JACKKNIFING WEIGHTED LEAST SQUARES ESTIMATORS

by

Jun Shao

Purdue University
Technical Report #87-57

Department of Statistics
Purdue University

December 1987

JACKKNIFING WEIGHTED LEAST SQUARES ESTIMATORS

Jun Shao

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

SUMMARY

In a heteroscedastic linear regression model, the ordinary jackknife estimator of the asymptotic covariance matrix of the weighted least squares estimator is proved to be inconsistent. A modified jackknife procedure is proposed and shown to produce consistent estimator of the asymptotic covariance matrix. Finite sample performances of the jackknife, the modified jackknife and the customary δ-method are discussed. Some empirical results are also presented.

Keywords: HETEROSCEDASTICITY; WEIGHTED LEAST SQUARES; ASYMPTOTIC COVARIANCE MATRIX; JACKKNIFE; δ-METHOD; CONSISTENCY.

1. INTRODUCTION

In this article we study the jackknife method for estimating asymptotic covariance matrix of the weighted least squares estimator in a regression problem. Consider the following heteroscedastic linear regression model:

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

where y_{ij} is the jth response at the ith design point x_i, x_i are known p-vectors, β is a p-vector of unknown parameters, and e_{ij} are mutually independent with $Ee_{ij}=0$ and $Ee_{ij}^2=\sigma_i^2$, $j=1,...,n_i$. The variances σ_i^2 are unknown and unequal (heteroscedastic). A matrix form of model (1.1) is

$$y = X\beta + e,$$

where

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

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

and

$$X = (\, x_{11} \cdots x_{1n_1} \cdots \quad \cdots x_{k1} \cdots x_{kn_k}\,)'_{n\times p}, \quad x_{ij}= x_i,\ j=1,...,n_i.$$

Note that X and y depend on the sample size n. Strictly, we have $X=X_n$ and $y=y_n$, but the subscript n is omitted for simplicity. The design matrix X is assumed to be of full rank. In most of practical applications, we need to estimate the parameter $\theta=g(\beta)$, $g: \mathbf{R}^p \to \mathbf{R}^q$, and make statistical inferences for θ such as setting confidence intervals or testing statistical hypotheses.

Fuller and Rao (1978) introduced a weighted least squares estimator of β obtained by the following two steps.

(1) Obtain the ordinary least squares estimator (OLSE) of β:

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

and estimates of σ_i^2:

$$v_i = n_i^{-1}\textstyle\sum_{j=1}^{n_i} r_{ij}^2, \quad r_{ij} = y_{ij} - x_i'\hat\beta. \tag{1.2}$$

(2) By using the reciprocals of v_i as weights, obtain the weighted least squares estimator (WLSE) of β:

$$\hat{\beta}^w = (X'WX)^{-1}(X'Wy) \tag{1.3}$$

where

$$W = block\ diag.(\ w_1 I_{n_1} \ldots w_k I_{n_k}\ ), \tag{1.4}$$

I_t is the $t \times t$ identity matrix and $w_i = v_i^{-1}$. The WLSE of $\theta = g(\beta)$ is then $\hat{\theta}^w = g(\hat{\beta}^w)$.

The weighted least squares method provides more efficient estimators than the ordinary least squares method in some situations (Fuller and Rao (1978)).

The number of replicates at a design point is usually small for a regression problem. In view of this, Fuller and Rao (1978) and Shao (1988a) obtained the asymptotic distribution of $\hat{\theta}^w$ ($k \to \infty$ and $n_i \le n_\infty$ for a fixed n_∞) under some conditions (see Proposition 1 in the next section).

For further statistical inferences for θ, one needs a consistent (as $k \to \infty$) estimator of the asymptotic covariance matrix of $\hat{\theta}^w$. A customary approach is to use the δ-method (described in Section 2). Another approach is to use the resampling methods such as the jackknife (Quenouille (1956), Tukey (1958)), the half-sampling (McCarthy (1969)) and the bootstrap (Efron (1979)), which involve resampling the original data and estimating the variances from the resamples. Because of the availability of inexpensive and friendly computing, the resampling methods have caught on very rapidly in recent years. The validation of the resampling methods were justified in many situations. See Efron (1982), Parr (1985), Shao and Wu (1986) and their references. For applications of the jackknife method in linear regression, see Miller (1974), Hinkley (1977), Wu (1986), Shao and Wu (1987) and Shao (1988b).

However, in the present problem the usual jackknife estimator of the asymptotic covariance matrix of the WLSE (described in Section 2) is *inconsistent* and tends to underestimate (Theorem 1 of Section 3), because of the inconsistency of the estimators of σ_i^2 defined in (1.2). Although there are several other estimators of σ_i^2 proposed by various authors (see Section 5(iv)), no consistent estimator of σ_i^2 is available unless $n_i \to \infty$ or we assume that $\sigma_i^2 = H(x_i)$, where $H(t)$ is a smooth function (Carroll (1982), Muller and Stadtmuller (1987)).

In Section 3, a modified jackknife estimator is proposed and shown to be consistent and asymptotically equivalent to the estimator obtained by the δ-method. Finite sample empirical comparisons of the variance estimators obtained by the jackknife, the modified jackknife and the δ-method are presented in Section 4. Section 5 contains some concluding remarks about the comparisons of the jackknife and the δ-method and some other issues.

2. THE JACKKNIFE AND THE δ-METHOD

We will use the following assumptions in proving the main results:

Assumption (a). There are positive constants σ_0^2, σ_∞^2 and c_∞ and positive integers n_0 and n_∞ such that $\sigma_0^2 \le \sigma_i^2 \le \sigma_\infty^2$, $n_0 \le n_i \le n_\infty$ and $\|x_i\| \le c_\infty$ for all i, where $\|x\| = (x'x)^{1/2}$.

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 δ are positive constants.

Most of the error distributions encountered in practice satisfy (2.2). See Shao (1988a). Condition (2.1) reflects certain degree of symmetry of the error distribution.

Proposition 1. Suppose that Assumptions (a)-(c) hold. Let $\hat{\beta}^w$ be defined as in (1.3) and $\hat{\theta}^w = g(\hat{\beta}^w)$. Assume that the function g is continuously differentiable at β and $\nabla g(\beta)$ is of full rank, where $\nabla g(\beta)$ is the gradient of g at β. Then

$$(V^g)^{-1/2}(\hat{\theta}^w - \theta) \to N(0, I_q) \quad \text{in distribution},$$

where $(V^g)^{-1/2}$ is the inverse of a square root of

$$V^g = \nabla g(\beta) V (\nabla g(\beta))' \tag{2.3}$$

with

$$V = (X'D_1X)^{-1} + 4(X'D_1X)^{-1}X'D_2X(X'D_1X)^{-1} \tag{2.4}$$

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

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

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

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

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

The proof of this result can be found in Shao (1988a). The matrix V^g is called the asymptotic covariance matrix of $\hat{\theta}^w$. An estimator of V^g based on the δ-method is

$$\hat{V}_\delta^g = \nabla g(\hat{\beta}^w)\hat{V}_\delta(\nabla g(\hat{\beta}^w))',$$

where

$$\hat{V}_\delta = (X'WX)^{-1} + 4(X'WX)^{-1}X'W_2X(X'WX)^{-1}$$

$$+ 4(X'WX)^{-1}X'W_2X(X'X)^{-1}X'W^{-1}X(X'X)^{-1}X'W_2X(X'WX)^{-1},$$

$$W_2 = block\ diag.(\ n_1^{-1}w_1 I_{n_1} \ldots n_k^{-1}w_k I_{n_k}\ ),$$

and W is defined in (1.4). Under Assumptions (a)-(c), Shao (1988a) proved the consistency of this estimator, i.e., $k\hat{V}_\delta^g - kV^g \to_p 0$, where $\to_p$ denotes convergence in probability.

Let $z_i=w_i^{1/2}x_i$, $Z=W^{1/2}X$ and $\eta_{ij}=w_i^{1/2}y_{ij}$. The jackknife estimator estimates V^g from the variations in $\hat{\theta}^w-\hat{\theta}_{(i,j)}^w$, where $\hat{\theta}_{(i,j)}^w=g(\hat{\beta}_{(i,j)}^w)$ and $\hat{\beta}_{(i,j)}^w$ is the WLSE of β after deleting the (i,j)th "data" z_i and η_{ij}. We will focus on the following weighted jackknife variance estimator (Wu, 1986):

$$\hat{V}_J^g = \sum_{i=1}^k \sum_{j=1}^{n_i} (1-h_i)(\hat{\theta}_{(i,j)}^w-\hat{\theta}^w)(\hat{\theta}_{(i,j)}^w-\hat{\theta}^w)', \tag{2.5}$$

where

$$h_i = \max_{i\le k} z_i'(Z'Z)^{-1}z_i \tag{2.6}$$

and $\{\ (1-h_i)/(n-p)\ \}$ are the weights. If $w_i\equiv 1$, then $\hat{\beta}^w=\hat{\beta}$ and $\hat{V}_J^g$ defined in (2.5) is simply the weighted jackknife estimator of the asymptotic covariance matrix of the OLSE $\hat{\theta}=g(\hat{\beta})$. The weighted jackknife provides better variance estimators than the unweighted jackknife (Shao and Wu (1987), Shao (1988b)).

To compute the jackknife estimator without repeatedly fitting model (1.1), we can use an updating formula to obtain the WLSE's (Miller, 1974):

$$\hat{\beta}^w_{(i,j)} = \hat{\beta}^w - (1-h_i)^{-1}(Z'Z)^{-1}z_i(\eta_{ij} - z_i'\hat{\beta}^w). \tag{2.7}$$

Shao and Wu (1987) proved the consistency of the jackknife variance estimator of the OLSE. However, we will show in the next section that $\hat{V}_J^g$ is inconsistent. A consistent modified jackknife estimator is also proposed.

3. ASYMPTOTIC RESULTS

We state some lemmas before proving the main results. The proofs of these lemmas are given in the Appendix.

Lemma 1. Let $u_i = n_i^{-1}\sum_{j=1}^{n_i} e_{ij}^2$. Under Assumptions (a) and (c),

$$k^{-1}E(\max_{i\le k} u_i^{-1}) \to 0.$$

Lemma 2. Let $\Delta_i = |v_i u_i^{-1} - 1|$, where v_i is defined in (1.2). Under Assumptions (a)-(c),

$$\max_{i\le k} \Delta_i \to_p 0.$$

Lemma 3. Let $z_i = v_i^{-1/2} x_i$ and h_i be defined in (2.6). Under Assumptions (a)-(c),

$$\max_{i\le k} [z_i'(\hat{\beta}^w - \beta)]^2 \to_p 0 \quad \text{and} \quad \max_{i\le k} h_i \to_p 0.$$

Lemma 4. Suppose that Assumptions (a)-(c) hold. Then

$$\max_{j\le n_i,\, i\le k} \| \hat{\beta}^w_{(i,j)} - \hat{\beta}^w \| \to_p 0.$$

We first consider the special case of $\hat{\theta}^w = \hat{\beta}^w$. Denote the jackknife estimator (2.5) by $\hat{V}_J$ in this case. The following result shows the inconsistency of $\hat{V}_J$.

Theorem 1. Suppose that Assumptions (a)-(c) hold. Then

$$k\hat{V}_J - k(X'D_1X)^{-1} \to_p 0. \tag{3.1}$$

Proof. From (2.5)-(2.7), $u_i = n_i^{-1}\sum_{j=1}^{n_i} e_{ij}^2$ and $\eta_{ij} - z_i'\hat{\beta}^w = w_i^{1/2}[e_{ij} - x_i'(\hat{\beta}^w - \beta)]$,

$$k\hat{V}_J = k(Z'Z)^{-1}\sum_{i=1}^k\sum_{j=1}^{n_i}(1-h_i)^{-1}z_iz_i'(\eta_{ij}-z_i'\hat{\beta}^w)^2(Z'Z)^{-1} = S_1 + S_2 - S_{12},$$

where

$$S_1 = k(Z'Z)^{-1}\sum_{i=1}^k n_i(1-h_i)^{-1}z_iz_i'w_iu_i(Z'Z)^{-1},$$

$$S_2 = k(Z'Z)^{-1}\sum_{i=1}^k n_i(1-h_i)^{-1}z_iz_i'w_i[x_i'(\hat{\beta}^w-\beta)]^2(Z'Z)^{-1},$$

and

$$S_{12} = 2k(Z'Z)^{-1}\sum_{i=1}^k\sum_{j=1}^{n_i}(1-h_i)^{-1}z_iz_i'w_ie_{ij}[x_i'(\hat{\beta}^w-\beta)](Z'Z)^{-1}.$$

From Lemmas 2 and 3, $\max_{i\le k}h_i \to_p 0$ and $\max_{i\le k}|w_iu_i - 1| \to_p 0$. Thus $S_1 - k(Z'Z)^{-1} \to_p 0$. From Lemma 4 of Shao (1988a), $k(Z'Z)^{-1} - k(X'D_1X)^{-1} \to_p 0$. Hence

$$S_1 - k(X'D_1X)^{-1} \to_p 0. \tag{3.2}$$

From Lemma 3, $k(Z'Z)^{-1} - k(X'D_1X)^{-1} \to_p 0$ and $k(X'D_1X)^{-1} \le [c_0n_0\tau(n_\infty)]^{-1}\sigma_\infty^2 I_p$, we have

$$S_2 \le k(Z'Z)^{-1}(1-\max_{i\le k}h_i)^{-1}\max_{i\le k}[z_i'(\hat{\beta}^w-\beta)]^2 \to_p 0, \tag{3.3}$$

Then (3.1) follows from (3.2)-(3.3) and the Cauchy-Schwarz inequality. □

The asymptotic covariance matrix of $\hat{\beta}^w$ is given by (2.4). From (3.1), the jackknife estimator $\hat{V}_J$ is inconsistent and tends to underestimate. The bias of $\hat{V}_J$ is approximately

$$-4(X'D_1X)^{-1}[X'D_2X + X'D_2X(X'X)^{-1}X'DX(X'X)^{-1}X'D_2X](X'D_1X)^{-1}$$

and can be estimated by

$$-4(\tilde{V}_J + \tilde{V}_J\hat{V}_J^{-1}\hat{U}_J\hat{V}_J^{-1}\tilde{V}_J),$$

where

$$\tilde{V}_J = \sum_{i=1}^k\sum_{j=1}^{n_i}(1-h_i)n_i^{-1}(\hat{\beta}_{(i,j)}^w - \hat{\beta}^w)(\hat{\beta}_{(i,j)}^w - \hat{\beta}^w)',$$

$$\hat{U}_J = \sum_{i=1}^k\sum_{j=1}^{n_i}(1-c_i)(\hat{\beta}_{(i,j)} - \hat{\beta})(\hat{\beta}_{(i,j)} - \hat{\beta})'$$

is the jackknife estimator of the asymptotic covariance matrix of the OLSE, $c_i = x_i'(X'X)^{-1}x_i$ and $\hat{\beta}_{(i,j)}$ is the OLSE of β after deleting x_i and y_{ij}. By adjusting the bias of $\hat{V}_J$, we obtain the following modified jackknife estimator:

$$\hat{V}_M = \hat{V}_J + 4\tilde{V}_J + 4\tilde{V}_J\hat{V}_J^{-1}\hat{U}_J\hat{V}_J^{-1}\tilde{V}_J.$$

Theorem 2. Under Assumptions (a)-(c), the modified jackknife estimator $\hat{V}_M$ is consistent, i.e.,

$$k\hat{V}_M - kV \to_p 0.$$

Proof. From Lemma 4 of Shao (1988a), $k^{-1}\sum_{i=1}^k z_i z_i' - k^{-1}(X'D_2X) \to_p 0$. Following the same proof as given in Theorem 1, we have

$$k\tilde{V}_J - k(X'D_1X)^{-1}X'D_2X(X'D_1X)^{-1} \to_p 0. \tag{3.4}$$

From Theorem 3 of Shao and Wu (1987),

$$k\hat{U}_J - k(X'X)^{-1}X'DX(X'X)^{-1} \to_p 0. \tag{3.5}$$

The result follows from (2.4), (3.1), (3.4) and (3.5). □

In the important special case of $n_i = m$ for all i, $\tilde{V}_J = m^{-1}\hat{V}_J$ and therefore the modified jackknife estimator simplifies considerably:

$$\hat{V}_M = (1+4m^{-1})\hat{V}_J + 4m^{-2}\hat{U}_J.$$

We now consider the nonlinear case, i.e., the estimation of V^g defined in (2.3). The jackknife estimator is defined in (2.5). The modified jackknife estimator is

$$\hat{V}_M^g = \hat{V}_J^g + 4\tilde{V}_J^g + 4\tilde{V}_J^g(\hat{V}_J^g)^{-1}\hat{U}_J^g(\hat{V}_J^g)^{-1}\tilde{V}_J^g, \tag{3.6}$$

where

$$\tilde{V}_J^g = \sum_{i=1}^k \sum_{j=1}^{n_i} (1-h_i)n_i^{-1}(\hat{\theta}_{(i,j)}^w - \hat{\theta}^w)(\hat{\theta}_{(i,j)}^w - \hat{\theta}^w)',$$

$$\hat{U}_J^g = \sum_{i=1}^k \sum_{j=1}^{n_i} (1-c_i)(\hat{\theta}_{(i,j)} - \hat{\theta})(\hat{\theta}_{(i,j)} - \hat{\theta})',$$

and $\hat{\theta}_{(i,j)} = g(\hat{\beta}_{(i,j)})$. For the case of $n_i \equiv m$,

$$\hat{V}_M^g = (1+4m^{-1})\hat{V}_J^g + 4m^{-2}\hat{U}_J^g. \tag{3.7}$$

Theorem 3. Suppose that Assumptions (a)-(c) hold and ∇g is continuous in a neighborhood of β. Then

(i) $k\hat{V}_J^g - k\nabla g(\beta)(X'D_1X)^{-1}(\nabla g(\beta))' \to_p 0$ and therefore $\hat{V}_J^g$ is inconsistent;

(ii) In the special case of $n_i \equiv m$, $\hat{V}_M^g$ defined in (3.7) is consistent, i.e.,

$$k\hat{V}_M^g - kV^g \to_p 0;$$

(iii) In general $\hat{V}_M^g$ defined in (3.6) is consistent if $g: \mathbf{R}^p \to \mathbf{R}^p$ and $(\nabla g(\beta))^{-1}$ exists.

Proof. From Theorem 4 of Shao and Wu (1987),

$$k\hat{U}_J^g - k\nabla g(\beta)(X'X)^{-1}X'DX(X'X)^{-1}(\nabla g(\beta))' \to_p 0.$$

From Lemma 4, for any $\delta>0$,

$$\lim_{k\to\infty} P\{\ \|\hat{\beta}_{(i,j)}^w - \hat{\beta}^w\| \le \delta \ \text{ for } all \ (i,j)\ \} = 1.$$

The same argument used in the proof of Theorem 4 of Shao and Wu (1987) yields (i) and

$$k\tilde{V}_J^g - k\nabla g(\beta)(X'D_1X)^{-1}X'D_2X(X'D_1X)^{-1}(\nabla g(\beta))' \to_p 0.$$

This proves the results. □

For the application of Theorem 3 in the general case where some n_i are not equal and $g: \mathbf{R}^p \to \mathbf{R}^q$, $q<p$ but $\nabla g(\beta)$ is of full rank, we can construct a function $h: \mathbf{R}^p \to \mathbf{R}^{p-q}$ such that $f=(g', h')': \mathbf{R}^p \to \mathbf{R}^p$ satisfies the conditions in the theorem. Let $\hat{V}_M^f$ be the modified jackknife estimator of the asymptotic covariance matrix of $f(\hat{\beta}^w)$. Write $\hat{V}_M^f$ as

$$\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix},$$

where V_{11} is a $q\times q$ matrix. Then V_{11} is a consistent estimator of the asymptotic covariance matrix of $\hat{\theta}^w = g(\hat{\beta}^w)$.

The above results show that the modified jackknife and the δ-method are asymptotically equivalent. Some finite sample empirical results are presented in the next section. Remarks about the comparisons of the modified jackknife and the δ-method are given in Section 5.

4. SIMULATION RESULTS

We report in this section some simulation results of the performances of the variance estimators obtained by the jackknife, the modified jackknife and the δ-method.

4.1. The models

(1) *Univariate quadratic regression.* The first model we considered is

$$y_{ij} = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_{ij}, \quad j=1,2,3,\ i=1,...,24, \tag{M1}$$

$$x_i = .3, .4, .5, .6, .7, .8, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 7, 8, 10, 12, 14, 16, 18.$$

The errors e_{ij} are independently distributed as $N(0,\sigma_i^2)$. Two variance patterns under consideration are: (i) $\sigma_i^2 = x_i/2$ and (ii) σ_i^2 are not related to the design and given by:

$$\sigma_i^2 = .21, .99, .60, .91, .97, .35, .38, .10, .36, .70, .99, .72, \tag{4.1}$$

$$.24, .74, .35, .74, .45, .76, .19, .42, .11, .96, .62, .39.$$

(2) *Multiple regression.* Two models are considered. A simple one is

$$y_{ij} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_{ij}, \quad j=1,2,3,\ i=1,...,24, \tag{M2}$$

where the design values x_{1i} and x_{2i} are given by Neter, Wasserman and Whitmore (1988, p.712). The errors are independent and have variances σ_i^2 given by (4.1). Both normal and non-normal errors are considered. For the non-normal situation, the distribution of $\sigma_i^{-1} e_{ij}$ is (i) a uniform distribution on $[-3^{1/2}, 3^{1/2}]$; (ii) a double exponential distribution with shape parameter $2^{1/2}$ and (iii) a distribution $F(t)$ with density

$$F'(t) = 3^{-1/3}\Gamma^{-1}(1/3)\,|t|^{-1/3}\exp(-t^2/3). \tag{4.2}$$

A more complex model is from Gunst and Mason (1980):

$$y_{ij} = \beta_0 + \textstyle\sum_{l=1}^{8} \beta_l x_{li} + e_{ij}, \quad j=1,...,n_i,\ i=1,...,40. \tag{M3}$$

The errors are independently normal. We consider both equal replication case (n_i=3 for all i) and unequal replication case. The n_i for unequal replication case, the error variances σ_i^2 and the values of x_{li} are given by Table 3(b).

(3) *Response surface.* We consider a bivariate third order polynomial response surface:

$$y_{ij} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \beta_4 x_{1i}^2 + \beta_5 x_{2i}^2 \qquad \text{(M4)}$$

$$+ \beta_6 x_{1i}^2 x_{2i} + \beta_7 x_{1i} x_{2i}^2 + \beta_8 x_{1i}^3 + \beta_9 x_{2i}^3 + e_{ij}, \quad j=1,\ldots,n_i,\ i=1,\ldots,24.$$

The errors are independently normal with $\sigma_i^2=(x_{1i}^2 + x_{2i}^2)^{1/2}/2$. Both equal replication ($n_i \equiv 4$) and unequal replication cases are considered. The values of n_i (for unequal replication case), x_{1i} and x_{2i} are given in Table 4(b).

4.2. The estimators

For the above four models, we consider the estimation of the asymptotic variance of each component of $\hat{\beta}^w$. The root mean squared errors (rmse) and biases of the variance estimators obtained by the jackknife, the modified jackknife and the δ-method are reported in Tables 1-4. These values are independent of the parameter β. The asymptotic variances are also given.

For models (M1) and (M2), we also study the rmse and biases of the estimators of the covariances of $\hat{\beta}^w$ and the asymptotic variances of $\hat{\theta}^w=g(\hat{\beta}^w)$ (nonlinear case), where $g(\beta)$ is equal to (i) $\|\beta\|$; (ii) $\beta_{max}=-\beta_1/2\beta_2$, which maximizes the quadratic function $\beta_0+\beta_1 x+\beta_2 x^2$ over x. The rmse, the biases, the asymptotic variances and the values of β under consideration (in nonlinear cases) are included in Tables 1 and 2.

All the results are based on 3000 simulations on a VAX 11/780 at the Purdue University.

4.3. Summary of the simulation results

(1) *The jackknife.* The rmse and biases of $\hat{V}_J^g$ are large. $\hat{V}_J^g$ is severely downward-biased. The relative bias of $\hat{V}_J^g$ can be as large as 77.6%.

(2) *The modified jackknife.* The modified jackknife generally improves the original jackknife and performs well except in the cases of estimating $Cov(\hat{\beta}_2^w, \hat{\beta}_3^w)$ and $Var\,\hat{\beta}_3^w$ under model (M1) with $\sigma_i^2= x_i/2$ and $Var\,\hat{\beta}_1^w$ under model (M3).

(3) *Comparisons of the modified jackknife and the δ-method (normal errors).* The modified jackknife out-performs the δ-method under models (M2), (M3), (M4) and model (M1) with σ_i^2 given by (4.1). For model (M1) with $\sigma_i^2=x_i/2$, no definite conclusion can be made. For the nonlinear situation, the modified jackknife estimators are better especially when β is close to a discontinuity point of ∇g (The improvement in rmse can be as high as 30%.). A discussion of this phenomenon is given in Section 5(ii).

(4) *Non-normal errors.* Table 2(b) contains the results for model (M2) with non-normal errors. The modified jackknife performs well in all three cases and is clearly better than the δ-method. The improvement in rmse can be as high as 47%. Note that the error distribution given by (4.2) does not satisfy Assumption (c) (see Shao (1988a)). Therefore, the performance of the modified jackknife is less susceptible to violations of the assumptions for the asymptotic theory.

(5) *The number of replicates.* Except for the computational complexity, the effect of unequal n_i is not appreciable. As expected, the variance estimators perform better when there are larger number of replicates.

5. CONCLUDING REMARKS

(i) The ordinary jackknife estimator $\hat{V}_J^g$ is not recommended in view of its both small sample performance and large sample property.

(ii) The modified jackknife and the δ-method are asymptotically equivalent and provide consistent variance estimators. In terms of the finite sample performances, the results in Section 4 show that for the models under consideration, the modified jackknife is preferred. A general finite sample theory is not available. However, one may decide whether to use the modified jackknife or the δ-method by considering the following aspects.

(a) *The discontinuity of* ∇g. When the true parameter β is close to a discontinuity point of ∇g, $\nabla g(\hat{\beta}^w)$ is highly fluctuating even if $\hat{\beta}^w$ is an efficient estimator of β, since a small difference between $\hat{\beta}^w$ and β may result in a large difference between $\nabla g(\hat{\beta}^w)$ and $\nabla g(\beta)$. In this situation, the δ-method may not provide an accurate variance estimator and therefore the modified jackknife is recommended.

Section 4 provides two examples: (1) $g(\beta)=\|\beta\|$ and $\nabla g(\beta)$ is discontinuous at $\beta=0$ and (2) $g(\beta)=\beta_{max}$ and $\nabla g(\beta)$ is discontinuous at $\beta_2=0$. This explains why the modified jackknife out-performs the δ-method when $\|\beta\|$ or $|\beta_2|$ is small.

(b) *The computational cost.* Because of equation (2.7), the modified jackknife does not involve a large number of computations, if the function g can be easily evaluated. If the evaluation of g is expensive, then the δ-method is preferred. On the other hand, the modified jackknife is preferred if the evaluation of ∇g is more expensive. Thus, one may select a variance estimator which is computationally cheaper.

(c) *The derivation of* ∇g. The use of the δ-method requires a theoretical derivation of ∇g, which may be very complicated. The jackknife avoids this theoretical derivation. In some situations g does not have an explicit form and therefore ∇g can not be directly evaluated. For example, $y_{ij}=\sum_{t=0}^{p}\beta_t x_i^t+e_{ij}$, $p\geq 4$ and $g(\beta)=\beta_{\max}$, a point maximizes $\sum_{t=0}^{p}\beta_t x^t$ over x. Therefore, the modified jackknife is preferred in this case.

(d) *Robustness.* Tukey and subsequent workers recognized the distribution robustness of the jackknife method. This is supported by our empirical results in Section 4 (see Summary (4)). In addition, since the modified jackknife does not depend on ∇g explicitly, its performance is less susceptible to violation of the smoothness assumption on the function g.

(iii) It is known that when the estimators of σ_i^2 are not consistent, the WLSE may not be more efficient than the OLSE. A comparison of the OLSE $\hat{\theta}=g(\hat{\beta})$ and the WLSE $\hat{\theta}^w=g(\hat{\beta}^w)$ can be carried out by estimating the relative efficiency $Var\,\hat{\theta}^w/Var\,\hat{\theta}$ (assuming g is real-valued) by a consistent estimator $\hat{V}_M^g/\hat{U}_J^g$.

(iv) Alternative estimators of σ_i^2 can be found in Hartley, Rao and Kiefer (1969), Rao (1970), Horn, Horn and Duncan (1975) and Shao (1988a). Let

$$\tilde{v}_i = v_i + \delta_i$$

with δ_i satisfying $\max_{i\leq k}|\delta_i|=O_p(k^{-1})$, where v_i is defined in (1.2). Then all the results in Section 3 still hold if the weights $w_i=v_i^{-1}$ is replaced by $w_i=\tilde{v}_i^{-1}$. An example is (Shao, 1988a)

$$\tilde{v}_i = v_i + c_i s^2,$$

where $c_i=x_i'(X'X)^{-1}x_i$ and $s^2=(n-p)^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}r_{ij}^2$ is the usual estimator of σ^2 when $\sigma_i^2=\sigma^2$ for all i.

ACKNOWLEDGEMENTS

I wish to thank the referees and the Editor for their helpful suggestions and comments.

APPENDIX

Proof of Lemma 1. Let $\delta>0$ be given in Assumption (c). Then

$$[k^{-1}E(\max_{i\le k}u_i^{-1})]^{1+\delta} \le k^{-(1+\delta)}E(\max_{i\le k}u_i^{-(1+\delta)}) \le k^{-(1+\delta)}\sum_{i=1}^{k}Eu_i^{-(1+\delta)}$$

$$\le n_\infty^{1+\delta}k^{-(1+\delta)}\sum_{i=1}^{k}E(\sum_{j=1}^{n_0}e_{ij}^2)^{-(1+\delta)} \le bn_\infty^{1+\delta}k^{-\delta} \to 0. \ \square$$

Proof of Lemma 2. Note that $\Delta_i=u_i^{-1}|[x_i'(\hat\beta-\beta)]^2-2x_i'(\hat\beta-\beta)(n_i^{-1}\sum_{j=1}^{n_i}e_{ij})|$ and

$$\max_{i\le k}[x_i'(\hat\beta-\beta)]^2u_i^{-1} \le c_\infty^2\|\hat\beta-\beta\|^2\max_{i\le k}u_i^{-1}.$$

The result follows from Lemma 1, $\|\hat\beta-\beta\|^2=O_p(k^{-1})$ and the Cauchy-Schwarz inequality. $\square$

Proof of Lemma 3. From Assumptions (a)-(c), there is a positive constant c such that

$$k^{-1}\max_{i\le k}z_i'z_i \le ck^{-1}\max_{i\le k}u_i^{-1} \to_p 0. \tag{A1}$$

From Proposition 1, $\|\hat\beta^w-\beta\|^2=O_p(k^{-1})$. Then the first assertion follows from (A1) and

$$\max_{i\le k}[z_i'(\hat\beta^w-\beta)]^2 \le \|\hat\beta^w-\beta\|^2\max_{i\le k}z_i'z_i.$$

By Lemma 4 of Shao (1988a), $k(Z'Z)^{-1}-k(X'D_1X)^{-1}\to_p 0$. Then by Assumptions (a) and (b), there is a positive constant a such that

$$P\{\, k(Z'Z)^{-1} > a\, I_p \,\} \to 0. \tag{A2}$$

On the set $\{\, k(Z'Z)^{-1} \le a\, I_p \,\}$, $\max_{i\le k}z_i'(Z'Z)^{-1}z_i \le ak^{-1}\max_{i\le k}z_i'z_i$. Then the second assertion follows from (A1) and (A2). $\square$

Proof of Lemma 4. From (2.7),

$$\max_{j\le n_i,i\le k}\|\hat\beta^w_{(i,j)}-\hat\beta^w\|^2 \le (1-\max_{i\le k}h_i)^{-2}(\max_{i\le k}z_i'(Z'Z)^{-2}z_i)\max_{j\le n_i,i\le k}r_{ij}^2.$$

On the set $\{\, k(Z'Z)^{-1}\le a\, I_p \,\}$, $k\max_{i\le k}z_i'(Z'Z)^{-2}z_i \le a\max_{i\le k}z_i'(Z'Z)^{-1}z_i$. Hence the result follows from (A2), Lemma 3 and

$$\max_{j\le n_i,i\le k}r_{ij}^2 \le 2\sum_{i=1}^{k}\sum_{j=1}^{n_i}e_{ij}^2 + 2\max_{i\le k}[x_i'(\hat\beta-\beta)]^2 = O_p(k). \ \square$$

REFERENCES

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

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist. 7*, 1-26.

_______ (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans.* SIAM, Philadelphia.

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.

Gunst, R. F. and Mason R. L. (1980). *Regression Analysis and Its Application.* Marcel Dekker, New York.

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.

Hinkley, D. V. (1977). Jackknifing in unbalanced situations. *Technometrics 19*, 285-292.

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

McCarthy, P. J. (1969). Pseudo-replication: half-samples. *Rev. ISI 37*, 239-263.

Miller, R. G. (1974). An unbalanced jackknife. *Ann. Statist. 2*, 880-891.

Muller, H.-G. and Stadtmuller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist. 15*, 610-625.

Neter, J., Wasserman, W. and Whitmore, G. A. (1988). *Applied Statistics*, 3rd edition. Allyn and Bacon.

Parr, W. C. (1985). Jackknifing differentiable statistical functionals. *J. R. Statist. Soc. B 47*, 56-66.

Quenouille, M. (1956). Notes on bias in estimation. *Biometrika 43*, 353-360.

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

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

_______(1988b). On resampling methods for variance and bias estimation in linear models. *Ann. Statist. 16*, to appear.

Shao, J. and Wu, C. F. J. (1986). Some general theory for the jackknife. Tech. Report #797, Dept. of Statistics, University of Wisconsin, Madison.

__________________(1987). Heteroscedasticity-robustness of jackknife variance estimators in linear models. *Ann. Statist. 15*, 1563-1579.

Tukey, J. (1958). Bias and confidence in not quite large samples. *Ann. Math. Statist. 29*, 614.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist. 14*, 1261-1350.

Table 1(a): The rmse and biases of $n\hat{V}_J^g$, $n\hat{V}_M^g$ and $n\hat{V}_\delta^g$ (model (M1) with σ_i^2 given by (4.1)).

	Linear		case	nV =		1.9428	−.5187	.0257	
							.2273	−.0131	
								.0008	
	$\hat{V}_J$			$\hat{V}_M$			$\hat{V}_\delta$		
rmse	1.2827	.3323	.0161	.7426	.1853	.0098	1.0030	.2670	.0138
		.1415	.0080		.0767	.0050		.1126	.0069
			.0005			.0004			.0005
bias	−1.2472	.3235	−.0156	−.1540	.0166	.0004	.5848	−.1701	.0089
		−.1380	.0077		.0026	−.0008		.0772	−.0046
			−.0005			.0001			.0003
			Nonlinear	case	(β_0=−.1)				
	$g=\lVert\beta\rVert$	β_1=.04 nV^g=2.0640	β_2=0	$g=\lVert\beta\rVert$	β_1=.04 nV^g=1.6762	β_2=.05	$g=\lVert\beta\rVert$	β_1=0 nV^g=1.5339	β_2=.05
	$\hat{V}_J^g$	$\hat{V}_M^g$	$\hat{V}_\delta^g$	$\hat{V}_J^g$	$\hat{V}_M^g$	$\hat{V}_\delta^g$	$\hat{V}_J^g$	$\hat{V}_M^g$	$\hat{V}_\delta^g$
rmse	1.4577	.8960	1.0420	1.1905	.8610	1.0795	1.0701	.8330	1.1121
bias	−1.4203	−.4086	.2924	−1.1443	−.3067	.2512	−1.0181	−.2050	.3368
	$g=\lVert\beta\rVert$	β_1=4 nV^g=.2538	β_2=−.5	$g=\beta_{max}$	β_1=4 nV^g=.0709	β_2=−.5	$g=\beta_{max}$	β_1=4 nV^g=.0340	β_2=−1
	$\hat{V}_J^g$	$\hat{V}_M^g$	$\hat{V}_\delta^g$	$\hat{V}_J^g$	$\hat{V}_M^g$	$\hat{V}_\delta^g$	$\hat{V}_J^g$	$\hat{V}_M^g$	$\hat{V}_\delta^g$
rmse	.1590	.0970	.1388	.0458	.0234	.0341	.0217	.0137	.0165
bias	−.1540	.0030	.0869	−.0447	−.0038	.0229	−.0210	−.0004	.0113

Table 1(b): The rmse and biases of $n\hat{V}_J^g$, $n\hat{V}_M^g$ and $n\hat{V}_\delta^g$ (model (M1) with $\sigma_i^2 = x_i/2$).

	Linear		case	nV =		2.7629	−1.4191	.0876	
							1.2545	−.0850	
								.0067	
	$\hat{V}_J$			$\hat{V}_M$			$\hat{V}_\delta$		
rmse	2.0269	1.0471	.0642	.9725	.4695	.0402	1.1870	.5517	.0376
		.8680	.0575		.6469	.0625		.4936	.0372
			.0044			.0075			.0034
bias	−1.9962	1.0368	−.0635	−.3873	−.0203	.0125	.6528	−.2679	.0154
		−.8564	.0566		.3648	−.0397		.2423	−.0170
			−.0042			.0051			.0015
			Nonlinear	case	(β_0=−.1)				
	$g=\lVert\beta\rVert$	β_1=.04 nV^g=3.5335	β_2=0	$g=\lVert\beta\rVert$	β_1=.04 nV^g=2.8219	β_2=.05	$g=\lVert\beta\rVert$	β_1=0 nV^g=2.1416	β_2=.05
	$\hat{V}_J^g$	$\hat{V}_M^g$	$\hat{V}_\delta^g$	$\hat{V}_J^g$	$\hat{V}_M^g$	$\hat{V}_\delta^g$	$\hat{V}_J^g$	$\hat{V}_M^g$	$\hat{V}_\delta^g$
rmse	2.7743	1.4626	1.5522	2.1722	1.2867	1.5780	1.5339	1.2178	1.7393
bias	−2.7425	−.8194	−.0869	−2.1324	−.4355	.1748	−1.4775	.1551	.7420
	$g=\lVert\beta\rVert$	β_1=4 nV^g=1.3275	β_2=−.5	$g=\beta_{max}$	β_1=4 nV^g=.3219	β_2=−.5	$g=\beta_{max}$	β_1=4 nV^g=.1703	β_2=−1
	$\hat{V}_J^g$	$\hat{V}_M^g$	$\hat{V}_\delta^g$	$\hat{V}_J^g$	$\hat{V}_M^g$	$\hat{V}_\delta^g$	$\hat{V}_J^g$	$\hat{V}_M^g$	$\hat{V}_\delta^g$
rmse	.9231	.6770	.5349	.2253	.1357	.1296	.1194	.0724	.0656
bias	−.9107	.3623	.2499	−.2222	.0591	.0672	−.1178	.0326	.0328

Table 2(a): The rmse and biases of $n\hat{V}_J^g$, $n\hat{V}_M^g$ and $n\hat{V}_\delta^g$ (model (M2) with normal errors).

Linear case $nV =$

2.5086	−.0627	.1119
	.0032	−.0139
		.0959

	$\hat{V}_J$			$\hat{V}_M$			$\hat{V}_\delta$		
rmse	1.6419	.0402	.0798	.9402	.0247	.1023	1.3238	.0336	.1118
		.0020	.0085		.0014	.0069		.0018	.0085
			.0583			.0399			.0534
bias	−1.5969	.0389	−.0682	−.1709	.0002	.0108	.7919	−.0187	.0233
		−.0019	.0080		.0003	−.0015		.0010	−.0046
			−.0561			.0079			.0330

Nonlinear case ($g=\|\beta\|$, $\beta_0=-.1$)

	$\beta_1=.04$, $\beta_2=.05$, $nV^g=1.7489$			$\beta_1=.04$, $\beta_2=2.5$, $nV^g=.0904$			$\beta_1=.4$, $\beta_2=5$, $nV^g=.0899$		
	$\hat{V}_J^g$	$\hat{V}_M^g$	$\hat{V}_\delta^g$	$\hat{V}_J^g$	$\hat{V}_M^g$	$\hat{V}_\delta^g$	$\hat{V}_J^g$	$\hat{V}_M^g$	$\hat{V}_\delta^g$
rmse	1.1956	.9591	1.3671	.0516	.0483	.0701	.0535	.0396	.5450
bias	−1.1310	−.1630	.5010	−.0483	.0186	.0487	−.0512	.0104	.0358

Table 2(b): The rmse and biases of $n\hat{V}_J$, $n\hat{V}_M$ and $n\hat{V}_\delta$ (model (M2) with non-normal errors).

(i) Uniform

	$nVar\,\beta_0^w=2.9912$			$nVar\,\beta_1^w=.0039$			$nVar\,\beta_2^w=.1180$		
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$
rmse	1.8604	.9398	1.3162	.0023	.0015	.0018	.0690	.0423	.0532
bias	−1.8187	−.0156	.8696	−.0022	.0004	.0011	−.0670	.0136	.0349

(ii) Double exponential

	$nVar\,\beta_0^w=1.5947$			$nVar\,\beta_1^w=.0021$			$nVar\,\beta_2^w=.0651$		
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$
rmse	1.0382	.8073	1.5222	.0013	.0012	.0019	.0403	.0345	.0549
bias	−.9839	−.0032	1.0501	−.0012	.0002	.0012	−.0380	.0066	.0361

(iii) $F(t)$ given by (4.2)

	$nVar\,\beta_0^w=1.7859$			$nVar\,\beta_1^w=.0022$			$nVar\,\beta_2^w=.0671$		
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$
rmse	1.1572	.8671	1.4430	.0014	.0014	.0019	.0402	.0398	.0575
bias	−1.1005	−.0159	1.0005	−.0013	.0003	.0012	−.0371	.0120	.0399

Table 3(a): The rmse and biases of $\hat{V}_J$, $\hat{V}_M$ and $\hat{V}_\delta$ (model (M3)).

Equal replication									
	$Var\,\beta_0^w$=.0101			$Var\,\beta_1^w$=.0046			$Var\,\beta_2^w$= .0034		
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$
rmse	.0064	.0026	.0038	.0027	.0057	.0047	.0019	.0019	.0021
bias	−.0063	−.0003	.0025	−.0016	.0003	.0002	−.0017	.0008	.0012
	$Var\,\beta_3^w$=.0011			$Var\,\beta_4^w$=.0477			$Var\,\beta_5^w$= .8290		
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$
rmse	.0006	.0006	.0006	.0280	.0170	.0229	.4569	.4461	.4983
bias	−.0005	.0003	.0004	−.0271	.0049	.0151	−.4258	.1879	.3036
	$Var\,\beta_6^w$=.2864			$Var\,\beta_7^w$=.1531			$Var\,\beta_8^w$= .0431		
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$
rmse	.1696	.1208	.1605	.0921	.0471	.0672	.0235	.0206	.0267
bias	−.1629	.0299	.0854	−.0901	.0072	.0455	−.0223	.0096	.0191
Unequal replication									
	$Var\,\beta_0^w$=.0092			$Var\,\beta_1^w$=.0045			$Var\,\beta_2^w$= .0030		
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$
rmse	.0058	.0023	.0030	.0026	.0052	.0045	.0017	.0013	.0016
bias	−.0057	−.0008	.0019	−.0016	.0003	.0002	−.0016	.0003	.0008
	$Var\,\beta_3^w$=.0010			$Var\,\beta_4^w$=.0443			$Var\,\beta_5^w$= .8040		
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$
rmse	.0006	.0004	.0005	.0263	.0135	.0191	.4456	.3950	.4615
bias	−.0005	.0001	.0003	−.0258	.0013	.0122	−.4184	.1443	.2800
	$Var\,\beta_6^w$=.2793			$Var\,\beta_7^w$=.1401			$Var\,\beta_8^w$= .0380		
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$
rmse	.1663	.1083	.1494	.0855	.0379	.0547	.0215	.0148	.0197
bias	−.1604	.0172	.0797	−.0839	−.0033	.0352	−.0206	.0042	.0134

Table 3(b): The values of x_{li}, σ_i^2 and n_i for unequal replication case (model (M3)).

i	x_{1i}	x_{2i}	x_{3i}	x_{4i}	x_{5i}	x_{6i}	x_{7i}	x_{8i}	σ_i^2	n_i
1	0.1020	0.6900	1.3300	1.2500	0.3600	0.5300	1.0600	0.5326	0.21	3
2	1.2200	7.2300	26.1200	9.5300	1.3200	2.5200	5.7400	3.6138	0.99	3
3	0.1390	1.3800	0.4600	0.3500	0.0600	0.0900	0.2700	0.2594	0.60	4
4	0.2210	6.3700	1.5300	1.1500	0.1600	0.4100	0.8300	1.0346	0.91	4
5	0.0120	0.	0.0100	0.0900	0.0100	0.0200	0.0700	0.0381	0.97	3
6	0.0010	0.5000	0.0300	0.2500	0.0200	0.0700	0.0700	0.3440	0.35	3
7	1.0460	1.2700	3.1300	3.9200	0.5600	0.6200	2.1200	1.4559	0.38	3
8	2.0320	0.4400	4.0900	5.4000	0.9800	1.0600	2.8900	4.0182	0.10	5
9	0.8950	0.5400	1.6800	1.1700	0.3200	0.2000	0.7600	0.4600	0.36	3
10	0.	0.	0.0200	0.	0.0100	0.	0.0700	0.1540	0.70	3
11	0.0250	0.0200	0.2400	0.7800	0.1500	0.2500	0.5000	0.6516	0.99	4
12	0.0970	0.1200	0.9100	1.3500	0.2400	0.2800	0.5900	0.6011	0.72	4
13	0.0010	0.	0.1500	0.4600	0.1100	0.3500	0.4000	0.1922	0.24	3
14	0.0040	0.0100	0.1800	0.2300	0.0800	0.1300	0.2800	0.0931	0.74	3
15	0.0420	0.0400	0.7800	0.4100	0.6100	0.8500	0.4900	0.0538	0.35	3
16	0.0870	1.6200	5.9900	0.1100	0.0300	0.0300	0.2300	0.0199	0.74	5
17	0.0020	0.	0.2600	0.2400	0.0600	0.1100	0.5000	0.0419	0.45	3
18	0.0020	0.0900	0.2900	0.1100	0.0200	0.0800	0.2500	0.1093	0.76	3
19	0.0480	0.1800	1.0100	0.2500	0.0400	0.2400	0.0800	0.0328	0.19	3
20	0.1310	1.2600	3.8700	0.0600	0.	0.0200	0.0400	0.0797	0.42	3
21	0.0040	0.	1.0300	0.4900	0.0900	0.1800	0.5900	0.1855	0.11	4
22	0.0010	0.0400	0.4600	0.1600	0.0200	0.1600	0.2400	0.1572	0.96	3
23	0.	0.	4.6800	0.5600	0.0200	0.1100	0.2100	0.0998	0.62	3
24	0.0070	0.	0.5200	0.3700	0.0500	0.2400	0.4300	0.2804	0.39	4
25	0.0050	0.0100	0.0600	0.9500	0.1100	0.3900	0.2900	0.2879	0.55	3
26	0.1740	1.1300	6.8500	0.6900	0.1800	0.1100	0.4300	0.6810	0.57	3
27	0.	0.	0.0600	0.3500	0.0400	0.0900	0.2300	0.3242	0.39	3
28	0.2330	1.5300	6.8200	4.0400	0.8500	1.3300	2.7000	2.6013	0.41	3
29	0.1550	0.5600	0.9400	0.7500	0.1700	0.3200	0.6600	0.4469	0.95	5
30	0.1200	0.7400	0.5500	1.2000	0.0800	0.1200	0.4900	0.2436	0.82	5
31	8.9830	0.3700	2.3600	0.7700	0.3800	0.1800	0.4900	0.4400	0.71	3
32	0.0590	0.5400	1.3800	0.5500	0.1100	0.1300	0.1800	0.3351	0.53	3
33	0.0720	1.1200	1.6900	2.2800	0.3900	0.3800	0.9900	1.3979	0.25	3
34	0.5710	0.7800	2.5400	1.6200	0.4300	0.4600	1.4700	2.0138	0.76	3
35	0.8530	10.0200	10.1700	4.1800	0.5700	1.1600	1.8200	1.9356	0.63	3
36	0.0050	0.	0.1700	0.1400	0.1300	0.0300	0.0800	0.1050	0.74	3
37	0.0110	0.3400	0.0300	0.2000	0.0400	0.0500	0.1400	0.2207	0.54	4
38	0.2580	0.0100	0.3300	0.4800	0.1300	0.1800	0.2800	0.0810	0.83	3
39	0.0690	0.1400	1.2600	1.0800	0.2000	0.9500	0.4100	0.1017	0.91	3
40	4.7900	20.4600	37.1900	0.3100	0.0700	0.0600	0.1800	0.0962	0.52	3

Table 4(a): The rmse and biases of $\hat{V}_J$, $\hat{V}_M$ and $\hat{V}_\delta$ (model (M4)).

Equal replication									
	$Var\beta_0^w$=.0901			$Var\beta_1^w$=.1255			$Var\beta_2^w$= 1.6701		
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$
rmse	.0522	.0424	.0505	.0734	.0415	.0529	.9747	.5247	.6666
bias	−.0496	.0137	.0309	−.0717	.0140	.0367	−.9550	.1993	.4620
	$Var\beta_3^w$=.5173			$Var\beta_4^w$=.1152			$Var\beta_5^w$= 2.4046		
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$
rmse	.2969	.1771	.1946	.0656	.0402	.0477	1.4095	.6996	.8726
bias	−.2909	.0911	.1291	−.0641	.0186	.0334	−1.3872	.2950	.6067
	$Var\beta_6^w$=.1449			$Var\beta_7^w$=.1271			$Var\beta_8^w$= .1146		
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$
rmse	.0824	.0549	.0537	.0726	.0476	.0477	.0678	.0301	.0412
bias	−.0807	.0318	.0348	−.0710	.0259	.0304	−.0668	.0106	.0301
	$Var\beta_9^w$=.2572								
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$						
rmse	.1505	.0749	.0899						
bias	−.1483	.0351	.0622						
Unequal replication									
	$Var\beta_0^w$=.0954			$Var\beta_1^w$=.1494			$Var\beta_2^w$= 1.8742		
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$
rmse	.0547	.0440	.0602	.0851	.0616	.0825	1.0613	.5977	.8941
bias	−.0512	.0100	.0383	−.0814	.0194	.0586	−1.0322	.1936	.6590
	$Var\beta_3^w$=.6210			$Var\beta_4^w$=.1375			$Var\beta_5^w$= 2.7555		
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$
rmse	.3382	.2451	.3110	.0740	.0578	.0754	1.5610	.8612	1.2697
bias	−.3265	.1267	.2245	−.0709	.0288	.0558	−1.5230	.3174	.9310
	$Var\beta_6^w$=.1732			$Var\beta_7^w$=.1468			$Var\beta_8^w$= .1363		
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$
rmse	.0931	.0754	.0867	.0804	.0562	.0698	.0778	.0426	.0645
bias	−.0895	.0421	.0603	−.0778	.0281	.0488	−.0759	.0143	.0491
	$Var\beta_9^w$=.2980								
	$\hat{V}_J$	$\hat{V}_M$	$\hat{V}_\delta$						
rmse	.1680	.0951	.1359						
bias	−.1639	.0397	.0991						

Table 4(b): The values of x_{1i}, x_{2i} and n_i for unequal replication case (model (M4)).

i	1	2	3	4	5	6	7	8	9	10	11	12
x_{1i}	−1	−1	−1	−1	−1	−1	−.5	−.5	−.5	−.5	−.5	−.5
x_{2i}	.1	.3	.5	1	1.5	2	.1	.3	.5	1	1.5	2
n_i	3	3	4	3	3	4	5	3	4	3	3	3
i	13	14	15	16	17	18	19	20	21	22	23	24
x_{1i}	.5	.5	.5	.5	.5	.5	1	1	1	1	1	1
x_{2i}	.1	.3	.5	1	1.5	2	.1	.3	.5	1	1.5	2
n_i	3	3	3	4	4	3	3	3	5	3	3	4