UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

DTIC FILE COPY

(2)

JMENTATION PAGE

1a. REPORT SE	1b. RESTRICTIVE MARKINGS
2a. SECURITY (**AD-A198 000**	3. DISTRIBUTION/AVAILABILITY OF REPORT: Distribution unlimited
2b. DECLASSIF	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): Technical Report #1	5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 88-0844
6a. NAME OF PERFORMING ORGANIZATION: Texas A&M University; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code): College Station, TX 77853	7b. ADDRESS (City, State and ZIP Code): Same as 8c
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR; 8b. OFFICE SYMBOL (If applicable): NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-85-0144
8c. ADDRESS (City, State and ZIP Code): Bolling Air Force Base, Washington, DC 20332 (Bld 410)	10. SOURCE OF FUNDING NOS. — PROGRAM ELEMENT NO.: 61102F; PROJECT NO.: 2304; TASK NO.: A6; WORK UNIT NO.:

DTIC ELECTE AUG 29 1988 S D

11. TITLE (Include Security Classification): AN ASYMPTOTIC THEORY FOR WEIGHTED LEAST SQUARES WITH WEIGHTS ESTIMATED BY REPLICATION

12. PERSONAL AUTHOR(S): Carroll, Raymond J. and Cline, D.B.H.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
~~technical~~ Journal	FROM 8/87 TO 8/88		16

16. SUPPLEMENTARY NOTATION: This document (The authors)

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	generalized least squares, heteroscedasticity, regression, replication

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

~~We~~ consider a heteroscedastic linear regression model with replication. To estimate the variances, one can use the sample variances or the sample average squared errors from a regression fit. ~~We~~ study the large sample properties of these weighted least squares estimates with estimated weights when the number of replicates is small. The estimates are generally inconsistent for asymmetrically distributed data. If sample variances are used based on m replicates, the weighted least squares estimates are inconsistent for m=2 replicates even when the data are normally distributed. With between 3 and 5 replicates, the rates of convergence are slower than the usual square root of N. With m ≥ 6 (7 or =) replicates, the effect of estimating the weights is to increase variances by (m-5)/(m-3), relative to weighted least squares estimates with known weights. (KP)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ OTIC USERS ☐	UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Major Brian Woodruff	(202) 767-5026	NM

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE. — UNCLASSIFIED — SECURITY CLASSIFICATION OF THIS PAGE

88 8 25 088 25 09

An Asymptotic Theory for Weighted Least Squares with Weights Estimated by Replication

Technical Report #1

Raymond J. Carroll and Daren B. H. Cline

SUMMARY

We consider a heteroscedastic linear regression model with replication. To estimate the variances, one can use the sample variances or the sample average squared errors from a regression fit. We study the large sample properties of these weighted least squares estimates with estimated weights when the number of replicates is small. The estimates are generally inconsistent for asymmetrically distributed data. If sample variances are used based on m replicates, the weighted least squares estimates are inconsistent for $m = 2$ replicates even when the data are normally distributed. With between 3 and 5 replicates, the rates of convergence are slower than the usual square root of N. With $m \geq 6$ replicates, the effect of estimating the weights is to increase variances by $(m-5)/(m-3)$, relative to weighted least squares estimates with known weights.

Some Key Words : Generalized Least Squares, Heteroscedasticity, Regression, Replication.

Section 1 : Introduction

Consider a heteroscedastic linear regression model with replication:

$$y_{ij} = \mathbf{x}_i^T \beta + \sigma_i \epsilon_{ij} \quad (i=1,\ldots,N),\ (j=1,\ldots,m). \tag{1.1}$$

In model (1.1), β is a vector with p-components, the ϵ_{ij} are independent and identically distributed mean zero random variables with variance one. The heteroscedasticity in the model is governed by the unknown σ_i. We have taken the number of replicates at each $\mathbf{x}_i$ to be the constant m primarily as a matter of convenience. In practice, it is fairly common that the number of design vectors N is large while the number of replicates m is small. Our intention is to construct an asymptotic theory in this situation for weighted least squares estimates with estimated weights.

As a benchmark, let $\hat{\beta}_{WLS}$ be the weighted least squares estimate with weights $1/\sigma_i^2$. Of course, since the σ_i are unknown this estimate cannot be calculated from data. If m is fixed and

$$S_{WLS} = \operatorname*{plim}_{N\to\infty} N^{-1} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^T / \sigma_i^2,$$

then

$$(Nm)^{1/2}(\hat{\beta}_{WLS} - \beta) \Rightarrow \text{Normal}(0,\ S_{WLS}^{-1}) \ . \tag{1.2}$$

One common method for estimating weights uses the inverses of the sample variances,

$$\hat{\sigma}_{i1}^2 = s_i^2 = (m-1)^{-1} \sum_{j=1}^{m} (y_{ij} - \bar{y}_i)^2 \ . \tag{1.3}$$

The resulting weighted least squares estimator will be denoted by $\hat{\beta}_{SV}$.

This method is particularly convenient because it involves sending only the estimated weights to a computer program with a weighting option. The obvious question is whether $\hat{\beta}_{SV}$ is any good, and whether the inferences made by the computer program have any reliability. In Sections 3 and 4, we answer both questions in the negative, at least for normally distributed data with

less than 10 replicates at each x. In many applied fields this is already folklore (Garden, et al., 1980). Yates and Cochran (1938) also have a nice discussion of the problems with using the sample variances to estimate the weights.

More precisely, for normally distributed data we are able to describe the asymptotic distribution of $\hat{\beta}_{SV}$ for every m. For $m \geq 6$, this is an easy moment calculation and we show that $\hat{\beta}_{SV}$ is more variable than $\hat{\beta}_{WLS}$ by a factor $(m-3)/(m-5)$. The same result was obtained by Cochran (1937) for the weighted mean. Not only is $\hat{\beta}_{SV}$ inefficient, but if one uses an ordinary weighted regression package to compute $\hat{\beta}_{SV}$, the standard errors from the package will be too small by a factor exceeding 20% unless $m \geq 10$. For example, if one uses $m = 6$ replicates, the efficiency with respect to weighted least squares with known weights is only 1/3, and all estimated standard errors should be multiplied by $\sqrt{3} = 1.732$. For $m \leq 5$, we use the theory of stable laws and Cline (1986a,b) to describe the asymptotic distributions. Perhaps the most interesting result here is that if only duplicates ($m = 2$) are used, weighted least squares with estimated weights is not even consistent. The results are outlined in Table 1.

[TABLE 1 AFTER THIS POINT]

A second method for estimating weights is to use the linear structure of the means. Write $\hat{\beta}_L$ for the unweighted least squares estimate and define the average squared error estimate by

$$\hat{\sigma}^2_{i2} = \hat{\sigma}^2_{i2}(\hat{\beta}_L) = m^{-1} \sum_{j=1}^{m} (y_{ij} - \mathbf{x}_i^T \hat{\beta}_L)^2 . \qquad (1.4)$$

The resulting weighted least squares estimate will be denoted by $\hat{\beta}_{EL}$.

A third method is the normal theory maximum likelihood estimate $\hat{\beta}_{ML}$.

which is a weighted least squares estimate with weights the inverse of

$$\hat{\sigma}^2_{i3} = \hat{\sigma}^2_{i2}(\hat{\beta}_{ML}). \tag{1.5}$$

This can be thought of as an iterated version of $\hat{\beta}_{EL}$.

These methods have been discussed in the literature for normally distributed errors. Bement and Williams (1969) use (1.3), and construct approximations (as $m \to \infty$) for the exact covariance matrix of the resulting weighted least squares estimate. They do not discuss asymptotic distributions as $N \to \infty$ with m fixed. Fuller & Rao (1978) use (1.4) while Cochran (1937) and Neyman & Scott (1948) use (1.5). Both find limiting distributions as $N \to \infty$ for fixed $m \geq 3$, although the latter two papers consider only the case that $\mathbf{x}^T\beta \equiv \mu$.

One striking result concerns consistency. The estimates $\hat{\beta}_{SV}$, $\hat{\beta}_{EL}$ and $\hat{\beta}_{ML}$ are always consistent for symmetrically distributed errors but generally not otherwise: see Theorems 1 and 3. In Section 5, we compute the limit distributions of $\hat{\beta}_{EL}$ and $\hat{\beta}_{ML}$. The relative efficiency of the two is contrasted in the normal case for $m \geq 3$, as follows.

Remark 1 : If ordinary least squares is less than 3 times more variable than weighted least squares with known weights, then $\hat{\beta}_{EL}$ is more efficient than maximum likelihood.

Remark 2 : If ordinary least squares is more than 5 times more variable than weighted least squares with known weights, then maximum likelihood is more efficient.

Further, for normally distributed data, maximum likelihood is more variable than weighted least squares with known weights by a factor $m/(m-2)$. This means a tripling of variance for $m = 3$ even when using maximum likelihood.

Section 2 : Assumptions and Canonical Decomposition

We will assume throughout that $(\mathbf{x}_i, \sigma_i)$ are independent and identically distributed bounded random vectors, independently distributed of the $\{\epsilon_{ij}\}$. We define $\mathbf{z}_i = \mathbf{x}_i/\sigma_i$ and $\hat{d}_i = \hat{\sigma}_i/\sigma_i$. For any weighted least squares estimator with estimated weights $\hat{w}_i = 1/\hat{\sigma}_i^2$,

$$\hat{\beta} - \beta = \left[N^{-1} \sum_{i=1}^{N} \mathbf{z}_i \mathbf{z}_i^T / \hat{d}_i^2\right]^{-1} N^{-1} \sum_{i=1}^{N} \mathbf{z}_i \bar{\epsilon}_i / \hat{d}_i^2 . \tag{2.1}$$

Assuming they exist, we note that the asymptotic covariance of the weighted and unweighted least squares estimators are, respectively,

$$S_{WLS}^{-1} = \{E(\mathbf{z}\mathbf{z}^T)\}^{-1} \tag{2.2}$$

and

$$S_L^{-1} = \{E(\mathbf{x}\mathbf{x}^T)\}^{-1} E(\sigma^2 \mathbf{x}\mathbf{x}^T) \{E(\mathbf{x}\mathbf{x}^T)\}^{-1}. \tag{2.3}$$

Section 3 : Weighting with Sample Variances

In this section, we describe consistency and asymptotic normality for weighted least squares estimates $\hat{\beta}_{SV}$ with the weights being the inverse of sample variances. We first describe the general case assuming that sufficient moments exist. We then look more closely at the case of normally distributed observations. In this setup,

$$\hat{d}_i^2 = (m-1)^{-1} \sum_{j=1}^{m} (\epsilon_{ij} - \bar{\epsilon}_i)^2 .$$

Define $\eta_{jk} = E(\bar{\epsilon}_i^j / \hat{d}_i^{2k})$ and $\upsilon_{jk} = E(|\bar{\epsilon}_i|^j / \hat{d}_i^{2k})$.

The first result indicates that we obtain consistency only when

$$\eta_{11} = E(\bar{\epsilon}_i / \hat{d}_i^2) = 0. \tag{3.1}$$

This is true for symmetrically distributed data, but generally not otherwise.

THEOREM 1 :

(a) *If* $v_{11} < \infty$ *and* $v_{01} < \infty$, *then*

$$\text{plim } \hat{\beta}_{SV} = \beta + (\eta_{01} S_{WLS})^{-1} \eta_{11} E(z),$$

so that consistency holds only if $E(z) = 0$ *or* (3.1) *holds.*

(b) *If* $v_{jk} < \infty$ *for* $j \leq 2$, $k \leq 2$ *and* (3.1) *holds, then* $(Nm)^{1/2} (\hat{\beta}_{SV} - \beta)$ *is asymptotically* Normal$\{0, m(\eta_{22}/\eta_{01}^2) S_{WLS}^{-1}\}$. □

Proof of Theorem 1 : This follows from the weak law of large numbers and the central limit theorem. □

For normally distributed observations, the assumption that $v_{jk} < \infty$ for $j, k \leq 2$ holds only if there are at least 6 replicates. In this case, we have the following corollary.

COROLLARY 1 : *Assume that the errors* ϵ_{ij} *are normally distributed. For* $m \geq 6$, $(Nm)^{1/2} (\hat{\beta}_{SV} - \beta)$ *is asymptotically* Normal$(0, (m-3)/(m-5) S_{WLS}^{-1})$. □

Comparing with (1.2), we see that the effect of using $m \geq 6$ replicates to estimate sample variances causes an inflation of variance by the factor $(m-3)/(m-5)$ over weighted least squares with known weights. Even with $m = 10$, this results in a 40% increase in variance.

If one uses a standard statistical package with weights $1/s_i^2$, then the resulting standard errors will also be asymptotically incorrect. Such packages estimate the asymptotic covariance matrix of $(Nm)^{1/2} (\hat{\beta}_{SV} - \beta)$ by $\hat{\sigma}_{SV}^2 \hat{S}_{WLS}^{-1}$, where

$$\hat{\sigma}_{SV}^2 = (Nm-p)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{m} (y_{ij} - x_i^T \hat{\beta}_{SV})^2/s_i^2 ;$$

$$\hat{S}_{WLS} = N^{-1} \sum_{i=1}^{N} x_i x_i^T / s_i^2 .$$

If $m \geq 6$ and if the data are normally distributed, then $\hat{\sigma}^2_{SV}$ converges in probability to $E(1/\hat{d}_i^2) = (m-1)/(m-3)$, while $(\hat{S}_{WLS} - (m-1)/(m-3)\ S_{WLS}) \to 0$. Thus, $(\hat{\sigma}^2_{SV}\ \hat{S}^{-1}_{WLS} - S^{-1}_{WLS}) \to 0$. Asymptotically, therefore, standard errors should be multiplied by $\{(m-3)/(m-5)\}^{1/2}$: see Table 1.

Section 4 : Sample Variances With $m \leq 5$ Replicates in the Normal Case

In this section, we consider normally distributed data with $m \leq 5$ replicates and the weights being the inverses of the sample variances. Here Theorem 1 does not apply since $\bar{\epsilon}_i/\hat{d}_i^2$ does not have finite variance. The results here are based on the work of Cline (1986). We first state a general result which may be of independent interest. The results for weighted least squares, assuming normal errors, are then derived as a corollary.

First, a few definitions are required. A positive function μ is regularly varying with exponent ρ, denoted by $\mu \in RV(\rho)$, if

$$\mu(yt)/\mu(t) \to y^{\rho} \text{ as } t \to \infty \text{ for all } y > 0.$$

Let $(z_i,\ u_i,\ w_i)$ be independent and identically distributed random variables with $z_i \in \mathbb{R}^p$ independent of $(u_i,\ w_i)$, u_i with a symmetric distribution and $w_i > 0$. Define $\mu_1(t) = E\{w\ I(w \leq t)\}$ and $\mu_2(t) = E\{(uw)^2\ I(uw \leq t)\}$. Let $(c_{1N},\ c_{2N})$ be constants satisfying, as $N \to \infty$, $N\ \mu_1(c_{1N})/c_{1N} \to 1$ and $N\ \mu_2(c_{2N})/c_{2N}^2 \to 1$.

If $\alpha_1 < 1$, then $S_1 = S_1(\alpha_1)$ will denote a positive stable random variable with Laplace transform

$$E\{\exp(-tS_1)\} = \exp\{-\Gamma(2-\alpha_1)\ t^{\alpha_1}/\alpha_1\}.$$

If $\alpha_1 = 1$ then $S_1 = 1$ almost surely. We will denote by $S_2 = S_2(\alpha_2)$ a symmetric stable random variable with characteristic function

$$E\{\exp(itS_2)\} = \exp\left\{-\Gamma(3-\alpha_2)\cos(\pi\alpha_2/2)\ |t|^{\alpha_2}/(\alpha_2(1-\alpha_2))\right\}.$$

Of course, if $\alpha_2 = 2$ then S_2 is standard normal.

THEOREM 2 : *Assume that* $\mu_1 \in RV(1-\alpha_1)$, $\mu_2 \in RV(2-\alpha_2)$ *and that*

$$\left[c_{1N}^{-1}\sum_{i=1}^{N} w_i,\ c_{2N}^{-1}\sum_{i=1}^{N} u_i w_i\right] \tag{4.1}$$

is asymptotically distributed as $\left[S_1(\alpha_1),\ S_2(\alpha_2)\right]$. *Suppose that for some* $\delta > 0$ *and all* i,j,

$$E(|z_{ij}|^{\gamma}) < \infty \text{ for } \gamma = \min(2,\ \max(2\alpha_1,\alpha_2) + \delta).$$

Then there exists Y_1, Y_2 ($Y_2 \in \mathbb{R}^p$, Y_1 *pxp positive definite*) *such that*

$$\mathbf{b}_N = (c_{1N}/c_{2N})\left[\sum_{i=1}^{N} z_i z_i^T w_i\right]^{-1}\sum_{i=1}^{N} z_i u_i w_i$$

is asymptotically distributed as $Y_1^{-1}Y_2$. *Further, for any* $\mathbf{b} \in \mathbb{R}^p$, $\mathbf{b}^T Y_1 \mathbf{b}$ *and* $\mathbf{b}^T Y_2$ *have the same distributions, respectively, as*

$$\left\{E\left[|\mathbf{b}^T\mathbf{z}|^{2\alpha_1}\right]\right\}^{1/\alpha_1} S_1 \text{ and } \left\{E\left[|\mathbf{b}^T\mathbf{z}|^{\alpha_2}\right]\right\}^{1/\alpha_2} S_2. \qquad \square$$

Remark 3 : In Theorem 2, Y_1 and Y_2 are not necessarily independent unless $\alpha_1 = 1$ or $\alpha_2 = 2$. In the former case, $Y_1 = E(zz^T)$ almost surely, while in the latter case Y_2 is normally distributed with mean zero and covariance $E(zz^T)$.

Proof of Theorem 2 : Consider first the case $\alpha_1 < 1$. From Theorem 1 of Cline (1986) we get

$$\left[c_{1N}^{-1}\sum_{i=1}^{N} z_i z_i^T w_i,\ c_{2N}^{-1}\sum_{i=1}^{N} z_i u_i w_i\right]$$

is asymptotically distributed as (Y_1, Y_2). In the case that $\alpha_1 = 1$, then $S_1 \equiv 1$ almost surely by Feller (1971, p. 236). From unpublished work of Cline and

from Gnedenko & Kolmogorov ((1954), p. 134), for each (j,k),

$$\text{plim } c_{1N}^{-1} \sum_{i=1}^{N} z_{ij} z_{ik} w_i = E(z_{ij} z_{ik}).$$

The joint convergence of the remaining terms, $c_{2N}^{-1} \Sigma z_i u_i w_i$, again follows from Theorem 1 of Cline (1986).

In either case, convergence of the ratio $\mathbf{b}_N$ follows. The limiting joint distribution is difficult to describe, but the stated marginal distributions of $\mathbf{b}^T Y_1 \mathbf{b}$ and $\mathbf{b}^T Y_2$ can be inferred from Proposition 3 of Breiman (1965) and Theorem 3 of Maller (1981). One may also conclude that Y_1 and Y_2 are independent if $\alpha_1 = 1$, since then Y_1 is degenerate. Also, Y_1 and Y_2 are independent if $\alpha_2 = 2$, since then Y_2 is Gaussian, and for such limits the non-Gaussian stable component is always independent of the Gaussian component (c.f., Sharpe, 1969). □

Special Cases : If $(t/a_1)^{\tau_1} \Pr(w > t) \to 1$ for $\tau_1 > 0$, then we have the following cases:

(i) If $\tau_1<1$, then $\alpha_1=\tau_1$, $c_{1N} = a_1 \{N/(1-\alpha_1)\}^{1/\alpha_1}$, and S_1 is positive stable.

(ii) If $\tau_1=1$, then $\alpha_1=1$, $c_{1N} = a_1 N \log(N)$, and $S_1 \equiv 1$.

(iii) If $\tau_1>1$, then $\alpha_1=1$, $c_{1N} = N E(w)$, and $S_1 \equiv 1$.

If $(t/a_2)^{\tau_2} \Pr(|uw| > t) \to 1$ for $\tau_2 > 0$, then we have the following cases:

(i) If $\tau_2<2$, then $\alpha_2=\tau_2$, $c_{2N} = a_2 \{N/(2-\alpha_2)\}^{1/\alpha_2}$, and S_2 is symmetric stable.

(ii) If $\tau_2=2$, then $\alpha_2=2$, $c_{2N} = 2^{1/2} a_2 N^{1/2} \log(N)$, and S_2 is Normal.

(iii) If $\tau_2>2$, then $\alpha_2=2$, $c_{2N} = N^{1/2} \left[E|uw|^2\right]^{1/2}$, and S_2 is Normal.

Consider the case of normally distributed errors in model (1.1), where we make the identifications $z_i = x_i/\sigma_i$, with $E(zz^T) = S_{WLS}$. Further, write

$$u_i = \bar{\epsilon}_i = m^{-1} \sum_{j=1}^{m} \epsilon_{ij}$$

and

$$w_i = 1/\hat{d}_i^2 = \left\{(m-1)^{-1} \sum_{j=1}^{m} (\epsilon_{ij} - \bar{\epsilon}_i)^2\right\}^{-1}.$$

Of course, u_i and w_i are independent and $E(u_i^2) < \infty$. Set $\alpha = (m-1)/2$, $a = \alpha \{\Gamma(1+\alpha)\}^{-1/\alpha}$ and $b = (2/\pi m)^{1/2} \{\Gamma((1+\alpha)/2)\}^{1/\alpha}$. Then $(t/a)^{\alpha} \Pr(w > t) \to 1$, $(t/ab)^{\alpha} \Pr(|uw| > t) \to 1$ and if $\alpha > 1$, $E(w) = \alpha/(\alpha-1)$. With the indicated choices of c_{1N} and c_{2N}, Theorem 1 of Cline (1986a) shows that (4.1) holds. Thus the conditions of Theorem 2 are met.

COROLLARY 2 : *In the normally distributed case, with S_1, S_2, Y_1, Y_2 as defined in Theorem 2, we have the following cases:*

Case 1 ($m = 2$) : $\alpha_1 = \alpha_2 = 1/2$, *and*

$(\hat{\beta}_{SV} - \beta)$ *is asymptotically distributed as* $\Gamma^2(3/4)/(9\pi)\ Y_1^{-1}\ Y_2$.

Case 2 ($m = 3$) : $\alpha_1 = \alpha_2 = 1$, *and*

$\log(N)\ (\hat{\beta}_{SV} - \beta)$ *is asymptotically distributed as* $\{2/(3\pi)\}^{1/2}\ Y_1^{-1}\ Y_2$.

Case 3 ($m = 4$) : $\alpha_1 = 1$, $\alpha_2 = 3/2$, *and*

$N^{1/3}\ (\hat{\beta}_{SV} - \beta)$ *is asymptotically distributed as* $2^{-1/2}\ \{\Gamma^2(1/4)/(18\pi)\}^{1/3}\ Y_1^{-1}\ Y_2$.

Case 4 ($m = 5$) : $\alpha_1 = 1$, $\alpha_2 = 2$ *and*

$N^{1/2}/\log(N)\ (\hat{\beta}_{SV} \quad \beta)$ *is asymptotically distributed as* $5^{-1/2}\ Y_1^{-1}\ Y_2$.

Case 5 ($m > 6$) : *Covered by Corollary 1 already.* □

Proof of Corollary 2 : In the notation of Theorem 2, $b_N = (c_{1N}/c_{2N})\ (\hat{\beta}_{SV} - \beta)$ is asymptotically distributed as $Y_1^{-1}\ Y_2$. Thus, in each case it suffices to construct the constants $(c_{1N},\ c_{2N})$.

(m = 2) : Here $\alpha = 1/2$, $c_{1N} = (8/\pi)\ N^2$ and $c_{2N} = \{8\ \Gamma^2(3/4)/(9\pi^2)\}\ N^2$.

(m = 3) : Here $\alpha = 1$, $c_{1N} = N\ \log(N)$ and $c_{2N} = \{2/(3\pi)\}^{1/2}\ N$.

(m = 4) : Here $\alpha = 3/2$, $c_{1N} = 3\ N$ and $c_{2N} = 2^{-1/2}\ \{3\ \Gamma^2(1/4)/(2\pi)\}^{1/3}\ N^{2/3}$.

(m = 5) : Here $\alpha = 2$, $c_{1N} = 2\ N$ and $c_{2N} = 2\ 5^{-1/2}\ N^{1/2}\ \log(N)$. □

Section 5 : Estimating Variances by Sample Average Squared Errors

One might reasonably conjecture that making use of the known linear structure for the means results in improvements over using only sample variances. We will show that this is the case, at least for normally distributed data. Let $\hat{\beta}_0$ be any estimate of β, and define

$$\hat{\sigma}_i^2(\hat{\beta}_0) = m^{-1} \sum_{j=1}^{m} (y_{ij} - x_i^T\ \hat{\beta}_0)^2$$

and $\hat{d}_i(\hat{\beta}_0) = \hat{\sigma}_i(\hat{\beta}_0)/\sigma_i$. We denote by $\hat{\beta}_G$ the weighted estimate with the estimated weights $1/\hat{\sigma}_i^2(\hat{\beta}_0)$. As defined in the introduction, $\hat{\beta}_{EL}$ uses $\hat{\beta}_0 = \hat{\beta}_L$, the ordinary unweighted least squares estimate, and $\hat{\beta}_{ML}$ used $\hat{\beta}_0 = \hat{\beta}_{ML}$. Our results here rely on the consistency of $\hat{\beta}_0$, and two other reasonable moment conditions for m large enough. Here are the assumptions.

$$\text{plim}\ \hat{\beta}_0 = \beta. \tag{5.1}$$

For each $c_1 > 0$, there exists $c_2 > 0$ such that

$$E\left\{ \sup_{\|\beta_* - \beta\| \le c_2} |\hat{d}_i^{-2}(\beta_*) - \hat{d}_i^{-2}(\beta)| \right\} \le c_1 \tag{5.2}$$

and such that

$$E\left\{ \sup_{\|\beta_* - \beta\| \le c_2} |\bar{\epsilon}_i| \left| \frac{\partial}{\partial \beta} \hat{d}_i^{-2}(\beta_*) - \frac{\partial}{\partial \beta} \hat{d}_i^{-2}(\beta) \right| \right\} \le c_1. \tag{5.3}$$

In addition, we assume the finite existence of

$$\eta_{jk} = E\{\bar{\epsilon}_i^j / \hat{d}_i^{2k}(\beta)\}, \quad \upsilon_{jk} = E\{|\bar{\epsilon}_i|^j / \hat{d}_i^{2k}(\beta)\} < \infty \quad \text{for } j,k \le 2. \tag{5.4}$$

The first result describes the consistency of $\hat{\beta}_G$.

THEOREM 3 : *Assume* (5.1) - (5.4). *Then*

$$\text{plim } \hat{\beta}_G = \beta + \eta_{11}(\eta_{01} S_{WLS})^{-1} E(z). \tag{5.5}$$

Thus, $\hat{\beta}_G$ *is consistent only if* $\eta_{11} = 0$ *or* $E(z) = 0$. *Further, as* $N \to \infty$,

$$N^{1/2} (\hat{\beta}_G - \beta) = A_N^{-1} [b_N + C_N N^{1/2} (\hat{\beta}_0 - \beta)], \tag{5.6}$$

where plim $A_N = \eta_{01} S_{WLS}$, $b_N = N^{-1/2} \Sigma z_i \bar{\epsilon}_i / \hat{d}_i^2(\beta)$ *and* plim $C_N = 2 \eta_{22} S_{WLS}$. □

Proof of Theorem 3 : Since the $\{z_i\}$ are bounded, the assumptions make possible the usual Taylor's series argument leading to (5.6) and (5.5) is an immediate consequence of (5.6). □

Assuming consistency of the maximum likelihood estimator, we can compute the limit distributions of $\hat{\beta}_{EL}$ and $\hat{\beta}_{ML}$.

THEOREM 4 : *Make the assumptions of Theorem* 3, *with the* $\{\epsilon_{ij}\}$ *being symmetrically distributed.*

(a) *Let* $V_1 = E(xx^T)$ *and* $V_2 = E(\sigma^2 xx^T)$ *be finite and positive definite. Let* $S_L^{-1} = V_1^{-1} V_2 V_2^{-1}$. *Then with* $\hat{\beta}_0$ *chosen as the unweighted least squares estimate,* $(Nm)^{1/2} (\hat{\beta}_{EL} - \beta)$ *is asymptotically* Normal$(0, S_{EL}^{-1})$, *where*

$$S_{EL}^{-1} = m(\eta_{22}/\eta_{01}^2) \{(1 + 4 \eta_{21}) S_{WLS}^{-1} + 4 \eta_{22} S_L^{-1}\} .$$

(b) *With* $\hat{\beta}_0 = \hat{\beta}_{ML}$, *the maximum likelihood estimate, then*

$(Nm)^{1/2} (\hat{\beta}_{ML} - \beta)$ *is asymptotically* Normal$(0, S_{ML}^{-1})$, *where*

$$S_{ML}^{-1} = m \eta_{22} (\eta_{01} - 2\eta_{22})^{-2} S_{WLS}^{-1}.$$ □

Proof of Theorem 4 : Parts (a) and (b) follow easily from Theorem 3, Slutsky's theorem and the fact that $(Nm)^{1/2}\ (\hat{\beta}_L - \beta)$ is asymptotically Normal$(0, S_L^{-1})$.. □

COROLLARY 3 : *For normally distributed observations with* $m \geq 3$, $(Nm)^{1/2}\ (\hat{\beta}_{ML} - \beta)$ *and* $(Nm)^{1/2}\ (\hat{\beta}_{EL} - \beta)$ *are asymptotically normally distributed with respective covariances*

$$\{m/(m-2)\}\ S_{WLS}^{-1} \text{ and } \left\{(1 + 2m^{-1} - 8m^{-2})\ S_{WLS}^{-1} + 4m^{-2}\ S_L^{-1}\right\}. \quad \square$$

Proof of Corollary 3 : By direct calculation as in Fuller & Rao (1978),

$$\eta_{01} = m/(m-2), \quad \eta_{21} = 1/m \text{ and } \eta_{22} = 1/(m-2). \quad \square$$

As noted by Fuller & Rao (1978), the asymptotic covariance of $\hat{\beta}_{EL}$ consists of a mixture of the weighted least squares covariance S_{WLS}^{-1} and the unweighted least squares covariance S_L^{-1}. Comparing $\hat{\beta}_{EL}$ with the maximum likelihood estimate $\hat{\beta}_{ML}$ depends on how much bigger S_L^{-1} is than S_{WLS}^{-1}. Detailed calculations verify Remarks 1 and 2 of the introduction. Thus, doing iterative weighted least squares may actually hurt, unless the starting value $\hat{\beta}_L$ is sufficiently bad.

Section 6 : Discussion

Our results can be summarized as follows :

(a) If nothing is known about the structure of the sample variances, then none of the common weighted estimates can be assumed to be consistent for data from an asymmetric distribution.

(b) Using sample variances as a basis for estimating weights is inefficient unless the number of replicates m is fairly large, e.g., $m \geq 10$.

(c) Using sample average squared errors from a preliminary fit to the regression function as a basis for estimating weights is typically more efficient than using sample variances. However, even here a fair number of replicates is helpful. For example, the maximum likelihood estimate for normally distributed data based on 6 replicates still has standard errors approximately 20% larger than ordinary weighted least squares theory would suggest.

There are at least two alternative methods for estimating the weights. The first is to model the variances parametrically, e.g.,

$$\sigma_i = \sigma(x_i^T \beta)^\theta.$$

See Carroll & Ruppert (1987) and Davidian & Carroll (1987). The second is to perform a nonparametric regression of (1.3) and (1.4) against the predictors and use this regression to estimate the weights (c.f. Carroll, 1982).

ACKNOWLEDGEMENT

The work of R. J. Carroll was supported by the Air Force Office of Scientific Research and the work of D. B. H. Cline was partially supported by the National Science Foundation. We thank the referee for the Cochran and the Yates and Cochran references.

REFERENCES

BEMENT, T. R. & WILLIAMS, J. S. (1969). Variance of weighted regression estimators when sampling errors are independent and heteroscedastic. *J. Am. Statist. Assoc.* **64**, 1369-1382.

BREIMAN, L. (1965). On some limit theorems similar to the arc-sine law.

Theory Prob. Applications 10, 323-331.

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

CARROLL, R. J. & RUPPERT, D. (1987). *Transformations and Weighting in Regression*. Chapman & Hall, London.

CLINE, D. B. H. (1986). Joint stable attraction of two sums of products. To appear, *J. Mult. Anal.*

COCHRAN, W. A. (1937). Problems arising in the analysis of a series of similar experiments. *J. Royal Stat. Soc., Suppl.* 4, 102-118.

DAVIDIAN, M. & CARROLL, R. J. (1987). Variance function estimation. *J. Am. Statist. Assoc.*, to appear.

FELLER, W. (1971). *An Introduction to Probability Theory and its Applications, Vol. II*. Wiley, New York.

FULLER, W. A. & RAO, J. N. K.(1978). Estimation for a linear regression model with unknown diagonal covariance matrix. *Ann. Statist.* 6, 1149-1158.

GARDEN, J. S., MITCHELL, D. G. & MILLS, W. N. (1980). Nonconstant variance regression techniques for calibration curve based analysis. *Analytical Chemistry* 52, 2310-2315.

GNEDENKO, B. V. & KOLMOGOROV, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Cambridge.

MALLER, R. A. (1981). A theorem on products of random variables, with applications to regression. *Aust. J. Statist.* 23, 177-185.

NEYMAN, J. & SCOTT, E. L. (1948). Consistent estimators based on partially consistent observations. *Econometrica* 16, 1-32.

SHARPE, M. (1969). Operator stable distributions on vector spaces. *Trans. Amer. Math. Soc.* 136, 51-65.

YATES, F. & COCHRAN, W. G. (1938). The analysis of groups of

experiments. *J. Agric. Sci.* **28**, 556-580, reprinted in *Experimental Design: Selected Papers of Frank Yates* (1970), Griffin, London.

TABLE 1

A summary of the results when the weights are the inverses of sample variances based on m replicates. The relative efficiency is calculated with respect to weighted least squares with known weights. The column labelled "Standard Error Factor" is the number one should multiply standard errors from a weighted least squares package by to obtain asymptotically correct standard errors.

(m)	Consistent?	Asymptotically Normal?	Rate of convergence	Relative Efficiency	Standard Error Factor
2	No	No	-	0	-
3	Yes	No	$\log(N)$	0	-
4	Yes	No	$N^{1/3}$	0	-
5	Yes	Yes	$N^{1/2}/\log(N)$	0	-
6	Yes	Yes	$N^{1/2}$	1/3	1.73
7	Yes	Yes	$N^{1/2}$	1/2	1.41
8	Yes	Yes	$N^{1/2}$	3/5	1.26
9	Yes	Yes	$N^{1/2}$	2/3	1.22
10	Yes	Yes	$N^{1/2}$	5/7	1.18