A Comparison of the Asymptotic Efficiency of Ordinary and Weighted Least Squares Estimators

by

Jun Shao
Purdue University

Technical Report #88-30

Department of Statistics
Purdue University

June, 1988

△

A COMPARISON OF THE ASYMPTOTIC EFFICIENCIES OF ORDINARY AND WEIGHTED LEAST SQUARES ESTIMATORS

Jun Shao
Department of Statistics, Purdue University
West Lafayette, IN 47907, U. S. A.

SUMMARY

We compare the asymptotic efficiency of the ordinary least squares estimator (OLSE) and the weighted least squares estimator (WLSE) in a heteroscedastic linear regression model with a large number of regressors but a small number of replicates at each regressor. The WLSE is constructed by estimating the error variances by the (within-group) average of squared residuals. It is shown that the OLSE is more efficient than the WLSE if the maximum number of replicates is not larger than two. Comparisons of the asymptotic efficiency of the WLSE and OLSE are given for the situation where there are three or more replicates at each regressor. A method of estimating the relative efficiency based on the observed data is also proposed and its performance is examined in a Monte Carlo study.

Key words: Weighted least squares, replicates, asymptotic efficiency, random variances.

1. INTRODUCTION

The linear regression is one of the most useful model in statistical applications. If the random errors have the same variance, the classical theory shows that the ordinary least squares estimator (OLSE) of the unknown vector of regression coefficients is the best linear estimator and asymptotically efficient. In many practical problems the error variances are unequal (heteroscedastic) and therefore the optimal properties of the OLSE are lost. If the variances are known, the best linear estimator is the weighted least squares estimator (WLSE) with the reciprocals of the variances as the weights. However, usually the error variances are unknown. A natural and frequently used approach is to obtain estimates of the error variances from the observed data and use the reciprocals of the variance estimates as the weights in the WLSE. Typically, one may face one of the following situations:

(A) The error variances vary smoothly with the regressors or the mean responses. In this case, consistent estimators of the error variances can be obtained and the WLSE is asymptotically more efficient than the OLSE (see Carroll, 1982).

(B) There is no deterministic relation between the error variances and regressors and there are some replicates at each regressor. Great effort has been expended in finding estimators of the error variances in this situation. See Hartley, Rao and Kiefer (1969), Rao (1970), Rao (1973), Horn, Horn and Duncan (1975), Fuller and Rao (1978) and Shao (1987). However, it is common in practice that the number of replicates is small for each regressor and therefore no consistent estimator of the error variance is available. If the weights in the WLSE are based on inconsistent variance estimators, there may be a cost due to not knowing the error variances, i.e., the WLSE may not be more efficient than the OLSE.

The purpose of this paper is to compare the efficiency of the WLSE and OLSE in situation (B). Most of the previous researches of this problem are limited to empirical studies. Simulation results (e.g., Jacquez, Mather and Crawford, 1968; Rao and Subrahmaniam, 1971; Fuller and Rao, 1978; Shao, 1987) showed that neither WLSE nor OLSE is always more efficient than the other estimator.

After an introduction of some necessary notations and a preliminary result in Section 2, we compare the asymptotic efficiency of the WLSE and the OLSE in Sections 3 and 4 under various situations. The WLSE we focus on is the one introduced by Fuller and Rao (1978). The overall conclusions are:

(i) The OLSE is more efficient than the WLSE if there are at most two replicates at each regressor;

(ii) When there are at least three replicates at each regressor, the WLSE is more efficient if the variation in the error variances is sufficiently large.

Therefore, to compare the efficiency, a study of the error variance pattern is necessary. Such a study usually requires some additional information about the error variances. We consider several cases in Sections 3 and 4 and establish some conditions under which the WLSE (or the OLSE) is more efficient. In Section 5, we discuss the estimation of the relative efficiency of the WLSE with respect to the OLSE based on the observed data. Some simulation results are also presented.

2. ASYMPTOTIC DISTRIBUTION OF THE WLSE

The following linear regression model will be studied in this paper:

$$y_{ij} = x_i'\beta + \sigma_i e_{ij}, \quad j=1,\dots,n_i, \ i=1,\dots,k, \ \sum_{i=1}^k n_i = n, \tag{2.1}$$

where β is a p-vector of unknown regression coefficients, x_i is the ith regressor (design) vector, x_i' is the transpose of x_i, y_{ij} is the jth response at the ith regressor, $\sigma_i e_{ij}$ is the random error and e_{ij} are independent and identically distributed (i.i.d.) with mean zero and variance one. It is assumed that the error variance σ_i^2 and the regressor x_i satisfy

$$\sigma_0 \le \sigma_i \le \sigma_\infty, \quad x_i'x_i \le c_\infty$$

for some positive constants σ_0, σ_∞ and c_∞. We consider the common practical situations where the number of replicates n_i is small, i.e., $1 \le n_0 \le n_i \le n_\infty < \infty$ for all i, while the total number of regressors k is large. Denote the response vector by

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

and the design matrix by

$$X = (x_1 \ldots x_1 \ldots \ldots x_k \ldots x_k)'_{n \times p}.$$

The design matrix is assumed to be of full rank. The OLSE of β is

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

We focus on the following WLSE (Fuller and Rao, 1978):

$$\hat{\beta}^w = (X'WX)^{-1}X'Wy, \qquad W = block\ diag.(w_1 I_{n_1} \ldots w_k I_{n_k}),$$

where I_{n_i} is the $n_i \times n_i$ identity matrix and w_i are the reciprocals of the error variance estimates

$$n_i^{-1}\sum_{j=1}^{n_i}(y_{ij} - x_i'\hat{\beta})^2.$$

This WLSE is more efficient than the WLSE with the variances σ_i^2 estimated by the customary estimators: the within-group sample variances (see Rao, 1973; Carroll and Cline, 1988). Fuller and Rao (1978) established the asymptotic distribution of $\hat{\beta}^w$ under the assumption that e_{ij} are normally distributed. Their result has been extended to the general situation by Shao (1988):

Proposition 1. Suppose that k^{-1}(the minimum eigenvalue of $X'X$) $\geq c_0$ for a positive constant c_0 and that the distribution of e_{ij} satisfies the moment conditions:

$$E(e_{12}/\sum_{j=1}^{n_i} e_{1j}^2) = 0, \quad E[e_{11}e_{12}/(\sum_{j=1}^{n_i} e_{1j}^2)^t] = 0, \ \ t=1,2,$$

$$E\,|e_{11}|^{2+\delta} \leq c \quad \text{and} \quad E(\sum_{j=1}^{n_0} e_{1j}^2)^{-(1+\delta)} \leq c,$$

where c and $\delta<1/2$ are positive constants. Then the WLSE $\hat{\beta}^w$ is asymptotically (as $k \to \infty$) normal with mean β and asymptotic covariance matrix

$$V_k^w = A_k^{-1} + 4A_k^{-1}B_kA_k^{-1} + 4A_k^{-1}B_kV_kB_kA_k^{-1}, \tag{2.2}$$

where $A_k = \sum_{i=1}^k \sigma_i^{-2} n_i^2 \tau(n_i) x_i x_i'$, $B_k = \sum_{i=1}^k \sigma_i^{-2} n_i \tau(n_i) x_i x_i'$, $\tau(n_i) = E(\sum_{j=1}^{n_i} e_{ij}^2)^{-1}$, and

$$V_k = (X'X)^{-1}\sum_{i=1}^k \sigma_i^2 n_i x_i x_i'(X'X)^{-1} \tag{2.3}$$

is the covariance matrix of the OLSE $\hat{\beta}$.

Note that the first two moment conditions in Proposition 1 are satisfied if the distributions of e_{ij} are symmetric. Examples of distributions satisfying the conditions in Proposition 1, which include most of the error distributions encountered in practice if $n_0 \geq 3$, can be found in Shao (1988).

The efficiency of the WLSE and OLSE are compared in Sections 3 and 4 by comparing the asymptotic covariance matrices given by (2.2) and (2.3).

3. EFFICIENCY COMPARISON: THE BALANCED CASE

When there are equal number of replicates at each regressor, i.e., $n_i = m$ for all i, the model (2.1) is said to be balanced. In this section we focus on this important special case. When the model is balanced, V_k^w reduces to

$$[m\tau(m)]^{-1}(1+4m^{-1})\tilde{V}_k + 4m^{-2}V_k, \quad \tilde{V}_k = m^{-1}(\textstyle\sum_{i=1}^k \sigma_i^{-2} x_i x_i')^{-1}. \tag{3.1}$$

If m is large, the WLSE is more efficient than the OLSE since $\tilde{V}_k \leq V_k$ and $\lim_{m\to\infty}[m\tau(m)]=1$ (Lemma 1 in the Appendix), where V_k is the covariance matrix of the OLSE. However, in practice m is seldomly large. If m is too small ($m = 1$ or 2), there is no gain in using the WLSE. This is because under the conditions of Proposition 1,

$$V_k^w - V_k = [m\tau(m)]^{-1}(1+4m^{-1})\tilde{V}_k + (4m^{-2}-1)V_k > 0$$

if $m \leq 2$. Hence the WLSE with $m = 1$ or 2 is not recommended. We now consider the case of $m \geq 3$.

3.1. General Situation

From (2.3) and (3.1),

$$V_k^w - V_k = [m\tau(m)]^{-1}(1+4m^{-1})[\tilde{V}_k - \alpha(m)V_k], \tag{3.2}$$

where $\alpha(m) = (m^2-4)\tau(m)/(m+4)$. Note that $\tilde{V}_k$ is the covariance matrix of

$$\tilde{\beta} = (X'\tilde{W}X)^{-1}X'\tilde{W}y, \qquad \tilde{W} = block\ diag.(\,\sigma_1^{-2}I_{n_1} \ldots \sigma_k^{-2}I_{n_k}\,),$$

which is the WLSE when σ_i^2 are known. (3.2) indicates that if $\tilde{\beta}$ is $\alpha(m)$ times less variable than the OLSE $\hat{\beta}$, then the WLSE $\hat{\beta}^w$ is more efficient than $\hat{\beta}$.

The value of $\alpha(m)$ depends on the distribution of e_{ij}. For several different distributions, the values of $\alpha(m)$ are given by Table 1 for m ranging from 3 to 8. A lower bound for $\alpha(m)$ is $(m^2-4)/(m^2+4m)$ since $m\tau(m) \geq 1$ (Lemma 1 in the Appendix). The values of this lower bound are also given in Table 1.

Table 1: Values of $\alpha(m)$

m	3	4	5	6	7	8
normal error	.714	.750	.778	.800	.818	.833
uniform error	.445	.536	.595	.643	.678	.709
R-distributed error *	.357	.500	.583	.640	.682	.714
lower bound	.238	.375	.467	.533	.584	.625

*The distribution of e_{ij} has a density $|t|\exp(-t^2)$.

As an example, suppose that $\tilde{\beta}$ is half times less variable than the OLSE. Then for normally distributed errors, the WLSE is more efficient than the OLSE as long as $m \geq 3$. For uniformly distributed errors, the WLSE is more efficient if $m \geq 4$. From the lower bounds in Table 1, if $m \geq 6$, then the WLSE is more efficient regardless of the error distributions.

Thus, comparing the efficiency of the WLSE and OLSE is equivalent to comparing $\tilde{V}_k$ and V_k. However, since $\tilde{V}_k$ and V_k depend on the regressors x_i and the variances σ_i^2 which are unknown, this can not be done without any further assumption. In the rest of this section, we consider the case where additional information about the $\bar{\sigma}_i$'s is available. A study of estimating $\tilde{V}_k$ and V_k based on the observed data is given in Section 5.

3.2. Estimation of Common Mean

A special case of (2.1) is

$$y_{ij} = \mu + \sigma_i e_{ij}, \quad j=1,\ldots,m,\ i=1,\ldots,k. \tag{3.3}$$

Although this is the simplest case, the following analysis is heuristic and the results will be used in more complex models. Under model (3.3), $\tilde{V}_k = m^{-1}(\sum_{i=1}^k \sigma_i^{-2})^{-1}$, $V_k = m^{-1}k^{-2}\sum_{i=1}^k \sigma_i^2$, and $\tilde{V}_k < \alpha(m)V_k$ iff

$$[\alpha(m)]^{-1} < (k^{-1}\textstyle\sum_{i=1}^k \sigma_i^{-2})(k^{-1}\sum_{i=1}^k \sigma_i^2) = k^{-1}+k^{-2}\sum_{i<j}(\rho_{ij}+\rho_{ij}^{-1}),$$

where $\rho_{ij}=\sigma_i^2/\sigma_j^2$. Note that $\Delta_k = k^{-1}+k^{-2}\sum_{i<j}(\rho_{ij}+\rho_{ij}^{-1})$ is a measure of variability of the σ_i's. $\Delta_k \geq 1$ and $\Delta_k=1$ iff $\sigma_i^2=\sigma_j^2$ for all i and j. The WLSE is more efficient than the OLSE iff $1 < \Delta_k\alpha(m)$, i.e., the σ_i's are substantially different. This coincides with the simulation results in Jacquez et al. (1968) and Fuller and Rao (1978). The asymptotic relative efficiency (ARE) of the WLSE (with respect to the OLSE) in this case is

$$\gamma_k = V_k^w/V_k = 1+(1-4m^{-2})[\alpha^{-1}(m)\Delta_k^{-1}-1].$$

However, Δ_k is unknown since the σ_i's are. Further information about σ_i's are required for comparing Δ_k and $\alpha(m)$. We discuss the following cases for illustration.

Situation I. σ_i is (nearly) constant for most i but is large for one or a few cases, which usually corresponds to response outliers.

For example, $\sigma_i^2=\sigma^2$ for $i \leq k-1$ but $\sigma_k^2 \neq \sigma^2$. Then

$$\Delta_k = (1-k^{-1})^2+k^{-2}+k^{-1}(1-k^{-1})(\sigma^2\sigma_k^{-2}+\sigma^{-2}\sigma_k^2) \to 1$$

since $(\sigma_k^{-2}+\sigma_k^2)/k \to 0$. Therefore the OLSE is more efficient for large k.

Thus, in this situation the OLSE is recommended.

Situation II. Consider the Cochran and Carroll variance model (Fuller and Rao, 1978, Section 3). In this model one-third of the k variances are equal to c, one-third variances are equal to 1 and the remaining one-third variances are equal to c^{-1}. Therefore,

$$\Delta_k^{-1} = 9(c+1+c^{-1})^{-2}.$$

The values of Δ_k^{-1} for c ranging from .2 to .6 are given in Table 2.

From Tables 1 and 2, for normal errors, the WLSE is more efficient than the OLSE if m=3, $c\leq.45$ or m=4, $c\leq.5$. For uniform errors, the WLSE is more efficient if m=3, $c\leq.3$ or m=4, $c\leq.35$. Regardless of the error distributions, the WLSE is more efficient if m=3, $c\leq.2$ or m=4, $c\leq.25$.

Table 2: Values of Δ_k^{-1}

c	.2	.25	.3	.35	.4	.45	.5	.55	.6
Δ_k^{-1}	.234	.327	.419	.508	.592	.667	.735	.793	.843

Situation III. The σ_i's are i.i.d. random variables defined on $(\sigma_0, \sigma_\infty)$. Let $a=E\sigma_i^2$ and $b=E\sigma_i^{-2}$. Then $E\Delta_k = k^{-1}+k^{-2}\sum_{i<j}2ab = ab+k^{-1}(1-ab)$. Hence as $k\to\infty$,

$$\Delta_k \to ab \quad a.s.$$

Note that ab is a measure of the variability of the σ_i's and $ab\geq1$. If $\alpha(m)>(ab)^{-1}$, then the WLSE is more efficient than the OLSE for large k. The ARE of the WLSE is

$$\gamma = 1+(1-4m^{-2})[\alpha^{-1}(m)(ab)^{-1}-1].$$

For example, suppose that σ_i is uniformly distributed on the interval $(\sigma_0, \sigma_\infty)$. Then $a=(\sigma_\infty^3-\sigma_0^3)/3(\sigma_\infty-\sigma_0)$, $b=1/\sigma_\infty\sigma_0$ and $ab = (\rho^2+\rho+1)/3\rho$, where $\rho=\sigma_\infty/\sigma_0$. Some values of $(ab)^{-1}$ for different ratios ρ are given in Table 3.

Table 3: Values of $(ab)^{-1}$

ρ	1.5	2	2.5	3	3.5	4	4.5
$(ab)^{-1}$	.947	.857	.769	.693	.627	.571	.524
ρ	5	5.5	6	7	8.5	10	11.5
$(ab)^{-1}$	.484	.449	.419	.368	.312	.270	.238

Note that the WLSE is more efficient if $\alpha(m)>(ab)^{-1}$. As an example, if ρ=2, the OLSE is more efficient than the WLSE; if ρ=3, the WLSE is more efficient for normal errors; if ρ=4, the WLSE is more efficient for uniform errors and $m\geq5$; if ρ=7 (e.g., σ_∞=.7 and σ_0=.1), the WLSE is more efficient for $m \geq 4$ (regardless of the error distributions).

In conclusion, one may determine the relative efficiency between the WLSE and OLSE through a careful study of the variation of the σ_i's and the error distribution by making use of the available information. Another approach is to estimate Δ_k and $\alpha(m)$ from the observed data (see Section 5).

3.3. Random Variances

If the σ_i's are i.i.d. random variables, the results in Situation III of Section 3.2 can be extended to the general model (2.1). Let $a = E\sigma_i^2$ and $b = E\sigma_i^{-2}$. From the strong law of large numbers, as $k \to \infty$,

$$kV_k - amk(X'X)^{-1} = mk(X'X)^{-1}\sum_{i=1}^k \sigma_i^2 x_i x_i'(X'X)^{-1} - amk(X'X)^{-1} \to 0 \quad a.s.$$

and

$$k\tilde{V}_k - b^{-1}mk(X'X)^{-1} = mk(\sum_{i=1}^k \sigma_i^{-2} x_i x_i')^{-1} - b^{-1}mk(X'X)^{-1} \to 0 \quad a.s.$$

Hence for large k, the WLSE is more efficient than the OLSE iff $\alpha(m) > (ab)^{-1}$. Thus, the results in Section 3.2 hold in this general case.

3.4. Random Regressors

Assume that x_i's are i.i.d. random vectors (independent of σ_i's) and that $\Sigma = Ex_i x_i'$ is nonsingular. Then

$$k(X'X)^{-1} = m^{-1}k(\sum_{i=1}^k x_i x_i')^{-1} \to m^{-1}\Sigma^{-1} \quad a.s.,$$

$$k^{-1}\sum_{i=1}^k \sigma_i^2 x_i x_i' - (k^{-1}\sum_{i=1}^k \sigma_i^2)\Sigma \to 0 \quad a.s.,$$

and therefore

$$kV_k - m^{-1}(k^{-1}\sum_{i=1}^k \sigma_i^2)\Sigma^{-1} \to 0 \quad a.s.$$

Similarly,

$$k\tilde{V}_k - m^{-1}(k^{-1}\sum_{i=1}^k \sigma_i^{-2})\Sigma^{-1} \to 0 \quad a.s.$$

Thus, for large k, the WLSE is more efficient iff $\Delta_k = (k^{-1}\sum_{i=1}^k \sigma_i^{-2})(k^{-1}\sum_{i=1}^k \sigma_i^2) > [\alpha(m)]^{-1}$.

Hence all the results in Section 3.2 hold in this case.

3.5. Related Regressors and Variances

Carroll (1982) studies the case that given x_i, σ_i^2 is a smooth deterministic function of x_i. Here, we consider the situation where x_i and σ_i^2 are related and are both random. For example, given x_i, σ_i^2 has a distribution with mean $q(x_i'x_i)$, where q is a known function. Assume (σ_i^2, x_i) are i.i.d. and $\Sigma = Ex_ix_i'$ is nonsingular. Then

$$k(X'X)^{-1} \to m^{-1}\Sigma^{-1} \quad a.s.$$

$$k^{-1}\sum_{i=1}^k \sigma_i^2 x_i x_i' \to E(\sigma_i^2 x_i x_i') \quad a.s.$$

and

$$k^{-1}\sum_{i=1}^k \sigma_i^{-2} x_i x_i \to E(\sigma_i^{-2} x_i x_i') \quad a.s.$$

Therefore, the WLSE is more efficient than the OLSE iff

$$[E(\sigma_i^{-2} x_i x_i')]^{-1} < \alpha(m)\Sigma^{-1}E(\sigma_i^2 x_i x_i)\Sigma^{-1},$$

or equivalently,

$$\alpha(m)E(\sigma_i^{-2} x_i x_i') > \Sigma[E(\sigma_i^2 x_i x_i)]^{-1}\Sigma. \qquad (3.4)$$

Note that in this case it is possible that some of the components of the WLSE are more efficient than the corresponding OLSE while the other components of the WLSE are not. In general, the evaluation of the matrices in (3.4) is not simple and involves the distribution of (σ_i^2, x_i). We study the following example for illustration.

Suppose that the model is a simple linear regression:

$$y_{ij} = \beta_1 + \beta_2 t_i + \sigma_i e_{ij}, \quad j=1,\dots,m,\ i=1,\dots,k,$$

where t_i are uniformly distributed on $[a,b]$. For simplicity, let $a=.5$ and $b=5$. Assume that given t_i, σ_i^2 is uniformly distributed on $[st_i, (2-s)t_i]$, where $0<s<1$ is a known constant. Note that $x_i=(1, t_i)'$. By a straightforward calculation (see the Appendix),

$$\Sigma[E(\sigma_i^2 x_i x_i')]^{-1}\Sigma = \begin{pmatrix} .466 & .991 \\ .991 & 2.718 \end{pmatrix} \tag{3.5}$$

and

$$\alpha(m)E(\sigma_i^{-2} x_i x_i') = \alpha(m)h(s)\begin{pmatrix} .512 & 1 \\ 1 & 2.75 \end{pmatrix}, \tag{3.6}$$

where $h(s)=[2(1-s)]^{-1}\log[(2-s)/s]$. Consider the normal distribution case. If s=.5, $.512h(s)$=.562 and $2.75h(s)$=3.021. From (3.5), (3.6) and Table 1, the OLSE is more efficient than the WLSE. If s=.25, $.512h(s)$=.664 and $2.75h(s)$=3.567. Then the WLSE of β_1 is more efficient than the OLSE iff $m\geq 3$ while the WLSE of β_2 is more efficient iff $m\geq 5$. If s=.2, then $2.75h(s)$=3.776 and the WLSE of β_2 is more efficient iff $m\geq 4$.

4. EFFICIENCY COMPARISON: THE UNBALANCED CASE

The analysis in the general unbalanced model ($n_i \neq n_j$ for some i, j) is more complicated although the ideas are similar to those in the balanced model. We consider three cases.

4.1. Estimation of Common Mean

For the simple model (3.3),

$$V_k^w = n^{-1}(\xi_k^{-1}+4\xi_k^{-2}\eta_k+4\xi_k^{-2}\eta_k^2\zeta_k),$$

where

$$\xi_k = n^{-1}\textstyle\sum_{i=1}^k \sigma_i^{-2} n_i^2 \tau(n_i), \quad \eta_k = n^{-1}\textstyle\sum_{i=1}^k \sigma_i^{-2} n_i \tau(n_i) \quad \text{and} \quad \zeta_k = n^{-1}\textstyle\sum_{i=1}^k \sigma_i^2 n_i. \tag{4.1}$$

Then the WLSE is more efficient than the OLSE iff

$$g_k = \xi_k^{-1}+4\xi_k^{-2}\eta_k+4\xi_k^{-2}\eta_k^2\zeta_k-\zeta_k > 0. \tag{4.2}$$

The ARE of the WLSE with respect to the OLSE is $\gamma_k = g_k\zeta_k^{-1}+1$.

Note that $n_\infty^{-1} \leq \xi_k^{-1}\eta_k \leq n_0^{-1}$, where n_0 and n_∞ are the minimum and maximum of the number of replicates, respectively. Then

$$(1+4n_\infty^{-1})\xi_k^{-1}-(1-4n_\infty^{-2})\zeta_k \leq g_k \leq (1+4n_0^{-1})\xi_k^{-1}-(1-4n_0^{-2})\zeta_k. \tag{4.3}$$

An immediate consequence from (4.3) is that if the maximum of the number of replicates is not larger than two, then the OLSE is more efficient than the WLSE.

Consider the case of $n_0 \geq 3$. Denote $(n^{-1}\sum_{i=1}^k \sigma_i^{-2} n_i)(n^{-1}\sum_{i=1}^k \sigma_i^2 n_i)$ by Δ_k and $(n_0^2-4)\tau(n_\infty)/(n_0+4)$ by $\alpha(n_0)$. From (4.3) and $n_0\tau(n_\infty)(n^{-1}\sum_{i=1}^k \sigma_i^{-2} n_i) \leq \xi_k$, we have

$$V_k^w - V_k < 0 \quad \text{if } \Delta_k \alpha(n_0) > 1.$$

Hence $\Delta_k \alpha(n_0) > 1$ is a sufficient condition under which the WLSE is more efficient than the OLSE. Similar results to those in Section 3.2 (Situations I-III) can then be established.

4.2. Random Variances

Under model (2.1) with i.i.d. σ_i's (independent of x_i), we have

$$n^{-1}\sum_{i=1}^k \sigma_i^2 n_i x_i x_i' - an^{-1}(X'X) \to 0 \quad a.s.,$$

$$n^{-1}\sum_{i=1}^k \sigma_i^{-2} n_i^2 \tau(n_i) x_i x_i' - bn^{-1}\sum_{i=1}^k n_i^2 \tau(n_i) x_i x_i' \to 0 \quad a.s.,$$

$$n^{-1}\sum_{i=1}^k \sigma_i^{-2} n_i \tau(n_i) x_i x_i' - bn^{-1}\sum_{i=1}^k n_i \tau(n_i) x_i x_i' \to 0 \quad a.s.,$$

$$nV_k - an(X'X)^{-1} \to 0 \quad a.s.,$$

where $a = E\sigma_i^2$ and $b = E\sigma_i^{-2}$. Let $G = \sum_{i=1}^k n_i^2 \tau(n_i) x_i x_i'$, $H = \sum_{i=1}^k n_i \tau(n_i) x_i x_i'$ and $F = b^{-1}(G^{-1} + 4G^{-1}HG^{-1}) + 4aG^{-1}H(X'X)^{-1}HG^{-1}$. Then $nV_k^w - nF \to 0$ *a.s.* Hence the WLSE is more efficient than the OLSE iff

$$F - a(X'X)^{-1} < 0. \tag{4.4}$$

To check (4.4), some information about a, b and $\tau(n_i)$ is required. Since $(X'X)^{-1} \leq \tau(n_0)H^{-1}$, $H \leq n_0^{-1}G$ and $G^{-1} \leq [n_0\tau(n_0)]^{-1}(X'X)^{-1}$, a sufficient condition for (4.4) is

$$ab[1 - 4n_0^{-2}\tau(n_0)/\tau(n_\infty)] > (1 + 4n_0^{-1})/[n_0\tau(n_\infty)].$$

4.3. Random Regressors

Let x_i's be i.i.d. random vectors and independent of σ_i's. Using the notations in (2.2), (2.3) and (4.1), we have

$$n^{-1}A_k - \xi_k\Sigma \to 0 \quad a.s.,$$

$$n^{-1}B_k - \eta_k\Sigma \to 0 \quad a.s.,$$

$$nV_k - \zeta_k\Sigma^{-1} \to 0 \quad a.s.,$$

and

$$nV_k^w - (\xi_k^{-1}+4\xi_k^{-2}+4\xi_k^{-2}\eta_k^2\zeta_k)\Sigma^{-1} \to 0 \quad a.s.,$$

where $\Sigma = Ex_ix_i'$ is assumed to be nonsingular. Therefore, the WLSE is more efficient iff

$$\xi_k^{-1}+4\xi_k^{-2}+4\xi_k^{-2}\eta_k^2\zeta_k - \zeta_k > 0,$$

which is the same as (4.2). Thus, the rest of the discussion is the same as that in Section 4.1.

5. ESTIMATING THE RELATIVE EFFICIENCY

When there is little information about the error variances, an alternative approach is to estimate the ARE of the WLSE with respect to the OLSE based on the data.

5.1. Estimators of the ARE

Let $w_i = n_i[\sum_{j=1}^{n_i}(y_{ij}-x_i'\hat{\beta})^2]^{-1}$, $\hat{V}_k = (X'X)^{-1}\sum_{i=1}^k w_i^{-1}n_ix_ix_i'(X'X)^{-1}$, $\hat{A}_k = \sum_{i=1}^k w_in_ix_ix_i'$, $\hat{B}_k = \sum_{i=1}^k w_ix_ix_i'$ and $\hat{V}_k^w = \hat{A}_k^{-1}+4\hat{A}_k^{-1}\hat{B}_k\hat{A}_k^{-1}+4\hat{A}_k^{-1}\hat{B}_k\hat{V}_k\hat{B}_k\hat{A}_k^{-1}$. From Shao (1988, Theorem 3), under the conditions of Proposition 1,

$$k(\hat{V}_k^w - V_k^w) \to 0 \quad \text{and} \quad k(\hat{V}_k - V_k) \to 0 \text{ in probability.}$$

Thus, a consistent estimator of the ARE of the WLSE $t'\hat{\beta}^w$ with respect to the OLSE $t'\hat{\beta}$, where t is a known p-vector, is

$$\hat{\gamma} = t'\hat{V}_k^w t / t'\hat{V}_k t.$$

In many situations (see Sections 3.2-3.4, 4.1-4.3), the ARE of the WLSE with respect to the OLSE is $\gamma_k = g_k\zeta_k^{-1}+1$ (see (4.1) and (4.2)). We now concentrate on the estimation of γ_k in the balanced model ($n_i \equiv m$). Since $\gamma_k = 1+(1-4m^{-2})[\alpha^{-1}(m)\Delta_k^{-1}-1]$ (see Section 3.2), its consistent estimator is

$$\hat{\gamma}_k = (1+4m^{-1})(\hat{a}\hat{b})^{-1}+4m^{-2},$$

where $\hat{a}=(k-1)^{-1}\sum_{i=1}^{k} w_i^{-1}$ and $\hat{b}=(k-p)^{-1}\sum_{i=1}^{k} w_i$.

Note that the WLSE is more efficient than the OLSE iff $\gamma_k<1$. Hence one may make a decision of using the WLSE (the OLSE) when $\hat{\gamma}_k<1$ ($\hat{\gamma}_k\geq 1$). Precision measures of $\hat{\gamma}_k$ are

$$p_I = P(\,\hat{\gamma}_k<1 \mid \gamma_k<1\,) \quad \text{and} \quad p_{II} = P(\,\hat{\gamma}_k\geq 1 \mid \gamma_k\geq 1\,), \tag{5.1}$$

which are the probabilities that one makes a correct decision. The performance of $\hat{\gamma}_k$ is examined in the following Monte Carlo study.

5.2. A Monte Carlo Study

We study the following model:

$$y_{ij} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \sigma_i e_{ij}, \quad j=1,2,3,\ i=1,\dots,24.$$

The e_{ij}'s are normally distributed and $\sigma_i=(h-l)u_i+l$, where u_i's are independent samples from the uniform distribution on [0,1]. Thus, σ_i's are independent samples from the uniform distribution on $[l,h]$. The values of x_{1i}, x_{2i} and u_i are listed in Table 4.

Under this model, the Monte Carlo approximations of p_I and p_{II} defined in (5.1) are given in Table 5 for some different values of l and h. The true values of γ_k and the bias and root mean squared error (rmse) of $\hat{\gamma}_k$ as an estimator of γ_k are also included. All the results are based on 5000 simulations on a VAX 11/780 at the Purdue University.

The following are some conclusions drawn from the simulation results:

(a) The precision p_I (or p_{II}) is small if γ_k is close to one. This is reasonable since when γ_k is close to one, the difference between the WLSE and OLSE is not appreciable. If the relative gain in using the WLSE (or the OLSE) is beyond 25%, it is quite safe to use $\hat{\gamma}_k$ for choosing an efficient estimator of β.

(b) It is clear that $\hat{\gamma}_k$ is upward biased as an estimator of γ_k. Since the OLSE is computationally simpler than the WLSE, it may be preferred if the WLSE is better but the relative gain is small (e.g., less than 10%). For this reason, an upward biased estimator of γ_k, which corresponds to a conservative estimator of the relative gain in using the WLSE, is acceptable.

Table 4. Values of x_{1i}, x_{2i} and u_i

i	1	2	3	4	5	6	7	8
x_{1i}	10	36	75	16	35	50	48	21
x_{2i}	0	8	5	4	5	8	4	1
u_i	.232	.965	.229	.089	.771	.600	.038	.863
i	9	10	11	12	13	14	15	16
x_{1i}	66	30	55	49	23	66	10	70
x_{2i}	10	3	9	6	3	7	2	10
u_i	.646	.191	.880	.438	.715	.715	.052	.952
i	17	18	19	20	21	22	23	24
x_{1i}	44	42	69	58	26	14	62	63
x_{2i}	0	4	12	10	2	4	9	3
u_i	.404	.951	.321	.265	.177	.027	.187	.155

Table 5. The precision, bias and rmse of $\hat{\gamma}_k$

l	h	γ_k	precision *	bias	rmse
1.4	1.6	1.216	.825	.036	.272
1.2	1.8	1.171	.788	.041	.266
1.0	2.0	1.086	.707	.049	.254
0.8	2.4	.939	.511	.062	.228
0.7	2.8	.838	.695	.068	.205
0.6	3.0	.769	.818	.071	.188
0.5	3.5	.684	.938	.075	.166
0.4	3.6	.634	.976	.076	.153

* The precision equals p_I (p_{II}) when $\gamma_k<1$ ($\gamma_k\geq1$).

APPENDIX

Lemma 1. Assume the conditions in Proposition 1. Then for any $m \geq n_0$, $m\tau(m) \geq 1$ and

$$\lim_{m\to\infty} m\tau(m)=1.$$

Proof. The first assertion follows from $m\tau(m) \geq (m^{-1}\sum_{j=1}^{m} Ee_{ij}^2)^{-1}=1$. Let g be the integer part of m/n_0 and $u_h=\sum_{j=1+(h-1)n_0}^{hn_0} e_{ij}^2$. Then $gn_0 \leq m \leq (g+1)n_0$ and

$$m(\sum_{j=1}^{m} e_{ij}^2)^{-1} \leq n_0(g+1)(\sum_{j=1}^{gn_0} e_{ij}^2)^{-1} \leq n_0(g+1)g^{-2}\sum_{h=1}^{g} u_h^{-1}.$$

By the strong law of large numbers, $m(\sum_{j=1}^{m} e_{ij}^2)^{-1} \to 1$ and $g^{-1}\sum_{h=1}^{g} u_h^{-1} \to Eu_h^{-1}=\tau(n_0)$ *a.s.* as $m\to\infty$. Then the second assertion follows from the dominated convergence theorem (see Royden, 1968, p.232). □

Proof of (3.5) *and* (3.6). Note that $Et_i=2.75$ and $Et_i^2=9.25$. Since the conditional distribution of σ_i^2 (given t_i) is uniform on $[st_i, (2-s)t_i]$, we have

$$E\sigma_i^2 = E[E(\sigma_i^2 \mid t_i)] = Et_i = 2.75,$$

$$E\sigma_i^2 t_i = E[E(\sigma_i^2 t_i \mid t_i)] = Et_i^2 = 9.25$$

and

$$E\sigma_i^2 t_i^2 = E[E(\sigma_i^2 t_i^2 \mid t_i)] = Et_i^3 = 34.719.$$

Note that $E(\sigma_i^{-2} \mid t_i)=[2(1-s)t_i]^{-1}\log[(2-s)/s]$. Then

$$E\sigma_i^{-2} = E[E(\sigma_i^{-2} \mid t_i)] = h(s)Et_i^{-1} = .512h(s),$$

$$E\sigma_i^{-2} t_i = E[E(\sigma_i^{-2} t_i \mid t_i)] = h(s)$$

and

$$E\sigma_i^{-2} t_i^2 = E[E(\sigma_i^{-2} t_i^2 \mid t_i)] = h(s)Et_i = 2.75h(s). \quad \square$$

REFERENCES

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

Carroll, R. J. and Cline, D. B. H. (1988). An asymptotic theory for weighted least-squares with weights estimated by replication. *Biometrika*, 75, 35-43.

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.

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.

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

Jacquez, J. A., Mather, F. J. and Crawford, C. R. (1968). Linear regression with nonconstant, unknown error variances: sampling experiments with least squares, weighted least squares and maximum likelihood estimators. *Biometrics,* 24, 607-626.

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

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

Rao, J. N. K. and Subrahmaniam, K. (1971). Combining independent estimators and estimation in linear regression with unequal variances. *Biometrics*, 27, 971-990.

Royden, H. L. (1968). *Real Analysis*, second edition. Macmillan, New York.

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

Shao, J. (1988). Asymptotic distribution of the weighted least squares estimator. Tech. Report #88-10, Dept. of Statistics, Purdue University.