ASYMPTOTIC DISTRIBUTION OF THE WEIGHTED LEAST SQUARES ESTIMATOR

by
Jun Shao
Purdue University

Technical Report #88-10

Department of Statistics
Purdue University

March 1988

*

ASYMPTOTIC DISTRIBUTION OF THE WEIGHTED LEAST SQUARES ESTIMATOR

by

Jun Shao

Purdue University

SUMMARY

This paper derives the asymptotic distribution of the weighted least squares estimator (WLSE) in a heteroscedastic linear regression model. A consistent estimator of the asymptotic covariance matrix of the WLSE is also obtained. The results are obtained under weak conditions on the design matrix and some moment conditions on the error distributions. It is shown that most of the error distributions encountered in practice satisfy these moment conditions. Some examples of the asymptotic covariance matrices are also given.

Keywords. Heteroscedasticity, weighted least squares, asymptotic distribution, asymptotic covariance matrix, consistency.

1. Introduction

The following linear regression model is widely used in practice:

$$y_{ij} = x_{ij}'\beta + e_{ij}, \quad j=1,...,n_i,\ i=1,...,k,\ \sum_{i=1}^{k} n_i = n, \tag{1.1}$$

where $\beta = (\beta_1 \ldots \beta_p)'$ is the unknown parameter of interest, p is a fixed integer, y_{ij} are the responses, $x_{ij} = (x_{ij1} \ldots x_{ijp})'$ are known design vectors, and e_{ij} are independent random errors. For each i, e_{ij}, $j=1,...,n_i$, have the same distribution with mean zero and variance σ_i^2. The σ_i^2 are *unknown* and *unequal*. Model (1.1) is called a *heteroscedastic* model because of the presence of unequal scale parameters σ_i. Let

$$e = (e_{11} \ldots e_{1n_1} \ldots\ldots e_{k1} \ldots e_{kn_k})'_{n\times 1},$$

$$y = (y_{11} \ldots y_{1n_1} \ldots\ldots y_{k1} \ldots y_{kn_k})'_{n\times 1},$$

and

$$X = (x_{11} \ldots x_{1n_1} \ldots\ldots x_{k1} \ldots x_{kn_k})'_{n\times p}.$$

A matrix form of Model (1.1) is then

$$y = X\beta + e$$

with the dispersion matrix

$$D = Var(e) = block\ diag.(\sigma_1^2 I_{n_1} \ldots \sigma_k^2 I_{n_k}), \tag{1.2}$$

where I_t is the $t\times t$ identity matrix. The design matrix X is assumed to be of full rank. Note that X and y depend on the sample size n, but the subscript n is omitted for simplicity. For each i, if x_{ij} are the same for all j, then y_{ij} is the jth replicate at the ith design point. Model (1.1) is slightly more general since x_{ij}, $j=1,...,n_i$, are not assumed to be the same.

For a regression problem, it is usually difficult to obtain a large number of replicates (see [3]). Therefore, throughout the paper we assume that $n_i \le n_\infty$, $i=1,...,k$, for a fixed integer n_∞.

The ordinary least squares estimator (OLSE) of β is

$$\hat{\beta} = (X'X)^{-1}X'y. \tag{1.3}$$

The OLSE may be improved by the weighted least squares estimator (WLSE) of β:

$$\hat{\beta}^w = (X'WX)^{-1}(X'Wy), \tag{1.4}$$

where

$$W = block\ diag.(\, w_1 I_{n_1} \ldots w_k I_{n_k}\,),$$

and w_i^{-1} are estimates of σ_i^2, $i=1,...,k$.

If σ_i^2 is a smooth function of the design or the mean response, a consistent estimator of σ_i^2 $(k\to\infty)$ can be obtained and the asymptotic distribution of $\hat{\beta}^w$ is the same as $\tilde{\beta}=(X'D^{-1}X)^{-1}X'D^{-1}y$ (see [1]). Therefore, $\hat{\beta}^w$ is more efficient than $\hat{\beta}$. In the general situation where σ_i^2 is not related to the design, no consistent estimator of σ_i^2 is available unless $n_i\to\infty$. If w_i^{-1} are inconsistent estimators of σ_i^2, the asymptotic distribution of $\hat{\beta}^w$ is not well known and is different from that of $\tilde{\beta}$. Thus, $\hat{\beta}^w$ may or may not be more efficient than $\hat{\beta}$. For comparing the efficiency of $\hat{\beta}^w$ and $\hat{\beta}$ and making statistical inferences based on $\hat{\beta}^w$, it is crucial to obtain the asymptotic distribution of $\hat{\beta}^w$.

Several types of estimators of σ_i^2 are proposed by various authors (e.g., Hartley, Rao and Kiefer [5]; Rao [7]; Rao [8]; Horn, Horn and Duncan [6]; Fuller and Rao [3]; Shao [9]) for the case where σ_i^2 is not assumed to be related to the design. The empirical results in [9] show that the WLSE is more efficient when the following estimator of σ_i^2 is used:

$$v_i^b = n_i^{-1}\textstyle\sum_{j=1}^{n_i} r_{ij}^2 + h_i s^2, \tag{1.5}$$

where $r_{ij} = y_{ij}-x_{ij}'\hat{\beta}$, $s^2=(n-p)^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} r_{ij}^2$ and $h_i=n_i^{-1}\sum_{j=1}^{n_i} x_{ij}'(X'X)^{-1}x_{ij}$.

Fuller and Rao [3] derived the asymptotic distribution of $\hat{\beta}^w$ with $w_i^{-1}=v_i$, where

$$v_i = n_i^{-1}\textstyle\sum_{j=1}^{n_i} r_{ij}^2. \tag{1.6}$$

However, the assumptions they made are rather restrictive: (1) the errors e_{ij} are normally distributed; (2) several matrices, such as $k^{-1}X'X$ and $k^{-1}X'DX$, converge to positive definite matrices. Since σ_i^2 are unknown and unequal, it is usually difficult to check if $k^{-1}X'DX$ has a positive definite limiting matrix.

Under a weak condition on the design matrix X and some moment conditions on the error distributions, this paper derives the asymptotic distribution of $\hat{\beta}^w$ with the reciprocals of the estimators (1.5) or (1.6) as weights.

The rest of the paper is organized as follows. In Section 2, we state some assumptions and prove some technical lemmas. The main results are established in Section 3. The last section studies some examples and derives a large class of error distributions satisfying the assumptions for the asymptotic theory.

2. Preliminaries

The following assumptions are used in deriving the asymptotic distribution of $\hat{\beta}^w$:

Assumption A. There are positive constants σ_0, σ_∞ and c_∞ and positive integers n_0 and n_∞ such that $\sigma_0 \le \sigma_i \le \sigma_\infty$, $n_0 \le n_i \le n_\infty$ and $\|x_{ij}\| \le c_\infty$ for all i and j, where $\|x\| = (x'x)^{1/2}$ is the Euclidean norm of x.

Assumption B. There is a positive constant c_0 such that

$$c_0 \le k^{-1}(\text{ the minimum eigenvalue of } X'X \text{ }).$$

Assumption C. The errors e_{ij} satisfy the following moment conditions:

$$E(e_{i1}/\textstyle\sum_{j=1}^{n_i} e_{ij}^2) = 0 \quad \text{and} \quad E[e_{i1}e_{i2}/(\sum_{j=1}^{n_i} e_{ij}^2)^t] = 0 \tag{2.1}$$

for all i and t=1,2, and

$$E\,|e_{i1}|^{2+\delta} \le b \quad \text{and} \quad E(\textstyle\sum_{j=1}^{n_0} e_{ij}^2)^{-(1+\delta)} \le b \tag{2.2}$$

for all i, where b and $\delta < 1/2$ are positive constants.

Assumption B together with the boundedness of $\|x_{ij}\|$ imply that $X'X$ diverges to infinity at the rate $O(k)$, i.e., there are positive constants c_1 and c_2 such that

$$c_1 I_p \le k^{-1}X'X \le c_2 I_p \quad \text{for } all \ k. \tag{2.3}$$

This is much weaker than $k^{-1}X'X$ converging to a positive definite matrix, a condition assumed in [3]. A sufficient condition for (2.1) is that the distributions of e_{ij} are symmetric about zero. It is shown in Proposition 1 of Section 4 that most of the error distributions encountered in practice satisfy (2.2). Define

$$\tau(n_i) = \sigma_i^2 E\,(\textstyle\sum_{j=1}^{n_i} e_{ij}^2)^{-1}. \tag{2.4}$$

Examples of the functions τ can be found in Section 4. It is easy to see that under Assumption A and (2.2),

$$0 < \tau(n_\infty) \leq \tau(n_i) \leq \tau(n_0) < \infty \quad \text{for } all\ i.$$

The results in Lemma 1 can be found in probability theory. The proofs are omitted.

Lemma 1. (i) Let $\{\ \xi_i\ \}$ be a sequence of independent random variables with $E\xi_i=\mu_i$. Put $\bar{\xi}=k^{-1}\sum_{i=1}^k \xi_i$ and $\bar{\mu}=k^{-1}\sum_{i=1}^k \mu_i$. If for a positive $\delta \leq 1$, $k^{-(1+\delta)}\sum_{i=1}^k E\,|\xi_i|^{1+\delta} \to 0$, then $\bar{\xi}-\bar{\mu} \to_p 0$, where $\to_p$ denotes convergence in probability.

(ii) Let $\{\ \xi_i\ \}$ be a sequence of random variables with $\sup_i E\,|\xi_i|^{1+\delta} < \infty$ for a $\delta>0$. Then $\lim_{k\to\infty} k^{-1}E\,[\max_{i\leq k} |\xi_i|]=0$.

Lemma 2. Suppose that Assumption A and (2.2) hold. Let $\{\ z_i\ \}$ be a nonrandom sequence such that $|z_i| \leq z_\infty$ for all i. Then

$$k^{-1}\sum_{i=1}^k z_i u_i^{-1} - k^{-1}\sum_{i=1}^k z_i n_i \tau(n_i)\sigma_i^{-2} \to_p 0,$$

where

$$u_i = n_i^{-1}\sum_{j=1}^{n_i} e_{ij}^2. \tag{2.5}$$

Proof. From (2.4), $E\,(z_i u_i^{-1}) = z_i n_i \tau(n_i)\sigma_i^{-2}$. From Assumption A and (2.2),

$$k^{-(1+\delta)}\sum_{i=1}^k E\,|z_i u_i^{-1}|^{1+\delta} \leq z_\infty^{1+\delta} k^{-(1+\delta)}\sum_{i=1}^k E u_i^{-(1+\delta)} \leq z_\infty^{1+\delta} n_\infty^{1+\delta} b k^{-\delta}.$$

The result follows from Lemma 1(i). □

Lemma 3. Suppose that Assumptions A and (2.2) hold. Let $\alpha > [2(1+\delta)]^{-1}$, where δ is given in (2.2). Then

$$k^{-3\alpha}\sum_{i=1}^{k}|e_{i1}|(\sum_{j=1}^{n_i}e_{ij}^2)^{-2}\rightarrow_p 0$$

and for any $1\le h\le n_i$,

$$k^{-3\alpha}\sum_{i=1}^{k}e_{i1}^2|e_{ih}|(\sum_{j=1}^{n_i}e_{ij}^2)^{-3}\rightarrow_p 0.$$

Proof. Let $\zeta_i=(\sum_{j=1}^{n_0}e_{ij}^2)^{-1}$. Then

$$|e_{i1}|(\sum_{j=1}^{n_i}e_{ij}^2)^{-2}\le(\sum_{j=1}^{n_i}e_{ij}^2)^{-3/2}\le\zeta_i^{3/2}$$

and

$$e_{i1}^2|e_{ih}|(\sum_{j=1}^{n_i}e_{ij}^2)^{-3}\le|e_{ih}|(\sum_{j=1}^{n_i}e_{ij}^2)^{-2}\le\zeta_i^{3/2}.$$

Hence, the result follows from

$$k^{-3\alpha}\sum_{i=1}^{k}\zeta_i^{3/2}\rightarrow_p 0,$$

which is implied by

$$E[(k^{-3\alpha}\sum_{i=1}^{k}\zeta_i^{3/2})^{2(1+\delta)/3}]\le k^{-2\alpha(1+\delta)}\sum_{i=1}^{k}E\,\zeta_i^{1+\delta}\le bk^{1-2\alpha(1+\delta)}\rightarrow 0$$

since $\alpha>[2(1+\delta)]^{-1}$. □

Lemma 4. Suppose that Assumptions A and B and (2.2) hold. Let

$$H=block\ diag.(\,z_1v_1^{-1}I_{n_1}\ldots z_kv_k^{-1}I_{n_k}\,)$$

and

$$G=block\ diag.(\,z_1n_1\tau(n_1)\sigma_1^{-2}I_{n_1}\ldots z_kn_k\tau(n_k)\sigma_k^{-2}I_{n_k}\,),$$

where v_i is defined in (1.6) and z_i satisfying $0<z_0\le z_i\le z_\infty<\infty$. Then

$$k^{-1}X'HX-k^{-1}X'GX\rightarrow_p 0.$$

Proof. The (t,s)th element of $k^{-1}X'HX$ is

$$k^{-1}\sum_{i=1}^{k}z_iv_i^{-1}\sum_{j=1}^{n_i}x_{ijt}x_{ijs}.$$

Let u_i be defined in (2.5). Then

$$|k^{-1}\sum_{i=1}^{k}z_i(v_i^{-1}-u_i^{-1})\sum_{j=1}^{n_i}x_{ijt}x_{ijs}|\le Ck^{-1}\sum_{i=1}^{k}|v_i^{-1}-u_i^{-1}|,$$

where C is a positive constant. From Lemma 2,

$$k^{-1}\sum_{i=1}^{k}z_iu_i^{-1}\sum_{j=1}^{n_i}x_{ijt}x_{ijs}-k^{-1}\sum_{i=1}^{k}z_in_i\tau(n_i)\sigma_i^{-2}\sum_{j=1}^{n_i}x_{ijt}x_{ijs}\rightarrow_p 0.$$

Hence the result follows from

$$k^{-1}\sum_{i=1}^{k}|v_i^{-1}-u_i^{-1}|\rightarrow_p 0.$$

Fix an α such that $[2(1+\delta)]^{-1} < \alpha < 2^{-1}$. Let $B_{ik} = \{ \max_j |e_{ij}| < k^{-\alpha} \}$, $A_{k1} = \cup_{i \le k} B_{ik}$, $A_{k2} = \{ \max_{i,j} |\phi_{ij}| > 2^{-1} n_\infty^{-1/2} k^{-\alpha} \}$ and $A_k = A_{k1} \cup A_{k2}$, where $\phi_{ij} = x_{ij}'(\hat{\beta} - \beta)$. Let A^c be the complement of the set A. On A_k^c,

$$\sum_{j=1}^{n_i} |\phi_{ij}| \le n_i 2^{-1} n_\infty^{-1/2} k^{-\alpha} \le 2^{-1} n_\infty^{1/2} \max_j |e_{ij}| \le 2^{-1} n_\infty^{1/2} \sum_{j=1}^{n_i} |e_{ij}|$$

and

$$\sum_{j=1}^{n_i} \phi_{ij}^2 \le n_i 4^{-1} n_\infty^{-1} k^{-2\alpha} \le 4^{-1} \max_j e_{ij}^2 \le 4^{-1} \sum_{j=1}^{n_i} e_{ij}^2,$$

which implies

$$v_i = n_i^{-1} \sum_{j=1}^{n_i} (e_{ij} - \phi_{ij})^2 \ge 2^{-1} n_i^{-1} \sum_{j=1}^{n_i} e_{ij}^2 - n_i^{-1} \sum_{j=1}^{n_i} \phi_{ij}^2 \ge 4^{-1} n_i^{-1} \sum_{j=1}^{n_i} e_{ij}^2 = 4^{-1} u_i .$$

Hence there is a constant $C > 0$ such that on A_k^c,

$$|v_i^{-1} - u_i^{-1}| = n_i^{-1} |u_i^{-1} v_i^{-1} \sum_{j=1}^{n_i} (2e_{ij} - \phi_{ij}) \phi_{ij}| \le C k^{-\alpha} u_i^{-2} \sum_{j=1}^{n_i} |e_{ij}| . \tag{2.6}$$

Thus, from Lemma 3,

$$\begin{aligned} k^{-1} [\sum_{i=1}^{k} |u_i^{-1} - v_i^{-1}|] I(A_k^c) &\le C k^{-1-\alpha} [\sum_{i=1}^{k} \sum_{j=1}^{n_i} |e_{ij}| u_i^{-2}] I(A_k^c) \\ &\le C k^{-3\alpha} \sum_{i=1}^{k} \sum_{j=1}^{n_i} |e_{ij}| u_i^{-2} \to_p 0, \end{aligned}$$

where $I(A)$ is the indicator function of the set A. It remains to show $P(A_k) \to 0$. Note that

$$P(A_k) \le \sum_{i=1}^{k} P(B_{ik}) + P(A_{k2})$$

and

$$\begin{aligned} \sum_{i=1}^{k} P(B_{ik}) &= \sum_{i=1}^{k} P(k^{\alpha} < (\max_j |e_{ij}|)^{-1}) \le \sum_{i=1}^{k} P(n_i^{-1} k^{2\alpha} < (\sum_{j=1}^{n_i} e_{ij}^2)^{-1}) \\ &\le \sum_{i=1}^{k} n_i^{(1+\delta)} k^{-2(1+\delta)\alpha} E (\sum_{j=1}^{n_0} e_{ij}^2)^{-(1+\delta)} \le b n_\infty^{(1+\delta)} k^{1-2(1+\delta)\alpha} \to 0. \end{aligned}$$

Under Assumptions A and B, $\|\hat{\beta} - \beta\| = O_p(k^{-1/2})$. Hence $k^{\alpha} \max_{i,j} |\phi_{ij}| \le c_\infty k^{\alpha} \|\hat{\beta} - \beta\| \to_p 0$ and

$$P(A_{k2}) \to 0.$$

This completes the proof. □

Lemma 5. Suppose that Assumptions A and B and (2.2) hold. Let $\phi_{ij} = x_{ij}'(\hat{\beta} - \beta)$. Then

$$R_{t1} = \sum_{i=1}^{k} n_i^{-1} u_i^{-2} \sum_{j=1}^{n_i} \phi_{ij}^2 \sum_{h=1}^{n_i} x_{iht} e_{ih} = o_p(k^{1/2}) \tag{2.7}$$

and

$$R_{t2} = \sum_{i=1}^{k} n_i^{-2} u_i^{-2} v_i^{-1} [\sum_{j=1}^{n_i} (2e_{ij} - \phi_{ij}) \phi_{ij}]^2 \sum_{h=1}^{n_i} x_{iht} e_{ih} = o_p(k^{1/2}). \qquad (2.8)$$

Proof. From Assumption A, there is a constant $C > 0$ such that

$$k^{-1/2} |R_{t1}| \leq C k^{-1/2} \| \hat{\beta} - \beta \|^2 \sum_{i=1}^{k} \sum_{h=1}^{n_i} |e_{ih}| u_i^{-2}.$$

Since $\| \hat{\beta} - \beta \|^2 = O_p(k^{-1})$, (2.7) follows from Lemma 3. Let A_k be defined as in the proof of Lemma 4. From (2.6), there are positive constants C_1 and C_2 such that

$$k^{-1/2} |R_{t2}| I(A_k^c) \leq C_1 k^{-1/2} (\max_{i,j} \phi_{ij}^2) \sum_{i=1}^{k} u_i^{-3} \sum_{h=1}^{n_i} |e_{ih}| (\sum_{j=1}^{n_i} |e_{ij}|)^2$$

$$\leq C_2 k^{-1/2} \| \hat{\beta} - \beta \|^2 \sum_{i=1}^{k} u_i^{-3} \sum_{j=1}^{n_i} \sum_{h=1}^{n_i} e_{ij}^2 |e_{ih}| \rightarrow_p 0$$

by Lemma 3 and $\| \hat{\beta} - \beta \|^2 = O_p(k^{-1})$. From the proof of Lemma 4, $P(A_k) \rightarrow 0$. Hence (2.8) holds. □

3. The main results

Let $\tau(n_i)$ be defined in (2.4). Define

$$D_1 = block\ diag.(\ \sigma_1^{-2} n_1 \tau(n_1) I_{n_1} \dots \sigma_k^{-2} n_k \tau(n_k) I_{n_k}\ ),$$

$$D_2 = block\ diag.(\ \sigma_1^{-2} \tau(n_1) I_{n_1} \dots \sigma_k^{-2} \tau(n_k) I_{n_k}\ ),$$

and

$$\Sigma_k = X'D_1X + 4X'D_2X + 4X'D_2X(X'X)^{-1}X'DX(X'X)^{-1}X'D_2X,$$

where D is given by (1.2). Note that under Assumptions A and B, there are positive constants C_0 and C_1 such that

$$C_0 I_p \leq k^{-1} \Sigma_k \leq C_1 I_p. \qquad (3.1)$$

We first derive the asymptotic distribution of $\hat{\beta}^w$ with $w_i^{-1} = v_i$ given by (1.6).

Theorem 1. Suppose that Assumptions A-C hold. Let $\hat{\beta}^w$ be defined in (1.4) with $w_i^{-1} = v_i$. Then

$$V_k^{-1/2}(\hat{\beta}^w - \beta) \to_d N(0, I_p), \tag{3.2}$$

where $\to_d$ denotes convergence in distribution, $V_k^{-1/2} = (V_k^{1/2})^{-1}$ and $V_k^{1/2}$ is a square root of

$$V_k = (X'D_1X)^{-1}\Sigma_k(X'D_1X)^{-1}. \tag{3.3}$$

Proof. We first show that

$$\Sigma_k^{-1/2}T_k \to_d N(0, I_p), \tag{3.4}$$

where

$$T_k = X'\tilde{W}e + 2X'D_2X(X'X)^{-1}X'e$$

with

$$\tilde{W} = block\ \ diag.(\, u_1^{-1}\ I_{n_1} \ldots\ u_k^{-1}\ I_{n_k}\,).$$

Under (2.1), $ET_k = 0$. Then

$$Var(T_k) = E(T_kT_k') = E(X'\tilde{W}ee'\tilde{W}X) + 2E[X'\tilde{W}ee'X(X'X)^{-1}X'D_2X]$$

$$+ 2E[X'D_2X(X'X)^{-1}X'ee'\tilde{W}X] + 4[X'D_2X(X'X)^{-1}X'DX(X'X)^{-1}X'D_2X].$$

Since $u_i^{-2}e_{ij}^2$, $j=1,\ldots,n_i$, have the same distribution,

$$E(u_i^{-2}e_{ij}^2) = n_i^{-1}\textstyle\sum_{j=1}^{n_i}E(u_i^{-2}e_{ij}^2) = Eu_i^{-1} = \sigma_i^{-2}n_i\tau(n_i). \tag{3.5}$$

Then the expected value of the (t, s)th element of $X'\tilde{W}ee'\tilde{W}X$ is

$$E[\textstyle\sum_{i=1}^k u_i^{-2}(\sum_{j=1}^{n_i}x_{ijt}e_{ij})(\sum_{j=1}^{n_i}x_{ijs}e_{ij})] = \sum_{i=1}^k\sum_{j=1}^{n_i}\sigma_i^{-2}n_i\tau(n_i)x_{ijt}x_{ijs}.$$

Hence

$$E(X'\tilde{W}ee'\tilde{W}X) = X'D_1X.$$

Similarly, the expected value of the (t, s)th element of $X'\tilde{W}ee'X$ is

$$E[\textstyle\sum_{i=1}^k u_i^{-1}(\sum_{j=1}^{n_i}x_{ijt}e_{ij})(\sum_{j=1}^{n_i}x_{ijs}e_{ij})] = \sum_{i=1}^k\sum_{j=1}^{n_i}E(u_i^{-1}e_{ij}^2)x_{ijt}x_{ijs} = \sum_{i=1}^k\sum_{j=1}^{n_i}x_{ijt}x_{ijs}$$

since $E(u_i^{-1}e_{ij}^2) = n_i^{-1}\sum_{j=1}^{n_i}E(u_i^{-1}e_{ij}^2) = 1$. Hence

$$E[X'D_2X(X'X)^{-1}X'ee'\tilde{W}X] = X'D_2X(X'X)^{-1}X'X = X'D_2X.$$

This shows

$$Var(T_k) = \Sigma_k.$$

Let l be a fixed nonzero p-vector and $\lambda_k = \Sigma_k^{-1/2} l / (l' \Sigma_k^{-1} l)^{1/2}$. Then $\| \lambda_k \| = 1$ and

$$d_k^2 = Var(\lambda_k' T_k) = (l' \Sigma_k^{-1} l)^{-1} l' \Sigma_k^{-1/2} \Sigma_k \Sigma_k^{-1/2} l = (l' \Sigma_k^{-1} l)^{-1} l' l.$$

Let λ_{ks} and η_{ks} be the sth elements of λ_k and $2(X'X)^{-1} X' D_2 X \lambda_k$, respectively. Then

$$\lambda_k' T_k = \sum_{i=1}^{k} \sum_{s=1}^{p} \sum_{j=1}^{n_i} (\lambda_{ks} u_i^{-1} x_{ijs} e_{ij} + \eta_{ks} x_{ijs} e_{ij}).$$

From Assumption A, $\| \eta_k \|^2 = 4 \lambda_k' X' D_2 X (X'X)^{-2} X' D_2 X \lambda_k \le C_1$, where C_1 is a positive constant. Let δ be given in (2.2). Then

$$E \, | \lambda_{ks} u_i^{-1} x_{ijs} e_{ij} + \eta_{ks} x_{ijs} e_{ij} |^{2+\delta} \le C_2 (E u_i^{-(1+\delta/2)} + E \, | e_{i1} |^{2+\delta}) \le C_3,$$

where C_2 and C_3 are positive constants. Hence

$$d_k^{-(2+\delta)} \sum_{i=1}^{k} \sum_{s=1}^{p} \sum_{j=1}^{n_i} E \, | \lambda_{ks} u_i^{-1} x_{ijs} e_{ij} + \eta_{ks} x_{ijs} e_{ij} |^{2+\delta} \le C_3 n_{\infty} p k d_k^{-(2+\delta)}.$$

From (3.1), $k d_k^{-(2+\delta)} \le C_0^{-1-\delta/2} k^{-\delta/2} \to 0$. Hence the Lindeberg's condition holds and therefore

$$d_k^{-1} \lambda_k' T_k = l' \Sigma_k^{-1/2} T_k / (l'l)^{1/2} \to_d N(0, 1).$$

Since l is arbitrary, the proof of (3.4) is completed.

Next, we show that

$$X'We - T_k = o_p(k^{1/2}). \tag{3.6}$$

From $v_i^{-1} = u_i^{-1} - u_i^{-2}(v_i - u_i) + u_i^{-2} v_i^{-1} (v_i - u_i)^2$, the tth element of $X'We$ is

$$\sum_{i=1}^{k} v_i^{-1} \sum_{j=1}^{n_i} x_{ijt} e_{ij} = \sum_{i=1}^{k} u_i^{-1} \sum_{j=1}^{n_i} x_{ijt} e_{ij} - R_{t1} + R_{t2} + 2R_{t3},$$

where R_{t1} and R_{t2} are defined in (2.7) and (2.8), respectively,

$$R_{t3} = \sum_{s=1}^{p} \sum_{i=1}^{k} n_i^{-1} u_i^{-2} (\sum_{j=1}^{n_i} x_{ijt} e_{ij}) (\sum_{h=1}^{n_i} x_{ihs} e_{ih}) (\hat{\beta}_s - \beta_s),$$

and $\hat{\beta}_s$ is the sth element of $\hat{\beta}$. From Lemma 5, $R_{t1} = o_p(k^{1/2})$ and $R_{t2} = o_p(k^{1/2})$. Note that the tth elements of $X'\tilde{W}e$ and $X' D_2 X (X'X)^{-1} X'e$ are, respectively,

$$\sum_{i=1}^{k} u_i^{-1} \sum_{j=1}^{n_i} x_{ijt} e_{ij} \quad \text{and} \quad \sum_{s=1}^{p} \sigma_i^{-2} \tau(n_i) \sum_{j=1}^{n_i} x_{ijt} x_{ijs} (\hat{\beta}_s - \beta_s).$$

Hence it remains to show that

$$\sum_{s=1}^{p}\sum_{i=1}^{k}\sigma_i^{-2}\tau(n_i)\sum_{j=1}^{n_i}x_{ijt}x_{ijs}(\hat{\beta}_s-\beta_s) - R_{t3} = o_p(k^{1/2}). \tag{3.7}$$

Since $\|\hat{\beta}-\beta\|=O_p(k^{-1/2})$, (3.7) follows from

$$\sum_{i=1}^{k}\sigma_i^{-2}\tau(n_i)\sum_{j=1}^{n_i}x_{ijt}x_{ijs} - \sum_{i=1}^{k}n_i^{-1}u_i^{-2}(\sum_{j=1}^{n_i}x_{ijt}e_{ij})(\sum_{h=1}^{n_i}x_{ihs}e_{ih}) = o_p(k). \tag{3.8}$$

From (2.1) and (3.5),

$$E[n_i^{-1}u_i^{-2}(\sum_{j=1}^{n_i}x_{ijt}e_{ij})(\sum_{h=1}^{n_i}x_{ihs}e_{ih})] = \sigma_i^{-2}\tau(n_i)\sum_{j=1}^{n_i}x_{ijt}x_{ijs}.$$

Also, $|n_i^{-1}u_i^{-2}(\sum_{j=1}^{n_i}x_{ijt}e_{ij})(\sum_{h=1}^{n_i}x_{ihs}e_{ih})|\le c_\infty^2 n_i^{-1}u_i^{-2}(\sum_{j=1}^{n_i}|e_{ij}|)^2\le c_\infty^2 u_i^{-1}$. Hence (3.8) follows from Lemma 1(i). This proves (3.6).

Finally, from (3.4) and (3.6),

$$\Sigma_k^{-1/2}X'We \to_d N(0, I_p). \tag{3.9}$$

Since $\hat{\beta}^w-\beta=(X'WX)^{-1}X'We$,

$$\Sigma_k^{-1/2}(X'D_1X)(\hat{\beta}^w-\beta) = \Sigma_k^{-1/2}(X'D_1X)(X'WX)^{-1}\Sigma_k^{1/2}\Sigma_k^{-1/2}X'We. \tag{3.10}$$

Note that

$$\Sigma_k^{-1/2}(X'D_1X)(X'WX)^{-1}\Sigma_k^{1/2} = \Sigma_k^{-1/2}(X'D_1X)[(X'WX)^{-1} - (X'D_1X)^{-1}]\Sigma_k^{1/2} + I_p.$$

From Lemma 4, $k[(X'WX)^{-1} - (X'D_1X)^{-1}] \to_p 0$. From (3.1) and $X'D_1X \le n_\infty\sigma_0^{-2}\tau(n_0)X'X$,

$$\Sigma_k^{-1/2}(X'D_1X)(X'WX)^{-1}\Sigma_k^{1/2} \to_p I_p.$$

Thus, from (3.9) and (3.10),

$$\Sigma_k^{-1/2}(X'D_1X)(\hat{\beta}^w-\beta) \to_d N(0, I_p). \tag{3.11}$$

That is, we have shown (3.2) when $(X'D_1X)^{-1}\Sigma_k^{1/2}$ is taken to be a square root of V_k. For an arbitrary square root $V_k^{1/2}$, (3.2) follows from (3.11) by using the same argument as in [2] (p. 349). This completes the proof. □

The next theorem shows that (3.2) also holds for $\hat{\beta}^w$ with σ_i^2 estimated by v_i^b given in (1.5).

Theorem 2. Suppose that Assumptions A-C hold. Let $\hat{\beta}^w$ be the WLSE with $w_i^{-1}=v_i^b$. Then

$$V_k^{-1/2}(\hat{\beta}^w-\beta) \to_d N(0, I_p).$$

Proof. Let

$$W_1 = block\ diag.(\ v_1^{-1}\ I_{n_1} \ldots\ v_k^{-1}\ I_{n_k}\ )$$

and

$$W_2 = block\ diag.(\ (v_1^b)^{-1}\ I_{n_1} \ldots\ (v_k^b)^{-1}\ I_{n_k}\ ).$$

The result follows from

$$X'W_2e = T_k + o_p(k^{1/2}) \tag{3.12}$$

and

$$(X'W_2X)^{-1}(X'W_1X) \to_p I_p, \tag{3.13}$$

where T_k is defined in the proof of Theorem 1. Following the proofs of Lemma 5 and Theorem 1, (3.12) follows from

$$\max_{i\le k} |v_i^b - v_i| = O_p(k^{-1}),$$

which is implied by $\max_{i\le k} h_i = O(k^{-1})$ under Assumptions A and B. For (3.13), it suffices to show that

$$\max_{i\le k} |v_i^b v_i^{-1} - 1| \to_p 0, \tag{3.14}$$

which follows from

$$\max_{i\le k} h_i v_i^{-1} \to_p 0.$$

Since $\max_{i\le k} h_i = O(k^{-1})$, it remains to show

$$k^{-1}\max_{i\le k} v_i^{-1} \to_p 0. \tag{3.15}$$

Let $B_k=\{\ \max_{i\le k} |\Delta_i u_i^{-1}| > 1/2\ \}$, where $\Delta_i=v_i-u_i=n_i^{-1}\sum_{j=1}^{n_i}\phi_{ij}^2-2n_i^{-1}\sum_{j=1}^{n_i}\phi_{ij}e_{ij}$. On B_k^c,

$$\max_{i\le k} v_i^{-1} \le 2^{-1}\max_{i\le k} u_i^{-1} \le 2^{-1}n_\infty \max_{i\le k}(\textstyle\sum_{j=1}^{n_0}e_{ij}^2)^{-1}.$$

Since $k^{-1}\max_{i\le k}(\sum_{j=1}^{n_0}e_{ij}^2)^{-1} \to_p 0$ by Lemma 1(ii), (3.15) follows from $P(B_k)\to 0$. Note that

$$|\Delta_i u_i^{-1}| \le n_i^{-1}\textstyle\sum_{j=1}^{n_i}[x_{ij}'(\hat{\beta}-\beta)]^2 u_i^{-1} + 2\{n_i^{-1}\sum_{j=1}^{n_i}[x_{ij}'(\hat{\beta}-\beta)]^2\}^{1/2}u_i^{-1/2}.$$

Hence $P(B_k) \to 0$ follows from $\|\hat{\beta}-\beta\|^2 = O_p(k^{-1})$ and

$$\max_{i\le k}\{n_i^{-1}\sum_{j=1}^{n_i}[x_{ij}'(\hat{\beta}-\beta)]^2 u_i^{-1}\} \le c_\infty^2 \|\hat{\beta}-\beta\|^2 \max_{i\le k} u_i^{-1}$$

$$\le c_\infty^2 n_\infty \|\hat{\beta}-\beta\|^2 \max_{i\le k}(\sum_{j=1}^{n_0} e_{ij}^2)^{-1} \to_p 0. \quad \square$$

From the above theorems, V_k defined in (3.3) is the asymptotic covariance matrix of $\hat{\beta}^w$. A consistent estimator of V_k is required for making statistical inference based on $\hat{\beta}^w$. Let w_i^{-1} be either v_i or v_i^b,

$$U = block\ diag.(\ n_1^{-1} w_1 I_{n_1} \ldots n_k^{-1} w_k I_{n_k}\ )$$

and

$$\hat{V}_k = (X'WX)^{-1} + 4(X'WX)^{-1}X'UX(X'WX)^{-1}$$
$$+ 4(X'WX)^{-1}X'UX(X'X)^{-1}X'W^{-1}X(X'X)^{-1}X'UX(X'WX)^{-1}.$$

The following theorem shows that $\hat{V}_k$ is consistent for V_k.

Theorem 3. Suppose that Assumptions A-C hold. Then

$$k(\hat{V}_k - V_k) \to_p 0.$$

Proof. From Lemma 4 and (3.14),

$$k^{-1}X'WX - k^{-1}X'D_1X \to_p 0$$

and

$$k^{-1}X'UX - k^{-1}X'D_2X \to_p 0.$$

It remains to show that

$$k^{-1}X'W^{-1}X - k^{-1}X'DX \to_p 0. \tag{3.16}$$

When $w_i^{-1} = v_i$, the (t, s)th element of $k^{-1}X'W^{-1}X$ is $k^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ijt}x_{ijs}v_i$. Let u_i be defined as in (2.5). From Lemma 1(i),

$$k^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ijt}x_{ijs}u_i - k^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ijt}x_{ijs}\sigma_i^2 \to_p 0.$$

From the proof of Theorem 2, $\max_{i\le k}|v_i u_i^{-1} - 1| = \max_{i\le k}|\Delta_i u_i^{-1}| = o_p(1)$. Hence (3.16) holds. The same argument shows that (3.16) also holds if $w_i^{-1} = v_i^b$. This completes the proof. $\square$

In some situations the parameter of interest is $\theta = g(\beta)$, where g is a function from $\mathbf{R}^p$ to $\mathbf{R}^q$ and is differentiable at β. A natural estimator of θ is $\hat{\theta} = g(\hat{\beta}^w)$. Let

$$V_k^g = \nabla g(\beta) V_k (\nabla g(\beta))',$$

where $\nabla g(\beta)$ is the gradient of g at β. We have the following result.

Theorem 4. Let $\hat{\theta} = g(\hat{\beta}^w)$ with $w_i^{-1}=$ either v_i or v_i^b. Suppose that Assumptions A-C hold, ∇g is continuous at β and $\nabla g(\beta)$ is of full rank. Then

$$(V_k^g)^{-1/2}(\hat{\theta}-\theta) \to_d N(0, I_q) \tag{3.17}$$

and

$$k\hat{V}_k^g - kV_k^g \to_p 0, \tag{3.18}$$

where $\hat{V}_k^g = \nabla g(\hat{\beta}^w)\hat{V}_k(\nabla g(\hat{\beta}^w))'$.

Proof. From Theorems 1 and 2 and the continuity of ∇g,

$$\hat{\theta}-\theta = \nabla g(\beta)(\hat{\beta}^w - \beta) + o_p(k^{-1/2}).$$

Let l be a fixed nonzero q-vector, $l_k = (V_k^g)^{-1/2} l$ and $\lambda_k = V_k^{1/2}(\nabla g(\beta))' l_k / (l_k' V_k^g l_k)^{1/2}$. Then

$$l'(V_k^g)^{-1/2}(\hat{\theta}-\theta)/(l'l)^{1/2} = l_k' \nabla g(\beta)(\hat{\beta}^w - \beta)/(l_k' V_k^g l_k)^{1/2} + l'(V_k^g)^{-1/2} o_p(k^{-1/2})/(l'l)^{1/2}$$

$$= \lambda_k' V_k^{-1/2}(\hat{\beta}^w - \beta) + o_p(1) \to_d N(0, 1)$$

by Theorems 1 and 2 and $\|\lambda_k\| = 1$. This proves (3.17). (3.18) follows from Theorem 3 and the continuity of ∇g. □

Let M denote a nonnegative definite matrix. If the map $M \to M^{1/2}$ is continuous with respect to the norm $\|M\| = [\mathrm{trace}(M'M)]^{1/2}$, then the square root $M^{1/2}$ is said to be continuous. An example of continuous square root is the Cholesky square root.

Statistical inferences such as testing hypothesis and setting confidence region for θ can be made by using the following result.

Corollary 1. Let $(\hat{V}_k^g)^{1/2}$ be a continuous square root of $\hat{V}_k^g$. Under the conditions of Theorem 4, we have

$$(\hat{V}_k^g)^{-1/2}(\hat{\theta}-\theta) \to_d N(0, I_q).$$

Proof. From (3.17), it suffices to show that

$$(\hat{V}_k^g)^{1/2}(V_k^g)^{-1/2} - I_q \to_p 0. \tag{3.19}$$

From (2.3) and (3.18), $(V_k^g)^{-1/2}\hat{V}_k^g(V_k^g)^{-1/2} - I_q \to_p 0$. Then (3.19) follows from (2.3), the continuity of the square root $(\hat{V}_k^g)^{1/2}$ and the result in [4] (page 85). □

4. Some examples of error distributions

The following are examples of the distributions of e_{ij} and the function τ defined in (2.4).

Example 1. If e_{ij} have normal distribution $N(0, \sigma_i^2)$ and $n_i \ge n_0 = 3$, then Assumption C holds and

$$\tau(n_i) = (n_i - 2)^{-1}.$$

Thus, the result in [3] is a special case of our results.

Example 2. Let the distribution of e_{ij} have a density

$$f_{ij}(t) = [\Gamma(\alpha)]^{-1}\lambda_i^{-\alpha}|t|^{2\alpha-1}\exp(-t^2/\lambda_i),$$

where $\alpha>0$ and $\lambda_i=\sigma_i^2/\alpha$. Then the density of $\sum_{j=1}^{n_i} e_{ij}^2$ is

$$f_i(t) = [\Gamma(n_i\alpha)]^{-1}\lambda_i^{-n_i\alpha}t^{n_i\alpha-1}\exp(-t/\lambda_i), \quad t \ge 0.$$

Thus, Assumption C holds if and only if

$$n_i\alpha \ge n_0\alpha > 1, \tag{4.1}$$

and if (4.1) holds,

$$\tau(n_i) = \alpha(n_i\alpha - 1)^{-1}.$$

When $\alpha>1$, (4.1) holds for $n_0=1$ and therefore there is no restriction on n_i's.

Example 3. Let e_{ij} be uniformly distributed on $[-3^{1/2}\sigma_i,\ 3^{1/2}\sigma_i]$. Then

$$Ee_{ij}^{-2}=\infty.$$

Note that $\varepsilon_{ij}=\sigma_i^{-1}e_{ij}$ are uniformly distributed on $[-3^{1/2},\ 3^{1/2}]$. Let $0\le\delta<1/2$, $m\ge2$ be an integer, $c=(3m)^{-(1+\delta)}$ and $d=3^{-(1+\delta)}$. Then

$$E(\textstyle\sum_{j=1}^{m}\varepsilon_{ij}^2)^{-(1+\delta)}=\int_c^\infty P(\sum_{j=1}^{m}\varepsilon_{ij}^2<t^{-1/(1+\delta)})dt \tag{4.2}$$

$$=(\pi/12)^{m/2}[\Gamma(m/2+1)]^{-1}\int_d^\infty t^{-m/(2+2\delta)}dt+\int_c^d P(\textstyle\sum_{j=1}^{m}\varepsilon_{ij}^2<t^{-1/(1+\delta)})dt, \tag{4.3}$$

which is infinity if $m=2$. When $m\ge3$, the first term in (4.3) is equal to

$$(\pi/12)^{m/2}[\Gamma(m/2+1)]^{-1}[m/(2+2\delta)-1]^{-1}3^{m/2-(1+\delta)}. \tag{4.4}$$

Thus, Assumption C holds if and only if $n_i\ge n_0=3$. If $n_0\ge3$,

$$\tau(n_i)=3^{-1}(\pi/4)^{n_i/2}[\Gamma(n_i/2+1)]^{-1}(n_i/2-1)^{-1}+\Delta(n_i),$$

where $\Delta(n_i)=\int_3^{3n_i}s^{-2}P(\sum_{j=1}^{n_i}\varepsilon_{ij}^2<s)ds$.

The following result gives a large class of distributions satisfying Assumption C.

Proposition 1. Assume that $n_i\ge n_0=3$ and e_{ij} has a density $f_i(t)$ which is symmetric about zero and satisfies $\int|t|^{2+\delta}f_i(t)dt\le C$ and $f_i(t)\le C$ when $t\in[-a,\ a]$, where a and C are positive constants and independent of i. Then Assumption C holds.

Proof. We only need to show that $E(\sum_{j=1}^{n_0}e_{ij}^2)^{-(1+\delta)}<\infty$. Let $A=[-a,\ a]\times[-a,\ a]\times[-a,\ a]$. For $0\le\delta<1/2$, from (4.2)-(4.4),

$$\int_A(t_1^2+t_2^2+t_3^2)^{-(1+\delta)}f_i(t_1)f_i(t_2)f_i(t_3)dt_1dt_2dt_3\le C^3\int_A(t_1^2+t_2^2+t_3^2)^{-(1+\delta)}dt_1dt_2dt_3<\infty.$$

The result follows from

$$\int_{A^c}(t_1^2+t_2^2+t_3^2)^{-(1+\delta)}f_i(t_1)f_i(t_2)f_i(t_3)dt_1dt_2dt_3\le a^{-2(1+\delta)}<\infty.\ \square$$

ACKNOWLEDGEMENTS

I would like to thank the referee for helpful comments.

REFERENCES

[1] Carroll, R. J. (1982). Adapting for heteroscedasticity in linear models. *Ann. Statist. 10*, 1224-1233.

[2] Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann Statist. 13*, 342-368.

[3] Fuller, W. A. and Rao, J. N. K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. *Ann. Statist. 6*, 1149-1158.

[4] Gourieroux, C. and Monfort, A. (1981). Asymptotic properties of the maximum likelihood estimator in dichotomous logit models. *J. Econometrics 17*, 83-97.

[5] Hartley, H. O., Rao, J. N. K. and Kiefer G. (1969). Variance estimation with one unit per stratum. *J. Amer. Statist. Assco., 64*, 841-851.

[6] Horn, S. D., Horn, R. A. and Duncan, D. B. (1975). Estimating heteroscedastic variances in linear models. *J. Amer. Statist. Assoc.*, 70, 380-385.

[7] Rao, C. R. (1970). Estimation of heteroscedastic variances in linear models. *J. Amer.* Statist. Assoc., 65, 161-172.

[8] Rao, J. N. K. (1973). On the estimation of heteroscedastic variances. *Biometrics*, 29, 11-24.

[9] Shao, J. (1987). Estimating heteroscedastic variances in linear models I and II. Tech. Report #87-42, 43, Dept. of Statistics, Purdue University.