Journal of the Korean Statistical Society (1999), **28**: *2, pp 225–234*

Efficient Prediction in the Semi-parametric Non-linear Mixed effect Model†

Beong-Soo So[1]

Abstract

We consider the following *semi-parametric* non-linear mixed effect regression model : $y_i = f(x_i;\beta)+\sigma u(x_i)+\sigma\epsilon_i, i=1,\cdots,n$, $y^* = f(x;\beta)+\sigma u(x)$ where $y' = (y_1,\cdots,y_n)$ is a vector of n observations, y^* is an unobserved new random variable of interest, $f(x;\beta)$ represents fixed effect of known functional form containing unknown parameter vector $\beta' = (\beta_1,\cdots,\beta_p)$, $u(x)$ is a random function of mean zero and the known covariance function $r(\cdot,\cdot)$, $\epsilon' = (\epsilon_1,\cdots,\epsilon_n)$ is the set of uncorrelated measurement errors with zero mean and unit variance and σ is an unknown dispersion (scale) parameter. On the basis of *finite-sample, small-dispersion* asymptotic framework, we derive an absolute lower bound for the asymptotic mean squared errors of prediction (AMSEP) of the *regular-consistent* non-linear predictors of the new random variable of interest y^* . Then we construct an optimal predictor of y^* which attains the lower bound irrespective of types of distributions of random effect $u(\cdot)$ and measurement errors ϵ .

Keywords: Small-dispersion asymptotics; Regular-consistency; Semi-parametric non-linear mixed effect model.

1. INTRODUCTION

Let us consider the following *semi-parametric* non-linear mixed effect regression model :

$$\begin{aligned} y_i &= f(x_i;\beta) + \sigma u(x_i) + \sigma\epsilon_i, \quad i=1,\cdots,n \\ y^* &= f(x;\beta) + \sigma u(x) \end{aligned} \tag{1.1}$$

where $y' = (y_1,\cdots,y_n)$ is a vector of n observations , y^* is a new unobserved random variable of interest , $f(x;\beta)$ is the non-linear regression function defined

† This research was supported by a grant from Korea Research Foundation 1993-1995

[1]Department of Statistics, Ewha Womans University, Seodaemun, Seoul, 120-750, Korea

on D which represents fixed effect and contains unknown parameter vector $\beta' = (\beta_1, \cdots, \beta_p)$, $u(x), x \in D$ is an arbitrary random function defined on D with zero mean and known covariance function $r(\cdot, \cdot)$ representing random effect , $\epsilon' = (\epsilon_1, \cdots, \epsilon_n)$ is a vector of uncorrelated measurement errors with zero mean and unit variance which is assumed to be uncorrelated with the random effect $u(x)$, $x \in D$ and σ is an unknown dispersion (scale) parameter.

Above model (1.1) , which incorporates both fixed effect $f(x, \beta)$ and random effect $u(x)$ and is called the mixed effect model in the literature, has many applications in such diverse fields as econometrics , mining , geology, animal breeding , meteorology , image restoration , agriculture , forestry , ecology and remote sensing. See Robinson (1991) for a general review of various applications of linear mixed effect model.

One of the most important problems in this model is to find the method of efficiently predicting a new unobserved random variable of interest y^* on the basis of past available data $(y_i, x_i), i = 1, \cdots, n$. For the important special case of linear mixed effect model, this problem has been settled satisfactorily by the classical theory of the best linear unbiased predictor (BLUP). See Robinson (1991) for the history and alternative derivations of BLUP as an optimal linear predictor. In the *non-linear* mixed effect model, the problem of efficient prediction has received relatively little attention in the literature. As a recent work in this direction , we can mention Gu (1990) who considered the problem of spline-smoothing in the non-Gaussian regression model as a natural extension of the semiparametric estimation method in the so-called partial-spline model of Engle at al. (1986) and Wahba (1984). We can also mention the works of Wedderburn (1974) and McCullagh (1983) on the robustness and the *large-sample* efficiency of the Quasi-Likelihood estimator in the generalized linear model (GLM) which depend only on the second-order structure and are independent of the types of distributions of observations.

On the other hand , in the *small-sample* non-linear problem , we can not resort to the usual large sample theory such as asymptotic normality in comparing performances of different prediction methods and should make a completely different approach to the asymptotics . In a pioneering work in this direction Villegas (1969) considered the problem of efficient estimation in the non-linear functional relation model with replicated observations for fixed n. Recently Jorgensen (1987) also investigated the possibility of small-sample low-noise asymptotics in the so-called exponential dispersion model.

In this paper we will focus on the problem of finding efficient predictor of the new random variable of interest y^* in the general non-linear mixed effect model (1.1) for a finite fixed sample size n and will develop new optimality theory from the perspective of small-dispersion asymptotics.

Motivated by the works of Villegas (1969) and Jorgensen (1987) on the small-dispersion asymptotics for finite n , we will apply the method of small-dispersion asymptotics to the prediction problem . This will provide us with the useful first-order approximations of various performance measures of competing predictors as dispersion parameter σ gets small. Specifically , after introducing relevant concept of consistency of the non-linear predictor , we will establish an absolute lower bound for the asymptotic mean squared errors of prediction (AMSEP) of the regular consistent predictors. Then we construct an optimal predictor which is based on the general non-linear least square (GNLS) estimator of the fixed effect $f(x;\beta)$ and the BLUP of its residuals and show that its AMSEP attains the lower bound irrespective of the type of the distributions of random effect $u(x)$ and the measurement errors ϵ_i .

This paper is organized as follows . In section 2 we introduce basic concept of regular consistent predictor as a natural extension of the linear unbiased predictor to our non-linear framework. Then we derive a fundamental semiparametric lower bound for the AMSEP of the regular consistent predictors which depends only on the second-order moments of the random effects and measurement errors. Next in section 3 , we construct an optimal predictor which is based on the non-linear least squares estimator of the fixed effect and the BLUP of its residuals and prove asymptotic optimality of the predictor . Finally in section 4 we discuss several possible extensions of the main optimality result to other interesting class of problems.

2. LOWER BOUND

First we introduce the basic concept of regular consistent predictor as a natural extension of the concept of linear-unbiased predictor of the linear mixed effect model.

Definition 2.1. *A predictor $h(y)$ of the new random variable y^* is called regular-consistent if $h(\cdot)$ is continuously differentiable and satisfies the following consistency condition :*

$$h[f(\beta)] = f(x;\beta) \quad for \;\; all \;\; \beta \in \Theta \tag{2.1}$$

where $y' = (y_1, \cdots, y_n)$, $f(\beta)' = (f(x_1;\beta), \cdots, f(x_n;\beta))$ and Θ is an open set *in* R^p.

Remark 2.1. We note that for the linear predictor , regular consistency reduces to the familiar unbiasedness of the predictor in the mixed effect linear model.

Remark 2.2. We note that the consistency condition (2.1) is equivalent to the following asymptotic unbiasedness condition :

$$E[h(y) - y^*] = o(1) \quad as \ \ \sigma \to 0, \ for \ any \ \beta \in \Theta.$$

Remark 2.3. Another equivalent definition of the consistency (2.1) is the following condition :

$$\lim_{\sigma \to 0} h(y) \ = \ f(x;\beta) \quad for \ any \ \beta \in \Theta.$$

Remark 2.4. Our model (1.1) may be considered as a semiparametric regression model because it incorporates both parametric component $f(x;\beta)$ which typically represents large-scale *global* variation of the unknown regression surface $m(x) = E[Y|x]$, $x \in D$ and the non-parametric component $\{u(x) \ , \ x \in D\}$ which describes residual small-scale *local* variation of the regression function $m(x)$ and is formally modeled by the stochastic process of known covariance function $r(\cdot,\cdot)$. In this partially-parametric framework , the problem of finding efficient predictor $h(y)$ of the new random variable $y^* = f(x;\beta) + \sigma u(x)$ at an arbitrary fixed $x \in D$ reduces to the problem of finding efficient non-parametric function estimator $\hat{m}(x)$ of the unknown regression function $m(x)$.

We also introduce the following definition of the AMSEP of the regular consistent predictor as a natural performance measure .

Definition 2.2. *AMSEP of the regular consistent predictor* $h(y)$ *of the new random variable* y^* *is formally defined as follows : If we have*

$$\sigma^{-1}[h(y) - y^*] \xrightarrow{d} Z \ as \ \ \sigma \to 0 \ ,$$

then

$$AMSEP[h(y)] = \sigma^2 E[Z^2]$$

where $\xrightarrow{d}$ *means convergence in distribution.*

We also make the following regularity condition for the model .

A$_1$: The mapping $f(\beta) = (f(x_1;\beta), \cdots, f(x_n;\beta)'$ from $\Theta \subset R^p$ to R^n is homeomorphic (one-to-one and bicontinuous) and continuously differentiable.

Now we establish the fundamental lower bound for the AMSEP of the regular consistent predictors of the new random variable y^* .

Theorem 2.1. *(Lower Bound for AMSEP) Let the condition A_1 be satisfied and let $F(\beta) = Df(\beta) = [\partial f(x_i;\beta)/\partial\beta_j]$ be a $n \times p$ Jacobian matrix of $f(\beta)$ with a full rank $p \leq n$. Let $h(\cdot)$ be a regular consistent predictor of y^* with $1 \times n$ gradient vector $\nabla h(y) = [\partial h/\partial y_i]$. Then we have :*

$$AMSEP[h(\cdot)] = (r(x,x) + \omega\Omega^{-1}\omega') + [\nabla h(f(\beta)) - \omega]\Omega^{-1}[\nabla h(f(\beta)) - \omega\Omega^{-1}]' \tag{2.2}$$

and

$$AMSEP[h] \geq (r(x,x) + \omega\Omega^{-1}\omega') + [\nabla f - F'\Omega^{-1}F][F'\Omega^{-1}F]^{-1}[\nabla f - F'\Omega^{-1}F]' \tag{2.3}$$

where $\nabla f = [\partial f(x;\beta)/\partial\beta_j]$ is the $1 \times p$ gradient vector of $f(x;\beta)$ with respect to β , ω is the $1 \times n$ vector of covariances $[cov(u(x), u(x_i))] = [r(x, x_i)]$ and Ω is a positive definite covariance matrix $Cov(u + \epsilon)$ of the $n \times 1$ random vector $u + \epsilon = (u(x_i) + \epsilon_i)$.

Proof: First we note that regular consistency of $h(y)$ as a predictor of y^* implies that ;

$$\begin{aligned} h(y) \; - \; y^* &= h(y) - h(f(\beta)) - \sigma u(x) \\ &= \nabla h(f(\beta))[y - f(\theta)] - \sigma u(x) + o(\sigma) \quad as\ \sigma \to 0 \,. \end{aligned} \tag{2.4}$$

Multiplying both sides of (2.4) by σ^{-1}, we have :

$$\sigma^{-1}[h(y) - y^*] = \nabla h(f(\beta))\sigma^{-1}[y - f(\theta)] - u(x) \; + o(1) \; as\ \sigma \to 0$$

This completes the proof of (2.2) because

$$AMSEP[h(\cdot)] = E[(\nabla h(f(\beta))(u + \epsilon) - u(x))^2]$$

which is the same as the expression in (2.2) . As for the proof of the lower bound of (2.3), we begin with the identity (2.1) :

$$h[f(\beta)] = f(x;\beta) \quad for\ all\ \ \beta\ \in\ \Theta.$$

Differentiating above identity with respect to β, we have the identity :

$$\nabla h(f(\beta))Df(\beta) \;=\; \nabla f(x;\beta). \tag{2.5}$$

Subtracting $\omega\Omega^{-1}F$ from (2.5) and postmultiplying it by a $p \times 1$ vector a , we have ;

$$(\nabla h - \omega\Omega^{-1})Fa = (\nabla f - \omega\Omega^{-1}F)a.$$

Then by the Cauchy-Schwartz inequality , we get the inequality ;

$$(\nabla h - \omega\Omega^{-1})\Omega(\nabla h - \omega\Omega^{-1})' \geq [(\nabla f - \omega\Omega^{-1}F)a]^2/(a'F'\Omega^{-1}Fa) \tag{2.6}$$

Now taking supremum of the lower bound of (2.6) with respect to $a \in R^p$ and noting the fact that $\sup_{a\in R^p}(a'b)^2/(a'Aa) = b'A^{-1}b$, we obtain the required inequality (2.3) immediately ;

$$AMSEP[h(y)] \geq (r(x,x)+\omega\Omega^{-1}\omega')+[\nabla f-\omega\Omega^{-1}F](F'\Omega^{-1}F)^{-1}[\nabla f-\omega\Omega^{-1}F]'.$$

□

Remark 2.5. If we let $h^* = \omega\Omega^{-1} + (\nabla f - F'\Omega^{-1}\omega')(F'\Omega^{-1}F)^{-1}F'\Omega^{-1}$, then we have the identity :

$$[\nabla h - \omega\Omega^{-1}]\Omega[\nabla h - \omega\Omega^{-1}]' \;=\; h^*\Omega h^{*'} \;+\; (\nabla h - h^*)\Omega(\nabla h - h^*)' \tag{2.7}$$

which follows immediately from (2.5) . Above identity provides us with not only alternative proof of the lower bound (2.3) but also important necessary and sufficient condition for the attainability of the lower bound :

$$\nabla h(f(\beta)) = h^*. \tag{2.8}$$

Now we introduce the definition of the efficiency of the regular consistent predictor as follows ;

Definition 2.3. *A regular consistent predictor $h(y)$ of y^* is said to be efficient if its AMSEP attains the lower bound (2.3) for all $\beta \in \Theta$.*

In the next section we will construct an efficient regular consistent predictor $h(y)$ of y^* .

3. EFFICIENT PREDICTOR

In order to construct an efficient regular consistent predictor of $y^* = f(x;\beta) + \sigma u(x)$ we have to find not only the efficient estimator of the fixed component $f(x;\beta)$ but also some efficient predictor of the random component $u(x)$. As a first step we introduce the general non-linear least squares (GNLS) estimator of β .

Definition 3.1. *GNLS estimator $\hat{\beta}$ of β is defined formally as the unique solution of the following system of normal equations :*

$$F(y,\theta) = Df(\beta)'\Omega^{-1}[y - f(\beta)] = 0 \tag{3.1}$$

if there exists an unique solution and is defined arbitrarily otherwise.

We also assume the following regularity condition :

$\mathbf{A_2}$: $F(\beta) = Df(\beta)$ is a continuously differentiable function of β in Θ.

Motivated by the theory of BLUP in the linear mixed effect model , we now introduce the following predictor $h(y)$ of y^* which is based on the GNLS estimator $\hat{\beta}$ of β and its residuals :

$$h^*(y) \;=\; f(x;\hat{\beta}) \;+\; \hat{u}(x) \tag{3.2}$$

where $\hat{u}(x) \;=\; \omega(x)\Omega^{-1}[y - f(\hat{\beta})]$ and $\omega(x) = (r(x,x_1),\cdots,r(x,x_n))$. Now we claim that the predictor $h^*(y)$ is an efficient regular consistent predictor of y^*.

Theorem 3.1. *Let the conditions A_1 and A_2 be satisfied. Then the predictor $h^*(y)$ is an efficient regular consistent predictor of y^* .*

Proof: First by the same argument as in theorem 2.1 of Ferguson (1958), we can establish the existence and uniqueness of the GNLS estimator $\hat{\beta}(y)$ of β which is defined on some neighborhood N of the set $S = \{f(\beta) \in R^n; \beta \in \Theta\}$ and is continuously differentiable such that

$$\hat{\beta}(f(\beta)) = \beta \;\; for\ all\ \beta \in \Theta$$

and

$$F(y,\hat{\beta}) = Df'(\hat{\beta})\Omega^{-1}(y - f(\hat{\beta})) = 0. \tag{3.3}$$

This establishes both regularity and consistency of the predictor $h^*(y)$ immediately. Now expanding $F(y,\hat{\beta})$ in (3.3) around β and rearranging terms, we have the result

$$\hat{\beta} - \beta \ = \ (F'\Omega^{-1}F)^{-1}F\Omega^{-1}(y - f(\beta)) \ + \ o(\sigma) \ \ as \ \sigma \to 0. \tag{3.4}$$

On the other hand , we note that

$$\begin{aligned} h^*(y) - y^* &= f(x,\hat{\beta}) + \omega(x)\Omega^{-1}(y - f(\hat{\beta})) - f(x,\beta) - \sigma u(x) & (3.5) \\ &= \nabla f(x,\beta)(\hat{\beta} - \beta) + \omega\Omega^{-1}(y - f(\beta) - Df(\hat{\beta} - \beta)) - \sigma u(x) + o(\sigma). \end{aligned}$$

Substituting (3.4) in (3.5) , we obtain immediately

$$h^*(y) - y^* = [\omega\Omega^{-1} + (\nabla f - \omega\Omega^{-1}F)(F'\Omega^{-1}F)^{-1}F'\Omega^{-1}](y - f(\beta)) - \sigma u(x) + o(\sigma). \tag{3.6}$$

Multiplying (3.6) by σ^{-1} we have

$$\sigma^{-1}[h^*(y) - y^*] = \nabla h^*(f(\beta))(u + \epsilon) - u(x) + o(1).$$

Since we have $\nabla h^*(f(\beta)) = h^*$, we conclude the efficiency of the predictor $h^*(y)$ directly from the sufficient condition (2.8) of Remark 2.5. □

Remark 3.1. As is noted in Remark 2.4 , the optimal predictor

$$h^*(x;y) = f(x;\hat{\beta}) + \omega(x)\Omega^{-1}(y - f(\hat{\beta})) \ , \ x \in D$$

, when written as an explicit function of $x \in D$ and the data $y = (y_1, \cdots, y_n)'$, can be considered as a kind of *non-linear smoother* $\hat{m}_n(x)$ of the unknown regression function $m(x) = f(x;\beta) + \sigma u(x)$ with a nice reproducing property of recovering original function $m(x)$ exactly for any finite n whenever there is no model error $u(x)$ and no measurement errors ϵ .

4. DISCUSSIONS

In this section we discuss several possible extensions of the finite-sample small-dispersion optimality results of previous sections to other class of interesting mixed non-linear models.

Remark 4.1. When we have correlated measurement errors $\epsilon' = (\epsilon_1, \cdots, \epsilon_n)$ with known covariance matrix $\Sigma = [\sigma_{ij}]$, we can easily obtain optimal predictor of the new random variable y^* if we let $\Omega = Cov(u + \epsilon) = R + \Sigma, R = [r(x_i, x_j)]$.

Remark 4.2. Furthermore if $Cov(\epsilon) = \Sigma(\beta)$ and the covariance function $r(\cdot,\cdot;\beta)$ of the random effect $u(x)$ depend explicitly on β as is the case in GLM , we can also construct an optimal regular consistent predictor $h^*(y)$ of the new random variable y^* which is based on the quasi-likelihood estimator (QLE) $\hat{\beta}$ of β given by the solution of the normal equation ;

$$Df'(\beta)\Omega^{-1}(\beta)(y - f(\beta)) = 0.$$

If there exists a consistent estimator $\bar{\Omega}$ of Ω ,we can also construct an alternative optimal regular consistent predictor $h^{**}(y)$ which is based on the modified GNLS estimator $\bar{\beta}$ of β defined by ;

$$Df'(\beta))\bar{\Omega}^{-1}(y - f(\beta)) = 0.$$

Remark 4.3. Finally let us consider the following type of very general mixed effect model which include our model (1.1) as special case ;

$$\begin{aligned} y_i &= f(x_i; \beta, u(x_i)) + \epsilon_i \quad i = 1, \cdots, n \\ y* &= f(x; \beta, u(x)) \end{aligned}$$

where $f(x; \beta, u(x))$ is a smooth function of (β, u) for each $x \in D$. Under suitable finite-sample small-dispersion asymptotic framework , we can expect to find an optimal non-linear predictor $h^*(y)$ of y^* . Optimality theory for this important class of problems will be pursued in a subsequent paper in detail.

References

Engle, R.F. , Granger, C.W.J. , Rice J.A. and Weiss A. (1986). "Semiparametric estimates of the relation between weather and electricity sale," *Journal of American Statistical Association* , 81, 310-320.

Ferguson T.S. (1958). "A method of generating BAN estimates with application to the estimation of bacterial densities," *The Annals of Mathematical Statistics.* 29, 1046-1062.

Gu C. (1990). "Adaptive Spline smoothing in non-Gaussian regression models," *Journal of the American Statistical Association.* , 85, 801-807.

Jorgensen B. (1987). "Exponential dispersion models," *Journal of Royal Statistical Society.* B 49, 127-162.

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

Robinson G.K. (1991). "That BLUP is a good thing ; the estimation of random effects," *Statistical Science.* , Vol. 6 , No 1, 15-51.

Villegas C. (1969). "On the least squares estimation of non-linear relations," *Annals of Statistics.* 40, 462-466.

Wahba G. (1984). "Partial spline models for the semiparametric estimation of functions of several variables," *Seminar Proceedings , Statistical Analysis of Time Series* , Japan-U.S. Joint Seminar.

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