1}^{T} z_j y_{t-1}(j)(y_t(j) - \pi_{tj}) - \Gamma^{-1}\gamma,$$

$$g_2^* = \sum_{j=1}^{n}\sum_{t=1}^{T} x_j y_{t-1}(j)(y_t(j) - \pi_{tj}),$$

and

$$g_3^* = \sum_{j=1}^{n}\sum_{t=1}^{T}(y_t(j) - \pi_{tj}).$$

The third estimating function g_3^* here corresponds to the common intercept β_0 in the model for θ_{tj}.

5 Marginal Quasi-likelihood Estimation

Let $\bar{\mu}_t(y(t-1), \beta) = E[y_t | y_{t-1}, \ldots, y_1, \beta]$, and $\bar{V}_t(y(t-1), \beta) = Var[y_t | y_{t-1}, \ldots, y_1, \beta]$. The optimal estimating equation for β is

$$\sum_{t=1}^{T} \left(\frac{\partial \bar{\mu}_t(y(t-1), \beta)}{\partial \beta'} \right)' \bar{V}_t^{-1}(y(t-1), \beta)(y_t - \bar{\mu}_t(y(t-1), \beta)) = 0. \quad (5.1)$$

Let $\bar{\beta}$ denote a solution of (5.1).

Define $\mu_t(y(t-1), \gamma, \beta) = E[y_t | y_{t-1}, \ldots, y_1, \gamma, \beta]$, and $V_t(y(t-1), \gamma, \beta) = Var[y_t | y_{t-1}, \ldots, y_1, \gamma, \beta]$. For fixed β, let

$$h_t(\gamma, \beta) = y_t - \mu_t(y(t-1), \gamma, \beta)$$

denote an elementary estimating function for γ.

Suppose γ has a prior density $\pi(\gamma|\alpha)$ with mean 0 and variance Γ. Consider the estimating equation

$$\sum_{t=1}^{T} \left(\frac{\partial \mu_t(y(t-1), \gamma, \beta)}{\partial \gamma'} \right)' V_t^{-1}(y(t-1), \gamma, \beta)\Big(y_t - \mu_t(y(t-1), \gamma, \beta)\Big)$$
$$+ E\left[\gamma \frac{\partial log \pi(\gamma|\alpha)}{\partial \gamma'}\right]' \Gamma^{-1}\gamma = 0. \quad (5.2)$$

For fixed β, (5.2) is an optimal estimating equation when only the random parameter γ is present in the model. We now substitute the marginal quasi-likelihood estimate $\bar{\beta}$ in (5.2), when β is unknown. Denote the resulting estimate of γ(obtained from (5.2) after replacing β by $\bar{\beta}$) by $\tilde{\gamma}$. The estimates $\bar{\beta}$ and $\tilde{\gamma}$ will be referred to as the marginal quasi-likelihood estimates(MQL). Note, however, that for fixed β, (5.2) is the same as the estimating equation (1.14) for γ corresponding to the joint optimal estimation of β and γ. It must be noted that (5.1) and (5.2) are not jointly optimal; on the other hand, (5.1) is optimal for β with respect to the marginal density of $y = (y_t, \ldots, y_1)$ and (5.2) is optimal for γ(when β is known) with respect to the joint density of y and γ.

Application to Autoregressive Processes: Suppose that $y_t(j), t = 1, 2, \ldots, T$, $j = 1, 2, \ldots, n$ are observed at time t on the j^{th} subject from the first order autoregressive process(AR(1) process):

$$y_t(j) = \phi(j) y_{t-1}(j) + \varepsilon_t(j), \quad (5.3)$$

and

$$\phi(j) = x_j'\beta + z_j'\gamma, \quad (5.4)$$

where $\gamma \sim N_q\Big(0, \Gamma\Big)$ and $\varepsilon_t = (\varepsilon_t(1), \ldots, \varepsilon_t(n))', t = 1, 2, \ldots, T$, are *iid* random vectors with $E[\varepsilon_t] = 0$ and $Var[\varepsilon_t] = V_n(\alpha)$. Assume that γ and ε_t

are independent for any t. Here β is a $(m \times 1)$ vector of fixed parameters, γ is a $(q \times 1)$ vector of random parameters, and x_j and z_j are vectors of known covariates.

Let

$$y_t = \Big(y_t(1), \ldots, y_t(n)\Big)',$$

and $Y_{t-1} = diag(y_{t-1}(j), j = 1, \ldots, n)$, i.e., Y_{t-1} is a diagonal matrix with $(j,j)^{th}$ diagonal element $y_{t-1}(j)$, $j = 1, \ldots, n$.

We consider optimal estimation for β, treating γ as a nuisance parameter when α is known. We assume that ε_t has a normal distribution for our illustration. Now

$$\bar{\mu}_t(y(t-1), \beta) = Y_{t-1}\{X'\beta + Z'E[\gamma|y_{t-1}, \ldots, y_1]\},$$

and

$$\bar{V}_t(y(t-1), \beta) = Y_{t-1}Z'Var[\gamma|y_{t-1}, \ldots, y_1]ZY_{t-1} + V_n(\alpha),$$

where $X = (x_1, \ldots, x_n)$ and $Z = (z_1, \ldots, z_n)$.

To find the conditional expectation and variance, we derive the posterior density $\pi(\gamma|y_{t-1}, \ldots, y_1)$ of ϕ, conditional on the past observations $y_{t-1}, y_{t-2}, \ldots, y_1$.

$$\begin{aligned}
\pi(\gamma|y_{t-1}, \ldots, y_1) &\propto p(y_{t-1}, \ldots, y_1|\gamma)\pi(\gamma) \\
&\propto exp\Bigg\{ -\frac{1}{2}\sum_{r=1}^{t-1}\Big(y_r - Y_{r-1}(X'\beta + Z'\gamma)\Big)' V_n(\alpha)^{-1} \\
&\qquad \Big(y_r - Y_{r-1}(X'\beta + Z'\gamma)\Big)\Bigg\} \cdot exp\Big\{-\frac{1}{2}(\gamma'\Gamma^{-1}\gamma)\Big\} \\
&= exp\Bigg\{ -\frac{1}{2}\Bigg[(\gamma - \Delta_{t-1}^{-1}E_{t-1})'\Delta_{t-1}(\gamma - \Delta_{t-1}^{-1}E_{t-1}) \\
&\qquad - E_{t-1}'\Delta_{t-1}^{-1}E_{t-1} + \sum_{r=1}^{t-1}\beta' XY_{r-1}V_n(\alpha)^{-1}Y_{r-1}X'\beta \\
&\qquad - \sum_{r=1}^{t-1}\beta' XY_{r-1}V_n(\alpha)^{-1}y_r - \sum_{r=1}^{t-1}y_r'V_n(\alpha)^{-1}Y_{r-1}X'\beta \\
&\qquad + \sum_{r=1}^{t-1}y_r^{-1}V_n(\alpha)^{-1}y_r\Bigg]\Bigg\},
\end{aligned}$$

where

$$\Delta_{t-1} = \sum_{r=1}^{t-1} ZY_{r-1}V_n(\alpha)^{-1}Y_{r-1}Z' + \Gamma^{-1}$$

and

$$E_{t-1} = \sum_{r=1}^{t-1} \left(ZY_{r-1}V_n(\alpha)^{-1}y_r - ZY_{r-1}V_n(\alpha)^{-1}Y_{r-1}X'\beta \right).$$

We then have $\bar{\mu}_t(y(t-1), \beta)$ and $\bar{V}_t(y(t-1), \beta)$ given by

$$\bar{\mu}_t(y(t-1), \beta) = Y_{t-1}\{X'\beta + Z'\Delta_{t-1}^{-1}E_{t-1}\},$$

and

$$\bar{V}_t(y(t-1), \beta) = Y_{t-1}Z'\Delta_{t-1}^{-1}ZY_{t-1} + V_n(\alpha).$$

Thus the optimal estimating equation for β is given as (5.1) and the optimal estimating function is computed as

$$g^* = \sum_{t=1}^{T} A_{t-1}^* B_{t-1}^{*-1} \left(y_t - Y_{t-1}\{X'\beta + Z'\Delta_{t-1}^{-1}E_{t-1}\} \right),$$

where

$$A_{t-1}^* = XY_{t-1} - X\Big(\sum_{r=1}^{t-1} Y_{r-1}V_n(\alpha)^{-1}Y_{r-1}\Big)Z'\Delta_{t-1}^{-1}ZY_{t-1}$$

and

$$B_{t-1}^* = Y_{t-1}Z'\Delta_{t-1}^{-1}ZY_{t-1} + V_n(\alpha).$$

Hence, from the above equation, the optimal estimate of β is found as

$$\begin{aligned} \bar{\beta} &= \left[\sum_{t=1}^{T} A_{t-1}^* B_{t-1}^{*-1} A_{t-1}^{*'} \right]^{-1} \\ &\quad \cdot \left[\sum_{t=1}^{T} A_{t-1}^* B_{t-1}^{*-1} \left\{ y_t - Y_{t-1}Z'\Delta_{t-1}^{-1}Z\Big(\sum_{r=1}^{t-1} Y_{r-1}V_n(\alpha)^{-1}y_r\Big) \right\} \right]. \end{aligned}$$

From (2.4), the information function corresponding to the optimal estimating function g^* is obtained as

$$\begin{aligned} I_{g^*}(\beta) &= E\left[\frac{\partial g^*}{\partial \beta'}\right]' E[g^* g^{*'}]^{-1} E\left[\frac{\partial g^*}{\partial \beta'}\right] \\ &= \sum_{t=1}^{T} E[A_{t-1}^* B_{t-1}^{*-1} A_{t-1}^{*'}]. \end{aligned}$$

Suppose that, for fixed β, we wish to estimate γ. We choose an elementary estimating function h_t for γ as

$$h_t = y_t - Y_{t-1}(X'\beta + Z'\gamma).$$

Then the optimal estimating equation for γ is given as (5.2) :

$$\sum_{t=1}^{T}(Y_{t-1}Z')'V_n(\alpha)^{-1}\Big(y_t - Y_{t-1}(X'\beta + Z'\gamma)\Big) - \Gamma^{-1}\gamma = 0.$$

Thus the above equation yields an estimate $\hat{\gamma}^*$ of γ when β is known.

$$\begin{aligned}\hat{\gamma}^* &= \left[\sum_{t=1}^{T} ZY_{t-1}V_n(\alpha)^{-1}Y_{t-1}Z' + \Gamma^{-1}\right]^{-1} \\ &\cdot \left[\sum_{t=1}^{T} ZY_{t-1}V_n(\alpha)^{-1}(y_t - Y_{t-1}X'\beta)\right]\end{aligned}$$

The estimate $\hat{\gamma}^*$ is optimal for γ when β is known. When β is unknown, we replace it by $\bar{\beta}$ in $\hat{\gamma}^*$ to obtain $\tilde{\gamma}$.

6 Concluding Remarks

In this paper, we have derived optimal estimating equations for estimating fixed and random parameters in mixed effects models with dependent observations. Among the issues that are not addressed in this paper are : (i) asymptotic distribution theory, (ii) estimation of variance components, and (iii) computational aspects. We hope to return to these topics in future work. Here, we content ourselves by offering the following remarks on these important issues.

(i) Asymptotic Distribution Theory

The consistency and asymptotic normality of the marginal quasi-likelihood estimate $\bar{\beta}$ in Section 5 can be established, under regularity conditions, from the general theory discussed by Heyde(1997). Appropriate extensions to develop asymptotic distribution theory for the joint estimation of β and γ are needed. Moreover, work on asymptotic efficiency of the estimates will be useful.

(ii) Estimation of Variance Components

Assuming $\gamma \sim N(0, \Gamma)$, one can estimate Γ from the marginal quasi-likelihood in Section 5. An extension of the REML approach discussed by Breslow and Clayton(1993) can be used in practice.

(iii) Computational Aspects

Extensions of Fisher scoring method discussed by Breslow and Clayton(1993) need to be developed. For most of the examples in our paper, however, we have obtained explicit solutions of the estimating equations.

Consistent parameter estimates of the information matrices can easily be obtained and hence the estimates of the standard errors of the estimates can be computed, in principle.

References

Basawa, I.V., Godambe, V.P., and Taylor, R.L., (Eds.) (1997). *Selected Proceedings of the Symposium on Estimating Functions.* IMS Lecture Notes-Monograph Series, Vol. 32.

Breslow, N.E. and Clayton, D.G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association* 88, 421, 9-25.

Chan, S. and Ghosh, M. (1998). Orthogonal Projections and the Geometry of Estimating functions. *Journal of Statistical Planning and Inference* 67, 227-245.

Desmond, A.F. (1997). Optimal Estimating Functions, Quasi-likelihood and Statistical Modelling. *Journal of Statistical Planning and Inference* 60, 77-121.

Durbin, J. (1960). Estimation of Parameters in Time-series Regression Models. *Journal of Royal Statistical Soc.* B22, 139-153.

Ferreira, P.E. (1981). Extending Fisher's Measure of Information. *Biometrika* 68, 3, 695-698.

Ferreira, P.E. (1982). Estimating Equations in the Presence of Prior Knowledge. *Biometrika* 69, 3, 667-669.

Godambe, V.P. (1960). An Optimum Property of Regular Maximum Likelihood Estimation. *Annals of Mathematical Statistics* 31, 1208-1211.

Godambe, V.P. (1985). The Foundations of Finite Sample Estimation in Stochastic Processes. *Biometrika* 72, 2, 419-428.

Godambe, V.P. (Ed.) (1991). *Estimating Functions.* Oxford University Press, Oxford.

Godambe, V.P. (1998). Linear Bayes and Optimal Estimation. To appear in *Annals of the Institute of Statistical Mathematics.*

Heyde, C.C (1997). *Quasi-Likelihood and Its Application, A General Approach to Optimal Parameter Estimation.* Springer-Verlag, New York.

Robinson, G.K. (1991). That BLUP is a Good Thing: The Estimation of Random Effects. *Statistical Science* 6, 1, 15-51.

Sutradhar, B.C. and Godambe, V.P. (1997). On Estimating Function Approach in the Generalized Linear Mixed Model. *IMS Selected Proceedings of the Symposium on Estimating Functions(Basawa I.V, Godame, V.P., and Taylor, R.L. Eds.)* 32, 193-213.

Waclawiw, M.A. and Liang, K.Y. (1993) Prediction of Random Effects in the Generalized Linear Model. *Journal of the American Statistical Association* 88, 421, 171-178.

Wefelmeyer, W. (1996). Quasi-likelihood Models and Optimal Inference. *Annals of Statistics* 24, 1, 405-422.