TECHNICAL REPORT NO. 958
MARCH 1996

LEAST SQUARES ESTIMATION OF REGRESSION PARAMETERS IN MIXED EFFECT MODELS WITH UNMEASURED COVARIATES

by

J. Shao, M. Palta, and R.P. Qu

University of Wisconsin-Madison

1This work was supported by grant CA-53786 from the National Cancer Institute.
Least Squares Estimation of Regression Parameters in Mixed Effect Models with Unmeasured Covariates

J. Shao, M. Palta, and R.P. Qu\textsuperscript{1}
University of Wisconsin-Madison

SUMMARY

We consider mixed effects models for longitudinal, repeated measures or clustered data. Unmeasured or omitted covariates in such models may be correlated with the included covariates, and create model violations when not taken into account. Previous research and experience with longitudinal data sets suggest a general form of model which should be considered when omitted covariates are likely, such as in observational studies. We derive ordinary and weighted least squares estimators, which are useful in models of the assumed structure, and make no assumptions on the distribution of random effects or error terms. Asymptotic properties of these estimators are also derived. The results shed light on the structure of least squares estimators in mixed effects models, and allow simple non-iterative estimation. We present an example of the relationship between fluid intake and output in very low birth weight infants, where the model is found to have the assumed structure.

KEY WORDS: Omitted covariates, repeated measures, random effects, clusters, longitudinal

\textsuperscript{1}This work was supported by grant CA-53786 from the National Cancer Institute.
1. INTRODUCTION

Omitted covariates in regression analysis of longitudinal data can be related to either follow-up time or to the characteristics of individuals who enter the study. For example, a recent report (Neuschaffer, Bush and Hale, 1992) showed that in a study of cholesterol levels in the elderly, there was both evidence that persons born in earlier calendar years had lower cholesterol and that cholesterol levels had risen with the calendar time at which the measurement was obtained. These trends may mark the presence of unknown individual level factors, such as childhood diet that may cause generational differences. Time related trends might be speculated to have arisen from population wide unmeasured covariates, such as availability of pre-processed, fat-rich foods. From an epidemiologic perspective, such factors confound regression estimators of aging effect with commonly used models, and are well known under the terms cohort and period effects.

In other situations, such as in the example of the relationship between fluid intake and output in very premature newborns given later in this paper, included covariates may simultaneously be markers for several unmeasured covariates. The weight of a neonate is a marker for body surface area, which in turn is positively related to a component of fluid loss not measured. However, although this component can be expected to be a major determinant of why neonates differ in measured fluid output, short term day to day variation in weight may be a marker for temporary fluid retention with subsequent diuresis. Thus, again omitted covariates are present, which have different relationships with influential unmeasured covariates within and between individuals.

From a statistical point of view, omitted covariates create violations of assumptions with commonly used methods such as mixed effects regression models. It is usually implicitly assumed in such mixed effects modeling that the distribution of the random effects does not depend on the fixed covariates. Palta and Yao (1991) and Palta and Qu (1995) demonstrated that omitted covariates can lead to violation of this assumption. Essentially this occurs when omitted and included covariates are correlated, but differentially so longitudinally and cross-sectionally. Palta, Yao and Velu (1994) give several examples of such situations, derive models which may arise when conditioning only on observed covariates, and develop procedures for testing for omitted covariates. In related work, Wishart (1938) and Mundlak (1978) both addressed the implications of the independence assumption in the mixed effect model with a random intercept, as did Dwyer et al.(1992), pages 27-33.
Palta and Yao (1991) derived the correct form of the regression function under certain assumptions on the covariate structure containing omitted covariates. In the present investigation we generalize this form and develop procedures for fitting such models that make minimal distributional assumptions. Our regression model is particularly relevant when omitted covariates may be present, but may also arise from other circumstances. In Section 2 we give specific examples of situations giving rise to such models. We then develop fitting procedures based on least squares and weighted least squares methods which, in contrast to likelihood based procedures, do not require distributional assumptions on the responses (Sections 3 and 4). The resulting estimator generalizes a results of Hsiao (1986) and provides insight into the construction of estimators from mixed effect models. We also derive asymptotic results which are useful for large sample inference. The weighted least squares method is shown to be asymptotically efficient when the covariance matrix of the response is correctly specified.

In Section 5, we apply the method to the data set of fluid output and intake in very low birth weight neonates.

2. MODELS

Consider independent response vectors \( y_1, \ldots, y_n \), where for each \( i \),

\[
y_i = \alpha 1 + X_i \beta + Z_i \gamma + e_i
\]

(2.1)
is a \( k_i \times 1 \) response vector, \( 1 \) is a vector of one’s (of appropriate order), \( X_i \) and \( Z_i \) are \( k_i \times p \) and \( k_i \times q \) matrices of values of random covariates, respectively, \( \alpha, \beta \) and \( \gamma \) are parameters or parameter vectors of appropriate orders, and \( e_i \) is a \( k_i \times 1 \) random error vector satisfying

\[
E(e_i|X_i, Z_i) = 0
\]

and

\[
V(e_i|X_i, Z_i) = \Sigma_i
\]

(which may depend on \( X_i \) and/or \( Z_i \)). Hence,

\[
E(y_i|X_i, Z_i) = \alpha 1 + X_i \beta + Z_i \gamma
\]

and

\[
V(y_i|X_i, Z_i) = \Sigma_i.
\]
Suppose that the $Z_i$ are values of $q$ unmeasured covariates. Taking the average over $Z_i$, we obtain

\[
E(y_i|X_i) = E[E(y_i|X_i, Z_i)|X_i] = \alpha 1 + X_i \beta + E(Z_i \gamma |X_i)
\]  
(2.2)

and

\[
V(y_i|X_i) = E[V(y_i|X_i, Z_i)|X_i] + V[E(y_i|X_i, Z_i)|X_i] = E(\Sigma_i|X_i) + V(Z_i \gamma |x_i).
\]  
(2.3)

Model fitting and testing can be done when $E(Z_i \gamma |X_i)$ and $V(Z_i \gamma |X_i)$ are of some special forms. Thus, we now focus on the model between the covariates $X_i$ and $Z_i$.

2.1. A Linear Model for $X_i$ and $Z_i$

The simplest model for covariates $X_i$ and $Z_i$ is the following linear model:

\[
E(Z_i \gamma |X_i) = A_i(X_i) \xi,
\]
(2.4)

where $A_i(X_i)$ is a $k_i \times s$ (for some $s < k_i$) matrix whose elements are functions of $X_i$, and $\xi$ is an $s \times 1$ vector of unknown parameters. Some examples are given as follows.

**Example 2.1.** Consider the case of $p = q = 1$. If $(X_i, Z_i)$ is multivariate normal as described in Palta and Yao (1991, p1356), then (2.4) holds with

\[
A_i(X_i) = (X_i, \bar{x}_i 1),
\]

where $\bar{x}_i$ is the average of the components of $X_i$, and

\[
\xi = \gamma \left( \begin{array}{c} R_2 \sqrt{C_2} \\
{r(R_1 \sqrt{C_1} - R_2 \sqrt{C_2})} \end{array} \right).
\]

The parameters $r$, $C_1$, $C_2$, $R_1$ and $R_2$ are given in Palta and Yao (1991). In this case, $s = 2$ and

\[
V(Z_i \gamma |X_i) = c_1 I + c_2 11',
\]

for some unknown constants $c_1$ and $c_2$ (which do not depend on $X_i$), where $I$ denotes the identity matrix (of appropriate order).
Example 2.2. Suppose that
\[ Z_i \gamma = A_i(X_i)\xi + u_i, \]
where \( E(u_i|X_i) = 0 \). That is, there is a linear model that describes the relationship between \( X_i \) and \( Z_i \gamma \). Note that Example 2.1 is a special case of Example 2.2. No assumption on the joint distribution of \( X_i \) and \( Z_i \) is made.

Combining (2.2) and (2.4), we obtain the following model between the response \( y_i \) and the covariate \( X_i \):
\[ E(y_i|X_i) = \alpha 1 + X_i \beta + A_i(X_i)\xi \tag{2.5} \]
and
\[ V(y_i|X_i) = E(\Sigma_i|X_i) + V(Z_i \gamma|X_i). \tag{2.6} \]

Note that the matrix \( A_i(X_i) \) may contain \( 1 \) and/or some columns of \( X_i \) (see Example 2.1). Let \( \bar{A}_i(X_i) \) be the submatrix of \( A_i(X_i) \) containing columns that are not in \( (1, X_i) \). Then (2.5) can be written as
\[ E(y_i|X_i) = \tilde{\alpha} 1 + X_i \tilde{\beta} + \bar{A}_i(X_i)\tilde{\xi}, \tag{2.7} \]
where \( \tilde{\alpha} \) is an unknown parameter and \( \tilde{\beta} \) and \( \tilde{\xi} \) are unknown parameter vectors. Note that the original parameter \( \beta \) (or \( \alpha \)) is estimable if and only if \( A_i(X_i) \) does not contain \( X_i \) (or 1).

2.2. Nonlinear Models for \( X_i \) and \( Z_i \)

In some cases, the relationship between \( X_i \) and \( Z_i \) is nonlinear. Assume that
\[ E(Z_i \gamma|X_i) = \phi_i( A_i(X_i)\xi ), \tag{2.8} \]
where \( A_i(X_i) \) and \( \xi \) are the same as before, and \( \phi_i \) is a vector-valued function. If \( \phi_i(t) = t \), then we have a linear model between \( X_i \) and \( Z_i \).

Example 2.3. Logit Model. Let \( Z_i \gamma = (z_{i1}, ..., z_{ik_i})' \) and \( X_i = (x_{i1}, ..., x_{ik_i})' \). Suppose that each \( z_{ij} \) is binary and
\[ g(Ez_{ij}) = a + bx_{ij} + cx_{ij}, \]
where \( a \), \( b \) and \( c \) are unknown parameters and \( g(t) = \log \frac{t}{1-t} \) or \( g(t) = \Phi^{-1}(t) \). Then the \( l \)th component of \( \phi_i \) is
\[ g^{-1}(\text{the } l \text{th row of } A_i(X_i)\xi) \]
with

\[ A_i(X_i) = (1, X_i, \bar{x}_i 1), \quad \xi = (a, b, c)' \]

**Example 2.4.** A special case of Example 2.3 arises when \( X_i \) is assumed normal and \( z_{ij} = z_i \) binary, as would occur when individuals fall into two groups with distinct outcome expectations. It can be easily shown that when this grouping is unknown or ignored, \( g \) is logistic with \( b = 0 \).

### 3. OLS MODEL FITTING AND TESTING

We start with the ordinary least squares (OLS) method which does not require any assumption on the covariance matrix \( \Sigma(y_i | X_i) \) in (2.6).

**3.1. Linear Model (2.7)**

Assume that the matrix

\[ W = (W'_1, W'_2, ..., W'_n)' \]

is of full rank, where

\[ W_i = \left(1, X_i, \bar{A}_i(X_i)\right). \]

Then the OLSE of \( \theta = (\bar{\alpha}, \bar{\beta}', \zeta')' \) is

\[ \hat{\theta}_{OLS} = (W'W)^{-1}W'y, \]

where \( y = (y'_1, ..., y'_n)' \).

The OLSE \( \hat{\theta}_{OLS} \) is a weighted average of the individual OLSE’s of the following models:

\[ E(y_i | X_i) = W_i \theta, \quad (3.1) \]

\( i = 1, ..., n \). Note that the \( W_i \) may not be of full rank, although \( W \) is of full rank. Assume that the rank of \( W_i \) is \( r \). Let \( R_i \) be the matrix containing \( r \) independent columns of \( W_i \). Then

\[ W_i = R_i Q_i \]

(3.2)

for some matrix \( Q_i \). For example, if

\[ W_i = (1, X_i, \bar{x}_{i1} 1, ..., \bar{x}_{ip} 1), \quad (3.3) \]
where $\bar{x}_{ij}$ is the average of the $j$th column of $X_i$ (e.g., Example 2.1), then

$$R_i = (1, X_i)_{k_i \times (p+1)} \quad \text{and} \quad Q_i = \begin{pmatrix} I_{(p+1) \times (p+1)} & \bar{x}_{i1} & \cdots & \bar{x}_{ip} \\ 0_{p \times 1} & \cdots & 0_{p \times 1} \end{pmatrix}.$$ 

Let $\hat{\delta}_i$ be the OLSE under model (3.1), i.e.,

$$\hat{\delta}_i = (R_i' R_i)^{-1} R_i' y_i.$$

Then $\hat{\delta}_i$ is unbiased for $\delta_i = Q_i \theta$ and

$$\hat{\theta}_{OLS} = (W'W)^{-1} \sum_{i=1}^{n} W_i' y_i = (W'W)^{-1} \sum_{i=1}^{n} Q_i' R_i y_i = (W'W)^{-1} \sum_{i=1}^{n} Q_i' (R_i' R_i) \hat{\delta}_i.$$  \hfill (3.4)

It is clear that $\hat{\theta}_{OLS}$ is unbiased for $\theta$. When $n \to \infty$, $\hat{\theta}_{OLS}$ is consistent and asymptotically normal under some mild conditions. The following result can be established by directly applying the central limit theorem and the law of large numbers.

**Theorem 3.1.** (i) Assume that $\sup_i k_i < \infty$,

$$\inf_i E(W_i' W_i) > 0, \quad \text{and} \quad \inf_i E[W_i' V(y_i | X_i) W_i] > 0.$$

Suppose also that there is a $\delta > 0$ such that

$$\sup_i E\|W_i' W_i\|^{1+\delta} < \infty \quad \text{and} \quad \sup_i E\|W_i' V(y_i | X_i) W_i\|^{1+\delta} < \infty,$$

where $\|A\|^2 = \text{tr}(A' A)$. Then, as $n \to \infty$,

$$V_{OLS}^{-1/2} (\hat{\theta}_{OLS} - \theta) \to N(0, I) \quad \text{in law},$$

where

$$V_{OLS} = (W'W)^{-1} \sum_{i=1}^{n} W_i' V(y_i | X_i) W_i (W'W)^{-1}.$$ 

(ii) A consistent estimator of $V_{OLS}$ is

$$\hat{V}_{OLS} = (W'W)^{-1} \sum_{i=1}^{n} W_i r_i r_i' W_i (W'W)^{-1},$$
where
\[ r_i = y_i - W_i \hat{\theta}_{OLS}. \] (3.5)

An approximate level \( a \) test for hypothesis
\[ H_0 : C\theta = 0 \quad \text{versus} \quad H_1 : C\theta \neq 0 \] (3.6)
rejects \( H_0 \) if
\[ \hat{\theta}_{OLS}'C'[C\hat{V}_{OLS}C]^{-1}C\theta_{OLS} > \chi_a^2(d), \] (3.7)
where \( C \) is a matrix of constants, \( d \) is the row-dimension of \( C \), and \( \chi_a^2(d) \) is the \( 1 - a \) percentile of the chi-square distribution with \( d \) degrees of freedom. For example, if \( W_i \) is given by (3.3) and \( \zeta = 0 \) (which means that the \( \bar{x}_{ij} \) have no significant effects) is to be tested, then
\[ C = \left(0_{p \times (p+1)}, I_{p \times p}\right). \]

3.2. Nonlinear Model (2.8)

The OLS method can be directly applied to the case where the relationship between \( X_i \) and \( Z_i \) is nonlinear, i.e., (2.8) holds with a nonlinear \( \phi_i \). Suppose that we can rewrite the mean function as
\[ E(y_i|X_i) = f_i(X_i, \theta) \]
with a function \( f_i \) differentiable in \( \theta \) and an estimable parameter vector \( \theta \) in the sense that
\[ \inf_i E[f_i(X_i, t) - f_i(X_i, \theta)]' [f_i(X_i, t) - f_i(X_i, \theta)] > 0 \] (3.8)
for any \( t \neq \theta \). If all of the original parameters are estimable, then \( \theta = (\alpha, \beta', \xi')' \) and
\[ f_i(X, \theta) = \alpha 1 + X\beta + \phi_i(a_i(X)\xi), \]
\[ \frac{\partial f_i}{\partial t}(X, t) = \left(1, X, \phi_i(a_i(X)\xi), \frac{\partial \phi_i}{\partial a}_{a=A_i(X)\xi} A_i(X)\right). \]

The OLSE of \( \theta \), denoted by \( \hat{\theta}_{OLS} \), is a solution to the minimization problem
\[ \sum_{i=1}^{n} \| y_i - f_i(X_i, t) \|^2 = \min. \]
over $t \in \Theta$ (the parameter space). Then $\hat{\theta}_{OLS}$ is a solution to

$$
\sum_{i=1}^{n} [y_i - f_i(X_i, t)] \frac{\partial f_i(X_i, t)}{\partial t} = 0.
$$

Asymptotic properties of the OLSE are given in the following result.

**Theorem 3.2.** (i) Assume condition (3.8) and either $\Theta$ is compact or $\inf_i \|f_i(X, t)\| \to \infty$ as $\|t\| \to \infty$ uniformly for $\|X\| \leq c_0$ with a constant $c_0$. Then

$$
\hat{\theta}_{OLS} \xrightarrow{p} \theta.
$$

(ii) Assume that (3.9) holds, $f_i(X, t)$ is twice differentiable in $t$, 

$$
\inf_i E \left\{ \frac{\partial f_i(X_i, \theta)}{\partial t} \left[ \frac{\partial f_i(X_i, \theta)}{\partial t} \right]' \right\} > 0,
$$

and

$$
\left\| \frac{\partial f_i(X_i, t)}{\partial t} \right\|^4 \leq b(X_i)
$$

and

$$
\left\| \frac{\partial^2 f_i(X_i, t)}{\partial t \partial t'} \right\|^2 \leq b(X_i)
$$

for all $i$ and $\|t - \theta\| \leq \epsilon_0$ with an $\epsilon_0 > 0$, where $b(X)$ is a function satisfying $\sup_i E[b(X_i)]^{1+\delta} < \infty$ for some $\delta > 0$. Then

$$
V_{OLS}^{-1/2}(\hat{\theta}_{OLS} - \theta) \xrightarrow{d} N(0, I)
$$

in law,

where

$$
V_{OLS} = \left[ W'(\theta) W(\theta) \right]^{-1} W'(\theta) G^{-1} W(\theta) [W'(\theta) W(\theta)]^{-1}
$$

and

$$
W(t) = (W_1(t), W_2(t), ..., W_n(t))', \quad W_i'(t) = \frac{\partial f_i(X_i, t)}{\partial t}.
$$

(iii) A consistent estimator of $V_{OLS}$ is

$$
\hat{V}_{OLS} = \left[ W'(\hat{\theta}_{OLS}) W(\hat{\theta}_{OLS}) \right]^{-1} W'(\hat{\theta}_{OLS}) \hat{\Sigma} W(\hat{\theta}_{OLS}) [W'(\hat{\theta}_{OLS}) W(\hat{\theta}_{OLS})]^{-1},
$$

where $\hat{\Sigma}$ is a block diagonal matrix whose $i$th diagonal block is

$$
[y_i - f_i(X_i, \hat{\theta}_{OLS})] [y_i - f_i(X_i, \hat{\theta}_{OLS})]',
$$

9
4. WLS MODEL FITTING AND TESTING

In general, the efficiency of the OLSE is lower than that of the best linear unbiased estimator (BLUE) of $\theta$. If the $V(y_i|X_i)$ are known, then the BLUE is of the form

$$
\hat{\theta}_{BLU} = (W'GW)^{-1}W'Gy,
$$

where

$$
G = \begin{pmatrix}
[V(y_1|X_1)]^{-1} & 0 & \cdots & 0 \\
0 & [V(y_2|X_2)]^{-1} & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & [V(y_n|X_n)]^{-1}
\end{pmatrix}
$$

is a block diagonal matrix.

Since the matrices $V(y_i|X_i)$ are unknown, replacing them by estimates $\hat{V}(y_i|X_i)$ leads to the weighted least squares estimator (WLSE)

$$
\hat{\theta}_{WLS} = (W'\hat{G}W)^{-1}W'\hat{G}y,
$$

where $\hat{G}$ is $G$ with $V(y_i|X_i)$ replaced by $\hat{V}(y_i|X_i)$.

Construction of the estimates $\hat{V}(y_i|X_i)$ depends on the structure of the covariance matrix $V(y_i|X_i)$, which can be modeled using several approaches. A common approach is to use the mixed effect regression model (Laird and Ware, 1982; Palta and Yao, 1991) and assume that

$$
V(y_i|X_i) = \sigma^2 I + U_iDU'_i, \quad (4.1)
$$

where $\sigma^2 > 0$ is an unknown parameter, $D$ is a nonnegative definite matrix of unknown parameters, and $U_i$ is a matrix containing some columns of $R_i$ defined in (3.2). This variance structure can be shown to hold under quite general conditions when the variance of model (2.1) is of similar form. Under (4.1), we may estimate $V(y_i|X_i)$ by estimating $\sigma^2$ and $D$. An unbiased estimator of $\sigma^2$ is

$$
\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - R_i\hat{\delta}_i)'(y_i - R_i\hat{\delta}_i)/\sum_{i=1}^{n}(k_i - s)}{\sum_{i=1}^{n}y_i'[I - R_i(R_i'R_i)^{-1}R_i']y_i/\sum_{i=1}^{n}(k_i - s)}, \quad (4.2)
$$

10
since

\[ E(\hat{\sigma}^2) = \sum_{i=1}^{n} \text{tr} \left\{ (I - R_i' R_i)^{-1} R_i' E(y_i y_i') \right\} / \sum_{i=1}^{n} (k_i - s) \]

\[ = \sum_{i=1}^{n} \text{tr} \left\{ (I - R_i' R_i)^{-1} R_i' [\sigma^2 I + R_i DR_i'] \right\} / \sum_{i=1}^{n} (k_i - s) \]

\[ = \sum_{i=1}^{n} \text{tr} \left\{ (I - R_i' R_i)^{-1} R_i' \sigma^2 \right\} / \sum_{i=1}^{n} (k_i - s) \]

\[ = \sigma^2. \]

This estimator is also consistent as \( n \to \infty \).

It is more difficult to obtain a good estimator of \( D \). We may use

\[ \hat{D} = \frac{1}{n} \sum_{i=1}^{n} (U_i' U_i)^{-1} U_i' r_i U_i (U_i' U_i)^{-1} - \frac{\hat{\sigma}^2}{n} \sum_{i=1}^{n} (U_i' U_i)^{-1}, \quad (4.3) \]

where \( r_i \) is the \( i \)th residual (from the OLS fitting) given by (3.5). This estimator is asymptotically unbiased and consistent as \( n \to \infty \). In fact

\[ \hat{D} = \bar{D} + o(1) \]

where

\[ \bar{D} = \frac{1}{n} \sum_{i=1}^{n} (U_i' U_i)^{-1} U_i' (y_i - W_i \theta) (y_i - W_i \theta)' U_i (U_i' U_i)^{-1} - \frac{\hat{\sigma}^2}{n} \sum_{i=1}^{n} (U_i' U_i)^{-1}. \]

Then, under (4.1),

\[ E(\bar{D}) = \frac{1}{n} \sum_{i=1}^{n} (U_i' U_i)^{-1} U_i' V (y_i | X_i) U_i (U_i' U_i)^{-1} - \frac{E(\hat{\sigma}^2)}{n} \sum_{i=1}^{n} (U_i' U_i)^{-1} \]

\[ = \frac{1}{n} \sum_{i=1}^{n} (U_i' U_i)^{-1} U_i' [\sigma^2 I + U_i D U_i'] U_i (U_i' U_i)^{-1} - \frac{\sigma^2}{n} \sum_{i=1}^{n} (U_i' U_i)^{-1} \]

\[ = D. \]

When \( U_i = R_i \), the estimator in (4.3) becomes

\[ \hat{D} = \frac{1}{n} \sum_{i=1}^{n} (\hat{\delta}_i - Q_i \hat{\theta}_{OLS})(\hat{\delta}_i - Q_i \hat{\theta}_{OLS})' - \frac{\hat{\sigma}^2}{n} \sum_{i=1}^{n} (R_i' R_i)^{-1}, \]

In this case, a different estimator of \( D \) given by Swamy (1970) is

\[ \bar{D} = \frac{1}{n-1} \sum_{i=1}^{n} \left( \hat{\delta}_i - \bar{\delta} \right) \left( \hat{\delta}_i - \bar{\delta} \right)' - \frac{\hat{\sigma}^2}{n} \sum_{i=1}^{n} (R_i' R_i)^{-1}, \]
where $\bar{\delta}$ is the average of the $\delta_i$. However, $\bar{D}$ is neither asymptotically unbiased nor consistent in general. It is asymptotically unbiased and consistent in some special cases, e.g., when $\xi = 0$ (which is often a null hypothesis we would like to test). This is because

$$
E(\bar{D}) = \frac{1}{n} \sum_{i=1}^{n} V(\delta_i) + \frac{1}{n-1} \sum_{i=1}^{n} (\delta_i - \bar{\delta})(\delta_i - \bar{\delta})' - \frac{\sigma^2}{n} \sum_{i=1}^{n} (R'_i R_i)^{-1},
$$

$$
= D + \frac{1}{n-1} \sum_{i=1}^{n} (\delta_i - \bar{\delta})(\delta_i - \bar{\delta})',
$$

where $\bar{\delta}$ is the average of the $\delta_i$, and under $\xi = 0$, $\delta_i \equiv \bar{\delta}$.

Let $\delta_i^W$ be the WLSE of $\delta_i$ under model (3.1), i.e.,

$$
\delta_i^W = (R'_i \hat{G}_i R_i)^{-1} R'_i \hat{G}_i y_i,
$$

where $\hat{G}_i$ is the ith diagonal block of the matrix $\hat{G}$. Then $\theta_{WLS}$ is a weighted average of the $\delta_i^W$:

$$
\hat{\theta}_{WLS} = (W' \hat{G} W)^{-1} \sum_{i=1}^{n} Q_i^p (R'_i \hat{G}_i R_i) \delta_i^W.
$$

If (4.1) holds, then $\hat{\delta}_i = \delta_i^W$ for all $i$ and

$$
\hat{\theta}_{WLS} = (W' \hat{G} W)^{-1} \sum_{i=1}^{n} Q_i^p (R'_i \hat{G}_i R_i) \delta_i.
$$

Comparing (3.4) with (4.4), we find that both the OLSE and WLSE are weighted averages of the $\delta_i$, but they use different weights. This generalizes a result of Hsiao (1986) to the case when covariates are included which vary only between individuals.

The WLSE $\hat{\theta}_{WLS}$ is asymptotically unbiased, consistent, and asymptotically normal, as $n \to \infty$. If (4.2) holds and $\sigma^2$ and $D$ are estimated by $\hat{\sigma}^2$ in (4.2) and $\bar{D}$ in (4.3), then $\hat{\theta}_{WLS}$ is asymptotically as efficient as the BLUE, i.e., there is no asymptotic efficiency loss due to estimating the unknown $V(y_i|X_i)$.

**Theorem 4.1.** (i) Assume the conditions in Theorem 3.1. Assume further that

$$
\max_{1 \leq i \leq n} \| \hat{G}_i G_i^{-1} - I \| \to_p 0,
$$

where $G_i$ and $\hat{G}_i$ are the ith diagonal blocks of $G$ and $\hat{G}$, respectively. Then

$$
V_{WLS}^{1/2} (\hat{\theta}_{WLS} - \theta) \to N(0, I) \text{ in law},
$$

where $V_{WLS}$ is the variance of $\hat{\theta}_{WLS}$ under model (3.1).
where
\[ V_{WLS} = (W'GW)^{-1}. \]

(ii) A consistent estimator of \( V_{WLS} \) is
\[ \hat{V}_{WLS} = (W'\hat{G}W)^{-1}. \]

In particular, if \( V(y_i|X_i) \) is estimated by \( \hat{\sigma}^2 I + U_i \hat{D} U_i' \), where \( \hat{\sigma}^2 \) and \( \hat{D} \) are given by (4.2) and (4.3), then (4.5) and (4.6) hold, provided that condition (4.1) and the following conditions hold:
\[ \sup_i E(Q_i Q_i') < \infty, \]
\[ \sup_i E((R_i R_i')^{-1})^{1+\delta} < \infty, \]
and
\[ \sup_i E(\epsilon_i \epsilon_i')^{1+\delta} < \infty \]
for some \( \delta > 0 \), where \( \epsilon_i = y_i - E(y_i|X_i) \).

The testing of hypotheses of the form (3.6) can be carried out by using (3.7) with \( \hat{\theta}_{OLS} \) and \( \hat{V}_{OLS} \) replaced by \( \hat{\theta}_{WLE} \) and \( \hat{V}_{WLS} \), respectively.

The WLSE is more efficient than the OLSE. But it is sensitive to the correctness of specifying the matrix \( U_i \) in (4.1). Our experience in applying the WLSE to actual data has shown that the WLSE may be very unstable when \( U_i \) contains close to the the maximum number of columns (i.e., when \( D \) is over parameterized). For this reason, the OLS is a valuable tool.

5. AN APPLICATION

The following example is taken from information up to four days of life in a data set of 756 very low birth weight (< 1501g) neonates hospitalized in seven neonatal intensive care units in Wisconsin and Iowa (Palta et al., 1991). We regress fluid output in ml/kg/day on fluid intake in ml/kg/day and the infant’s daily weight in grams. It is known that smaller infants lose more fluid through the body surface as the body surface area is larger relative to weight than for smaller infants. As body surface area is related to body volume approximately as 2:3, weight to the 2/3 power was used in the regression. Initial regression of output on intake and body weight\(^{2/3}\) yielded significant coefficients for both by ordinary and weighted least squares including a random effect for the intercept \( p < .00001 \). The regression coefficients with their standard errors are given below for the ordinary and weighted least squares procedures.
OLS          WLS
Intercept    -54.372  -57.466
Fluid intake (ml/kg/day)  0.509 (0.020)  0.480 (0.014)
Weight^{2/3} in grams    0.862 (0.054)  0.939 (0.051)

The within individual variation was estimated by (4.2) as 1363 and the between infant variation in intercept at 4.5. Various residual plots (not shown) indicated no violation of linearity assumptions in this model. It was not possible to add random slopes to the WLS model without creating oversaturation.

Comparison of the results in the above table indicates that WLS and OLS gave very similar results and that there was a small gain in standard error of the fluid intake coefficient when using WLS. There was little change in the standard error of the coefficient of weight^{2/3}. Fitting individual fixed intercepts to each infants’ data showed their distribution to be distinctly non-normal (see the qq-plot in Figure 1).

To examine model fit, and to allow for the possibility of omitted covariates, we fitted the above model with the additional terms mean intake and mean (weight^{2/3}) with these means taken across all observations on a neonate, i.e. allowing

\[ W_i = (1, X_i, \bar{x}_{ij}1, ..., \bar{x}_{ij}1), \]

where \( \bar{x}_{ij} \) is the average of the jth column of \( X_i \). Using formulas (3.2) and (4.4) we fitted OLS and WLS estimators. The resulting models are given below:

OLS          WLS
Intercept    -52.79   -52.24
Fluid intake (ml/kg/day)  0.515 (0.026)  0.506 (0.021)
Weight^{2/3} in grams    2.23 (0.226)  2.19 (0.202)
Mean fluid intake        0.033 (0.0386)  0.041 (0.0296)
Mean (weight^{2/3})      -1.445 (0.230)  -1.406 (0.210)

Again, WLS demonstrates a small gain in efficiency. Interestingly, for this particular model where only the intercept is associated with a random effect, errors are assumed independent and where the means of all time dependent covariates are included, WLS is more efficient than OLS only when numbers of observations \( k_i \) are unequal (Weerakkody and Johnson, 1992). In our example, the \( k_i \) are not equal as 42 infants had fluid measurements on less than 4 days. The variance components estimated by our method for this model were 893 for the within infant variation and 375 for the between infant variation. It is of interest
that the coefficient of mean (weight$^{2/3}$) is statistically significant and that the coefficient of weight$^{2/3}$ has changed considerably. It is easy to hypothesize variables which may confound the weight–output relationship. Cross-sectionally several treatments such as heating lamps and UV treatment for jaundice increase fluid loss through the skin. However, as these treatments would be most often administered to the smallest neonates they cannot explain the negative sign of the coefficient of mean weight$^{2/3}$. Instead, it may be suspected that within infant factors related to weight are involved. It is possible, for example, that an upward deviation in daily weight from mean weight may be a marker for fluid retention at time of weighing. This may in turn lead to change of diuretic regimen or natural diuresis. Hence there may be an influence of daily weight additional to that generated by the body surface relationship.

REFERENCES


