BIOMETRICS 57, 126–134
March 2001

A Covariance Estimator for GEE with Improved Small-Sample Properties

Lloyd A. Mancl[1,*] and Timothy A. DeRouen[1,2]

Departments of [1]Dental Public Health Sciences and [2]Biostatistics,
University of Washington, Box 357475, Seattle, Washington 98195, U.S.A.
*email: lman@biostat.washington.edu

SUMMARY. In this paper, we propose an alternative covariance estimator to the robust covariance estimator of generalized estimating equations (GEE). Hypothesis tests using the robust covariance estimator can have inflated size when the number of independent clusters is small. Resampling methods, such as the jackknife and bootstrap, have been suggested for covariance estimation when the number of clusters is small. A drawback of the resampling methods when the response is binary is that the methods can break down when the number of subjects is small due to zero or near-zero cell counts caused by resampling. We propose a bias-corrected covariance estimator that avoids this problem. In a small simulation study, we compare the bias-corrected covariance estimator to the robust and jackknife covariance estimators for binary responses for situations involving 10–40 subjects with equal and unequal cluster sizes of 16–64 observations. The bias-corrected covariance estimator gave tests with sizes close to the nominal level even when the number of subjects was 10 and cluster sizes were unequal, whereas the robust and jackknife covariance estimators gave tests with sizes that could be 2–3 times the nominal level. The methods are illustrated using data from a randomized clinical trial on treatment for bone loss in subjects with periodontal disease.

KEY WORDS: Correlated binary data; Logistic regression; Periodontal disease; Simulation study.

1. Introduction

Correlated observations are common to many biomedical applications due to multiple observations of a response from an individual. The generalized estimating equation methodology (GEE) has gained considerable popularity since its introduction (Liang and Zeger, 1986) as a regression method for correlated responses. GEE uses an empirical or robust covariance estimator of sandwich form to estimate the covariance matrix of the regression coefficients (Liang and Zeger, 1986). The estimator is robust with respect to misspecification of the covariance matrix of the correlated responses, which is typically unknown, and allows valid inference provided the number of individuals is sufficiently large. However, the robust estimator may be biased when the number of subjects is small since ordinary residuals (observed value minus fitted value) are used to estimate the unknown covariance matrix. The data example we focus on is a randomized clinical trial on radiographic bone loss in periodontal disease, in which measurements of bone loss and probing depth were taken at two locations on each tooth for 30 subjects. Because of the time and expense involved in obtaining the measurements on a subject, dental studies often involve relatively small numbers of subjects with a moderate to large number of observations per subject.

Several authors have studied the performance of GEE for binary responses for sample sizes involving 10–50 subjects (Lipsitz, Laird, and Harrington, 1990; Emrich and Piedmonte, 1992; Sharples and Breslow, 1992; Lipsitz et al., 1994; Qu, Piedmonte, and Williams, 1994; Gunsolley, Getchell, and Chinchilli, 1995; Sherman and le Cessie, 1997). In addition, the small-sample performance of GEE has been investigated by Paik (1988) for first-order autoregressive gamma responses and Feng, McLerran, and Grizzle (1996) for multivariate normal responses. The simulation studies have shown that the robust estimator tends to underestimate the variance of regression coefficients to a varying degree when the number of subjects is less than 50 and that hypothesis tests and confidence intervals using the robust estimates are too liberal. Paik (1988), Lipsitz et al. (1990), and Qu et al. (1994) studied various jackknife estimators, and Sherman and le Cessie (1997) studied several bootstrap estimators as alternative covariance estimators to the GEE robust estimator, which are also robust to misspecification of the covariance matrix. Both jackknife and bootstrap methods tend to perform better than GEE with the robust covariance estimator with respect to maintaining the nominal test size except when the number of subjects is 20 or less and the response is binary. The poor performance of the resampling methods is probably related to zero or small cell counts caused by resampling (Lipsitz et al., 1990; Sherman and le Cessie, 1997).

In this paper, we propose and evaluate a bias-corrected covariance estimator that avoids the small cell count problem of the resampling methods. In Section 2, we briefly review the GEE methodology. In Section 3, we derive a bias-corrected covariance estimator. In Section 4, we give alternative ap-

proaches to covariance estimation, and in Section 5 we illustrate the methods with a data example on radiographic bone loss. In Section 6, we present the results of our simulation study on test-size accuracy, and we provide a discussion in Section 7.

2. Generalized Estimating Equations

The data consists of correlated observations $\{y_{ij}, x_{ij}\}$, $j = 1, 2, \ldots, n_i$, for each of the $i = 1, 2, \ldots, K$ subjects, where y_{ij} is the response measure and x_{ij} is a $p \times 1$ vector of covariates. Of scientific interest is the relationship of the covariates with the response mean, whereas the correlation between observations is considered a nuisance. The mean $\mu_{ij} = \mathrm{E}(y_{ij} \mid x_{ij})$ is related to x_{ij} by $g(\mu_{ij}(\boldsymbol{\beta})) = x_{ij}^{\mathrm{T}}\boldsymbol{\beta}$, where g is a known link function and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients. The variance of y_{ij} is given by $\mathrm{var}(y_{ij}) = \phi \cdot h(\mu_{ij})$, where h is a known function of μ_{ij} and ϕ is a possibly unknown scale parameter. Let $\mathbf{A}_i = \mathrm{diag}\{h(\mu_{i1}), \ldots, h(\mu_{in_i})\}$ and $\mathbf{V}_i = \phi \mathbf{A}_i^{1/2} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}$ define the working covariance matrix for $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^{\mathrm{T}}$, where $\mathbf{R}_i(\boldsymbol{\alpha})$ is the $n_i \times n_i$ working correlation matrix for $\mathbf{y}_i$. $\mathbf{R}_i$ is assumed to be fully specified by an unknown parameter vector, $\boldsymbol{\alpha}$, which is the same for all subjects. GEE estimates, $\hat{\boldsymbol{\beta}}$, are given by the solution to the estimating equations

$$\sum_{i=1}^{K} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = 0, \tag{1}$$

where $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{in_i})^{\mathrm{T}}$ and $\mathbf{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}^{\mathrm{T}}$ (Liang and Zeger, 1986). Liang and Zeger (1986) proposed a sandwich estimator to estimate the covariance matrix of $\hat{\boldsymbol{\beta}}$,

$$\left(\sum_{i=1}^{K} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathbf{D}_i \right)^{-1} \left(\sum_{i=1}^{K} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathrm{cov}(\mathbf{y}_i) \mathbf{V}_i^{-1} \mathbf{D}_i \right) \times \left(\sum_{i=1}^{K} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathbf{D}_i \right)^{-1}, \tag{2}$$

where $(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^{\mathrm{T}}$ is typically used to estimate $\mathrm{cov}(\mathbf{y}_i)$. This estimate is commonly called the robust covariance estimator since it is consistent even if the working covariance matrix is not the true covariance matrix of $\mathbf{y}_i$. The bias-corrected robust covariance estimator described in the next section is also robust to misspecification of the covariance matrix of $\mathbf{y}_i$. If $\mathbf{V}_i$ correctly specifies the covariance matrix of $\mathbf{y}_i$, then a consistent estimator for the covariance matrix of $\hat{\boldsymbol{\beta}}$ is given by $(\Sigma_{i=1}^{K} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathbf{D}_i)^{-1}$, which is called the model-based or naive covariance estimator.

3. Bias-Corrected Robust Covariance Estimator

In practice, to calculate the GEE robust covariance estimator, (2) the residuals, $\mathbf{r}_i = \mathbf{y}_i - \hat{\boldsymbol{\mu}}_i$, are used to estimate $\mathrm{cov}(\mathbf{y}_i)$. It is well known that the residuals tend to be too small (i.e., the fitted values tend to be closer to the observed values than the true values). Hence, the robust estimator would be expected to underestimate the covariance of $\hat{\boldsymbol{\beta}}$, and the bias of the robust estimator due to underestimation of $\mathrm{cov}(\mathbf{y}_i)$ would be large when the bias of the residuals is large (e.g., when the number of subjects is small).

An alternative robust covariance estimator for $\hat{\boldsymbol{\beta}}$ is proposed that is intended to reduce the bias of the residual estimator, $\mathbf{r}_i \mathbf{r}_i^{\mathrm{T}}$. To derive an approximation for the bias of $\mathbf{r}_i \mathbf{r}_i^{\mathrm{T}}$, consider a first-order Taylor series expansion of the residual vector, $\mathbf{r}_i$, about $\boldsymbol{\beta}$ given by

$$\mathbf{r}_i = \mathbf{e}_i + \frac{\partial \mathbf{e}_i}{\partial \boldsymbol{\beta}^{\mathrm{T}}} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \tag{3}$$

where $\mathbf{e}_i = \mathbf{e}_i(\boldsymbol{\beta}) = \mathbf{y}_i - \boldsymbol{\mu}_i$ and $i = 1, 2, \ldots, K$. By squaring (3) and taking the expectation, one obtains

$$\begin{aligned} \mathrm{E}\left[\mathbf{r}_i \mathbf{r}_i^{\mathrm{T}}\right] &= \mathrm{E}\left[\mathbf{e}_i \mathbf{e}_i^{\mathrm{T}}\right] + \mathrm{E}\left[\mathbf{e}_i (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\mathrm{T}} \frac{\partial \mathbf{e}_i^{\mathrm{T}}}{\partial \boldsymbol{\beta}}\right] \\ &\quad + \mathrm{E}\left[\frac{\partial \mathbf{e}_i}{\partial \boldsymbol{\beta}^{\mathrm{T}}} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \mathbf{e}_i^{\mathrm{T}}\right] \\ &\quad + \mathrm{E}\left[\frac{\partial \mathbf{e}_i}{\partial \boldsymbol{\beta}^{\mathrm{T}}} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\mathrm{T}} \frac{\partial \mathbf{e}_i^{\mathrm{T}}}{\partial \boldsymbol{\beta}}\right]. \end{aligned}$$

Using the first-order approximation

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \approx \left(\sum_{i=1}^{K} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathbf{D}_i \right)^{-1} \sum_{i=1}^{K} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathbf{e}_i,$$

the expectation is approximated by

$$\begin{aligned} \mathrm{E}[\mathbf{r}_i \mathbf{r}_i^{\mathrm{T}}] &\approx \mathrm{cov}[\mathbf{y}_i] - \mathrm{cov}[\mathbf{y}_i] \mathbf{H}_{ii}^{\mathrm{T}} - \mathbf{H}_{ii}\, \mathrm{cov}[\mathbf{y}_i] \\ &\quad + \sum_{j=1}^{K} \mathbf{H}_{ij} \mathrm{cov}[\mathbf{y}_i] \mathbf{H}_{ij}^{\mathrm{T}} \\ &= (\mathbf{I}_i - \mathbf{H}_{ii}) \mathrm{cov}[\mathbf{y}_i] (\mathbf{I}_i - \mathbf{H}_{ii})^{\mathrm{T}} \\ &\quad + \sum\nolimits_{j \neq i} \mathbf{H}_{ij} \mathrm{cov}[\mathbf{y}_i] \mathbf{H}_{ij}^{\mathrm{T}}, \end{aligned} \tag{4}$$

where $\mathbf{H}_{ij} = \mathbf{D}_i (\Sigma_{l=1}^{K} \mathbf{D}_l^{\mathrm{T}} \mathbf{V}_l^{-1} \mathbf{D}_l)^{-1} \mathbf{D}_j^{\mathrm{T}} \mathbf{V}_j^{-1}$, $\mathbf{I}_i$ is an identity matrix of the same dimension as $\mathbf{H}_{ii}$ and the summation $\Sigma_{j \neq i}$ is over all $j = 1, 2, \ldots, K \neq i$. In order to derive a tractable approximation to the bias, we assume that the contribution to the bias of the sum in expression (4) is negligible. By definition, the elements of $\mathbf{H}_{ij}$ are between zero and one, usually close to zero, so it may be reasonable to assume that the summation makes only a small contribution to the bias. The matrix $\mathbf{H}_{ii}$ is an expression for the leverage of the ith subject (Preisser and Qaqish, 1996).

Assuming the expected value of the residual estimator is approximated by

$$\mathrm{E}\left[\mathbf{r}_i \mathbf{r}_i^{\mathrm{T}}\right] \approx (\mathbf{I}_i - \mathbf{H}_{ii}) \mathrm{cov}\left[\mathbf{y}_i\right] (\mathbf{I}_i - \mathbf{H}_{ii})^{\mathrm{T}} \tag{5}$$

gives the bias-corrected covariance estimator

$$\begin{aligned} &\mathrm{var}_{\mathrm{bias-corrected}}(\hat{\boldsymbol{\beta}}) \\ &\quad = \mathbf{F}^{-1} \left\{ \sum_{i=1}^{K} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} (\mathbf{I}_i - \mathbf{H}_{ii})^{-1} \mathbf{r}_i \mathbf{r}_i^{\mathrm{T}} \right. \\ &\qquad \left. \times (\mathbf{I}_i - \mathbf{H}_{ii}^{\mathrm{T}})^{-1} \mathbf{V}_i^{-1} \mathbf{D}_i \right\} \mathbf{F}^{-1}, \end{aligned} \tag{6}$$

where $\mathbf{F} = \Sigma_{i=1}^{K} \mathbf{D}_i^{\mathrm{T}} \mathbf{V}_i^{-1} \mathbf{D}_i$. Keener, Kmenta, and Weber (1991) have proposed an analogous bias-corrected covariance estimator to reduce the bias of White's (1980) heteroskedasticity-consistent estimator for covariance matrix of the least-squares coefficients under heteroskedasticity of unknown form.

It is possible to derive an approximation of the expected value of the residual estimator involving the elements of $\mathbf{H}_{ij}$, $i \neq j$. Let $\mathbf{r} = (\mathbf{r}_1^{\mathrm{T}}, \mathbf{r}_2^{\mathrm{T}}, \ldots, \mathbf{r}_K^{\mathrm{T}})^{\mathrm{T}}$, then using an approach analogous to the derivation of approximation (4), one can show that $\mathrm{E}[\mathbf{r}\mathbf{r}^{\mathrm{T}}] \approx (\mathbf{I}-\mathbf{H})\mathrm{cov}[\mathbf{y}](\mathbf{I}-\mathbf{H})^{\mathrm{T}}$, where $\mathbf{y} = (\mathbf{y}_1^{\mathrm{T}}, \mathbf{y}_2^{\mathrm{T}}, \ldots, \mathbf{y}_K^{\mathrm{T}})^{\mathrm{T}}$, $\mathbf{H} = \mathbf{D}(\Sigma_{l=1}^{K} \mathbf{D}_l^{\mathrm{T}}\mathbf{V}_l^{-1}\mathbf{D}_l)^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{V}^{-1}$, $\mathbf{D} = (\mathbf{D}_1^{\mathrm{T}}, \mathbf{D}_2^{\mathrm{T}}, \ldots, \mathbf{D}_K^{\mathrm{T}})^{\mathrm{T}}$, $\mathbf{V}$ = block diagonal $(\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_K)$, and $\mathbf{I}$ is an identity matrix of the same dimension as $\mathbf{H}$. Computation of a bias-corrected estimator based on the full hat matrix $\mathbf{H}$ was problematic due to the near singularity of the matrix $(\mathbf{I} - \mathbf{H})$. However, the simulation results presented in Section 5 indicate that approximation (5) provides an adequate correction for the bias of the robust covariance estimator.

4. Other Approaches to Covariance Estimation

For comparison purposes, we include a jackknife covariance estimator in our simulation study. Several authors have proposed jackknife estimators for GEE (Paik, 1988; Lipsitz et al., 1990; Qu et al., 1994). Delete-one jackknife estimators are typically used, where subsets of the data are obtained by deleting a single subject's response vector, but delete-two or -three jackknife estimators have also been studied (Qu et al., 1994). The optimal choice for the number of responses to delete when using the jackknife method is a topic for further research. A delete-one jackknife method was used in this study because of the potential convergence problems with the estimation procedure based on subsets of the data when cell counts are zero or near zero and because the delete-one method has been the most widely used.

Delete-one estimates, $\hat{\boldsymbol{\beta}}_{(i)}$, are obtained by computing the GEE estimates for each of the K subsets of data. The jackknife estimate, $\hat{\boldsymbol{\beta}}_J$, is defined by

$$\hat{\boldsymbol{\beta}}_J = \hat{\boldsymbol{\beta}} - \frac{(K-1)}{K}\sum_{i=1}^{K}\left(\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}}\right), \tag{7}$$

where $\hat{\boldsymbol{\beta}}$ is the full-data estimate using GEE (Wu, 1986). The estimated covariance of $\hat{\boldsymbol{\beta}}_J$ is given by

$$\mathrm{var}(\hat{\boldsymbol{\beta}}_J) = \frac{(K-1)}{K}\sum_{i=1}^{K}\left(\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}}_{(.)}\right)\left(\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}}_{(.)}\right)^{\mathrm{T}}, \tag{8}$$

where $\hat{\boldsymbol{\beta}}_{(.)} = (\Sigma_{i=1}^{K}\hat{\boldsymbol{\beta}}_{(i)})/K$ (Efron, 1982). Weighted jackknife estimators have also been studied, where delete-one estimates are weighted by $w_i = |\Sigma_{j\neq i}\mathbf{D}_j^{\mathrm{T}}\mathbf{V}_j^{-1}\mathbf{D}_j|$. Paik (1988) found the unweighted and weighted jackknife estimators performed similarly. We studied only the unweighted jackknife estimator.

Lipsitz et al. (1990) have studied a one-step jackknife covariance estimator defined by performing one iteration of delete-one-subject jackknife estimation starting from the full-data GEE estimates. The one-step jackknife covariance estimator is equal to the bias-corrected covariance estimator (6) minus the additional term $(\Sigma_{i=1}^{K} J_i)(\Sigma_{i=1}^{K} J_i)^{\mathrm{T}}/K$, where $\mathbf{J}_i = (\Sigma_{j=1}^{K}\mathbf{D}_j^{\mathrm{T}}\mathbf{V}_j^{-1}\mathbf{D}_j)^{-1}\mathbf{D}_i^{\mathrm{T}}\mathbf{V}_i^{-1}(\mathbf{I}_i - \mathbf{H}_{ii})^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)$. Lipsitz et al. (1990) have shown that the one-step jackknife covariance estimator, and therefore the bias-corrected covariance (6), is asymptotically equivalent to the robust covariance estimator (2).

We also evaluate a simple-to-implement modification to the GEE robust estimator (degrees-of-freedom–adjusted estimator), which consists of multiplying the GEE robust estimator by $K/(K-p)$, where K is the number of subjects and p is the number of regression parameters. MacKinnon and White (1985) have considered a similar modification to White's (1980) heteroskedasticity-consistent estimator using a degrees-of-freedom correction conventionally used to obtained unbiased estimates of the scale parameter, ϕ.

Another approach that has been used with the GEE robust estimator and jackknife estimators to improve the accuracy of the test size is to use a Student's t- or F-distribution instead of the asymptotic normal (or chi-square) distribution to compute the statistical significance (Paik, 1988; Lipsitz et al., 1990; Lipsitz et al., 1994; Qu et al., 1994). The determination of the degrees of freedom has been rather arbitrary, but if it is desired to have the test size equal to or less than the specified nominal level, it is reasonable to use the number of subjects minus the number of coefficients in the regression model as the denominator degrees of freedom and the number of parameters in the null hypothesis as the numerator degrees of freedom. Our simulation results indicated that the test size is more dependent on the number of subjects than the cluster size or the correlation.

5. Data Example

The data comes from a randomized clinical trial designed to measure the effect of a systemically administered drug on radiographic bone loss in subjects with moderate to severe periodontal disease. Thirty subjects were randomized to treatment regimens that involved oral ingestion of placebo or drug and were followed over a 2-year period, with radiographs taken of the upper and lower jaw every 6 months. The percent bone remaining was measured separately at two sites on each tooth for every subject. Because subjects did not have the same number of teeth, the number of sites per subject ranged from 34 to 64 and the average was 50 sites per subject. For this example, a binary response is used to indicate if there was bone loss over the follow-up period at a particular site, and the probability of bone loss at a site ($p_{ij} = \mathrm{E}(y_{ij} \mid x_{ij})$) was modeled as $\mathrm{logit}(p_{ij}) = \alpha + x_1\beta_{\text{treatment}} + x_2\beta_{\text{prior disease}}$, where the covariates are a subject-level covariate (x_1) indicating whether the subject was randomized to placebo or drug and a site-level covariate (x_2) indicating whether the site had clinical indications of prior disease activity based on probing depth at baseline. The regression parameters were estimated using GEE with independence and dental working correlation structures. The dental correlation structure is motivated in part due to the experience that the pairwise correlation varies within a mouth and that sites from the same tooth or quadrant, and to a lesser extent, from opposing and symmetrical sites tend to show a stronger dependence than sites from other locations in the mouth.

Table 1 shows the regression results for bone loss using GEE with independence and dental working correlation structures. The magnitude of the estimated treatment effect is similar between the two working correlations, but the treatment effect is highly significant if you ignore the correlation (i.e., if you use the model-based variance estimate for an independence working correlation) and it is not significant if you account for the correlation. It is also noteworthy that the estimated effect of prior disease activity on bone loss depends on

Table 1

Logistic regression results for radiographic bone loss using GEE with the robust, degrees-of-freedom–adjusted, bias-corrected, jackknife, and model-based covariance estimators. P-values computed using the asymptotic chi-square distribution and F-distribution with 1 and 27 d.f.

Method	Parameter	Parameter estimate	Covariance estimator	SE estimate	P-value	
					χ	F
GEE with independence working correlation	Treatment	0.381	Robust	0.327	0.24	0.25
			Degrees-of-freedom adjusted	0.345	0.27	0.28
			Bias corrected	0.352	0.28	0.29
			Jackknife	0.358	0.29	0.30
			Model based	0.130	0.0035	0.0070
	Prior disease	0.319	Robust	0.205	0.12	0.13
			Degrees-of-freedom adjusted	0.216	0.14	0.15
			Bias corrected	0.220	0.15	0.16
			Jackknife	0.220	0.15	0.16
			Model based	0.144	0.027	0.035
GEE with dental working correlation	Treatment	0.377	Robust	0.323	0.24	0.25
			Degrees-of-freedom adjusted	0.341	0.27	0.28
			Bias corrected	0.347	0.28	0.29
			Jackknife	0.353	0.31	0.32
			Model based	0.333	0.26	0.28
	Prior disease	0.359	Robust	0.166	0.031	0.040
			Degrees-of-freedom adjusted	0.175	0.040	0.050
			Bias corrected	0.174	0.039	0.049
			Jackknife	0.179	0.045	0.055
			Model based	0.142	0.011	0.018

which correlation matrix is used in GEE. The magnitude of the effect is smaller and nonsignificant with an independence working correlation and larger and significant with a dental working correlation matrix. The dental correlation structure estimates were 0.32 for observations from the same tooth, 0.18 for observations from symmetrical teeth, 0.09 for observations from the quadrant, and 0.11 for observations from opposing teeth and from other teeth. The difference in the estimates for the effect of prior disease activity between the two working correlations is due to differences in the weighting of the between-subject and within-subject information of the prior disease activity effect (Mancl and Leroux, 1998; Neuhaus and Kalbfleisch, 1998). In this case, the within-subject comparisons are most relevant, and hence, it would be appropriate to report the results based on a dental working correlation since the dental working correlation puts greater weight on the within-subject comparisons than the independence working correlation.

The standard error estimates and p-values for Wald-type tests are shown in Table 1 for the GEE robust estimator and the alternative robust covariance estimators. The alternative estimators all give standard error estimates larger than the GEE robust estimator, as would be predicted from the simulation results in the next section. Use of the alternative covariance estimators had only a small effect on the standard error estimates and p-values. An important finding is that the statistical significance of the effect of prior disease activity on bone loss using the GEE robust estimator and dental working correlation was confirmed with the alternative covariance estimators, which strengthens the conclusion that prior disease activity is associated with bone loss.

6. Simulations

The small-sample performance of the covariance estimators was investigated for binary responses by simulation. Correlated binary responses were generated using the method of Emrich and Piedmonte (1991) for samples sizes of 10, 20, 30, and 40 subjects (K) with 16, 32, or 64 observations per subject (n). Simulations were run with equal and unequal cluster sizes. For the cases of unequal cluster sizes, data were generated to have an average of 32 observations per subject. Cluster sizes of 16, 32, and 64 observations per subject mimic studies in which 2, 4, and 8 teeth from each quadrant of the mouth are monitored with observations taken at two sites on each tooth. Correlated binary responses were generated from specified underlying dental correlation structures between observations within a mouth. The values of the correlations were chosen so that the average correlation ($\bar{\rho}$) between observations within a mouth was approximately 0.1 or 0.3 (see footnote to Table 2). An average correlation of 0.1 is typical of

Table 2

Observed fraction of Wald-type test statistics rejecting individual hypothesis H_0: $\beta_l = \beta_{l(\text{true})}$, $l = 1$ *or* 2, *at a nominal* 0.05 *level and an average of* 32 *observations per subject*

			Working correlation[a]							
			$\bar{\rho} = 0.1$[a]				$\bar{\rho} = 0.3$[a]			
			Independence		Dental		Independence		Dental	
K	n	Covariance estimator	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
10	Equal	Robust	0.132	0.140	0.122	0.125	0.139	0.154	0.123	0.116
		Degrees-of-freedom adjusted	0.085	0.092	0.074	0.073	0.091	0.098	0.083	0.069
		Bias corrected	0.063	0.071	0.064	0.081	0.072	0.066	0.070	0.078
		Jackknife	0.100	0.117	0.098	0.107	0.114	0.112	0.105	0.108
		Model based	0.327	0.164	0.120	0.073	0.553	0.282	0.098	0.065
10	Unequal	Robust	0.179	0.158	0.167	0.167	0.165	0.179	0.144	0.204
		Degrees-of-freedom adjusted	0.121	0.107	0.112	0.113	0.121	0.101	0.098	0.134
		Bias corrected	0.064	0.069	0.066	0.095	0.069	0.065	0.070	0.118
		Jackknife	0.081	0.085	0.085	0.092	0.143	0.127	0.129	0.163
		Model based	0.381	0.197	0.147	0.097	0.544	0.342	0.129	0.124
20	Equal	Robust	0.071	0.089	0.066	0.081	0.093	0.105	0.091	0.089
		Degrees-of-freedom adjusted	0.051	0.069	0.047	0.062	0.068	0.081	0.071	0.068
		Bias corrected	0.042	0.058	0.044	0.066	0.061	0.057	0.060	0.073
		Jackknife	0.056	0.078	0.058	0.070	0.066	0.073	0.069	0.075
		Model based	0.302	0.160	0.067	0.055	0.527	0.308	0.081	0.058
20	Unequal	Robust	0.104	0.111	0.085	0.115	0.109	0.136	0.087	0.158
		Degrees-of-freedom adjusted	0.080	0.087	0.067	0.095	0.087	0.107	0.075	0.128
		Bias corrected	0.060	0.064	0.059	0.091	0.063	0.072	0.059	0.103
		Jackknife	0.063	0.069	0.057	0.081	0.058	0.077	0.051	0.077
		Model based	0.360	0.240	0.104	0.082	0.586	0.411	0.098	0.121
30	Equal	Robust	0.072	0.079	0.069	0.075	0.069	0.084	0.064	0.071
		Degrees-of-freedom adjusted	0.058	0.064	0.052	0.064	0.058	0.065	0.057	0.054
		Bias corrected	0.047	0.060	0.049	0.065	0.054	0.054	0.052	0.063
		Jackknife	0.059	0.070	0.058	0.069	0.056	0.062	0.054	0.069
		Model based	0.322	0.169	0.068	0.054	0.538	0.305	0.062	0.047
30	Unequal	Robust	0.083	0.088	0.072	0.096	0.088	0.089	0.076	0.115
		Degrees-of-freedom adjusted	0.070	0.078	0.061	0.077	0.078	0.079	0.062	0.105
		Bias corrected	0.055	0.062	0.058	0.072	0.063	0.054	0.055	0.086
		Jackknife	0.056	0.063	0.053	0.068	0.058	0.054	0.051	0.073
		Model based	0.380	0.203	0.069	0.054	0.562	0.378	0.078	0.094
40	Equal	Robust	0.066	0.063	0.063	0.071	0.072	0.064	0.074	0.061
		Degrees-of-freedom adjusted	0.058	0.054	0.057	0.058	0.061	0.053	0.065	0.052
		Bias corrected	0.054	0.051	0.052	0.060	0.057	0.046	0.061	0.055
		Jackknife	0.060	0.057	0.058	0.072	0.060	0.051	0.063	0.055
		Model based	0.330	0.161	0.065	0.054	0.536	0.302	0.070	0.051
40	Unequal	Robust	0.075	0.087	0.064	0.082	0.074	0.094	0.060	0.101
		Degrees-of-freedom adjusted	0.064	0.076	0.052	0.074	0.065	0.086	0.046	0.090
		Bias corrected	0.055	0.064	0.050	0.074	0.051	0.068	0.043	0.075
		Jackknife	0.055	0.064	0.047	0.064	0.050	0.068	0.042	0.064
		Model based	0.336	0.218	0.067	0.064	0.576	0.380	0.065	0.098

[a] An average correlation of 0.1 was given by specifying the correlation between observations from the same tooth as 0.3, from the same quadrant as 0.2, from symmetrical teeth as 0.15, from opposing teeth as 0.08, and from other teeth as 0.06. An average correlation of 0.3 was given by specifying the correlation between observations from the same tooth as 0.5, from the same quadrant as 0.4, from symmetrical teeth as 0.3, from opposing teeth as 0.28, and from other teeth as 0.25.

the correlation between sites for measures of naturally occurring periodontal disease, whereas an average correlation of 0.3 represents a strong correlation between sites that can be produced by experimentally induced disease (e.g., ligature-induced periodontal disease in animal studies). An additional set of simulations run with an exchangeable correlation of 0.5 indicated that the results on test size presented in Tables 2 and 3 will hold for higher correlation values. The choice of using the correlation coefficient to parameterize the dependence between observations is arbitrary, but a set of simulations run

Table 3

Observed fraction of Wald-type test statistics rejecting individual hypotheses H_0: $\beta_l = \beta_{l(\text{true})}$, $l = 1$ *or* 2, *and joint hypothesis* H_0: $\beta_1, \beta_2 = \beta_{1(\text{true})}, \beta_{2(\text{true})}$ *at a nominal* 0.05 *level, an average of* 32 *observations per subject, and a dental working correlation. Critical values computed using chi-square and* F*-distribution.*

K	n	Covariance estimator	Critical value	$\bar{\rho} = 0.1$			$\bar{\rho} = 0.3$		
				β_1	β_2	β_1,β_2	β_1	β_2	β_1,β_2
10	Equal	Robust	χ	0.122	0.125	0.198	0.123	0.116	0.219
			F	0.073	0.072	0.106	0.092	0.068	0.099
		Degrees-of-freedom adjusted	χ	0.074	0.073	0.120	0.083	0.069	0.122
			F	0.047	0.040	0.057	0.048	0.034	0.046
		Bias corrected	χ	0.064	0.081	0.117	0.070	0.078	0.117
			F	0.031	0.044	0.047	0.037	0.040	0.047
		Jackknife	χ	0.098	0.107	0.167	0.105	0.110	0.187
			F	0.059	0.064	0.093	0.067	0.060	0.099
		Model based	χ	0.120	0.073	0.113	0.098	0.065	0.090
			F	0.066	0.025	0.039	0.050	0.029	0.027
	Unequal	Robust	χ	0.167	0.167	0.261	0.144	0.204	0.272
			F	0.109	0.113	0.150	0.096	0.129	0.141
		Degrees-of-freedom adjusted	χ	0.112	0.113	0.176	0.098	0.134	0.167
			F	0.070	0.064	0.096	0.057	0.068	0.073
		Bias corrected	χ	0.066	0.095	0.143	0.070	0.118	0.128
			F	0.039	0.052	0.058	0.036	0.061	0.057
		Jackknife	χ	0.086	0.092	0.165	0.130	0.164	0.243
			F	0.057	0.053	0.086	0.088	0.109	0.154
		Model based	χ	0.147	0.097	0.150	0.129	0.124	0.175
			F	0.097	0.035	0.079	0.080	0.079	0.076
40	Equal	Robust	χ	0.063	0.071	0.077	0.073	0.061	0.075
			F	0.057	0.058	0.066	0.067	0.053	0.058
		Degrees-of-freedom adjusted	χ	0.057	0.058	0.066	0.065	0.052	0.058
			F	0.051	0.051	0.054	0.059	0.043	0.049
		Bias corrected	χ	0.052	0.060	0.065	0.061	0.055	0.061
			F	0.050	0.053	0.053	0.050	0.047	0.050
		Jackknife	χ	0.058	0.072	0.071	0.063	0.058	0.072
			F	0.052	0.061	0.062	0.050	0.049	0.056
		Model based	χ	0.065	0.054	0.066	0.070	0.051	0.056
			F	0.060	0.045	0.057	0.064	0.045	0.044
	Unequal	Robust	χ	0.064	0.082	0.097	0.060	0.101	0.106
			F	0.052	0.076	0.078	0.048	0.091	0.090
		Degrees-of-freedom corrected	χ	0.052	0.074	0.080	0.046	0.090	0.091
			F	0.047	0.071	0.065	0.042	0.082	0.076
		Bias corrected	χ	0.050	0.074	0.074	0.043	0.075	0.078
			F	0.044	0.068	0.058	0.039	0.069	0.058
		Jackknife	χ	0.047	0.064	0.067	0.042	0.064	0.066
			F	0.044	0.057	0.053	0.037	0.057	0.055
		Model based	χ	0.067	0.064	0.070	0.065	0.098	0.102
			F	0.058	0.058	0.057	0.060	0.084	0.084

with data generated using the odds ratio to parameterize the dependence between responses and to describe the working correlation matrix gave similar results to those shown in Tables 2 and 3.

A logistic regression model was used for the marginal mean of y_{ij}, $\text{logit}(\text{E}[y_{ij} \mid x_i, w_{ij}]) = \beta_0 + \beta_1 x_i + \beta_2 w_{ij}$, where x_i was a binary subject-level covariate ($x_i = 0$ or 1 with equal probability) and w_{ij} was a binary site-level covariate ($w_{ij} = 0$ or 1), $i = 1, 2, \ldots, K$, $j = 1, 2, \ldots, n_i$. The site-level covariate distribution was generated to mimic the unequal distribution of the site-level covariate among the subjects in the data example of Section 5. For clusters sizes of 32 and 64, $\Sigma_j w_{ij} =$

4, 12, 20, or 28 with equal probabilities, and for cluster size of 16, $\Sigma_j w_{ij} = 4$ or 12 with 0.25 probabilities and $\Sigma_j w_{ij} = 16$ with 0.50 probability. The values of the regression parameters specified in the simulations ($\beta_0, \beta_1, \beta_2 = -1.6$, 0.38, and 0.35) were also chosen to correspond with the data example in Section 5. For each data configuration, 1000 simulations were generated, and for each simulation, the estimates were obtained using independence and dental working correlations. Simulations were performed on a DECstation 5000/200 using programs written in Fortran.

6.1 *Size of Individual and Joint Hypothesis Tests*

The performance of the covariance estimators was evaluated by computing the observed fraction of Wald-type test statistics rejecting the individual null hypotheses H_0: $\beta_l = \beta_{l(\text{true})}$, $l = 1$ or 2, and the joint null hypothesis H_0: $\beta_1, \beta_2 = \beta_{1(\text{true})}, \beta_{2(\text{true})}$. The observed fractions for the individual hypotheses are shown in Table 2 for a 0.05 nominal level. At a true nominal 0.05 level and 1000 simulations, we would expect the estimated test size to be between 0.036 and 0.064 (95% confidence interval). As others have observed, the estimated test sizes tended to be inflated when using the GEE robust estimator (2). The estimated test sizes were substantially greater than 0.05, i.e., 0.10—0.20, for cases involving 10 subjects or for cases involving 20 subjects with unequal cluster sizes. In general, the estimated test sizes were more inflated for unequal cluster sizes than for equal cluster sizes, and the estimated test sizes became more inflated as the number of subjects decreased. In contrast, the average correlation value and working correlation structure had little effect on the estimated test sizes. Also, the estimated test size was not greatly affected by the number of observations per subject (results not shown).

For cases involving equal cluster sizes, the estimated test sizes with the degrees-of-freedom adjustment (degrees-of-freedom–adjusted estimator) were usually between 0.036 and 0.064 with 20 or more subjects and only a little inflated (0.069 to 0.098) with 10 subjects (Table 2). However, when the cluster sizes were unequal, the estimated test sizes were frequently greater than 0.10 with 10 or 20 subjects. With the bias-corrected estimator, the estimated test sizes were smaller and almost always closer to 0.05 compared with the degrees-of-freedom–adjusted and robust estimators. Also, the estimated size was usually less than 0.10 even when the number of subjects was 10 both for equal and unequal cluster sizes. The estimated test sizes with the jackknife estimator were similar to the bias-corrected estimator when the number of subjects was 20 or greater. When the number of subjects was 10, the sizes were often greater than 0.10. The jackknife estimator did not perform well when there were convergence problems with the delete-one estimates due to zero or near-zero cell counts. Test sizes were also estimated using the one-step jackknife estimator of Lipsitz et al. (1990). The estimated sizes for the one-step jackknife estimator were between the test sizes for the degrees-of-freedom–adjusted and bias-corrected estimators (results not shown). The estimated test sizes with the model-based estimator using the dental working correlation were usually closer to 0.05 compared with the robust estimator, but still the estimated test sizes were frequently greater than 0.064. The inflation of the test size with the model-based estimator using the true correlation structure is partially due to the underestimation of the correlation when the number of independent clusters is small.

Although the bias-corrected covariance estimator gave test sizes closer to 0.05 than the GEE robust, degrees-of-freedom–adjusted, and jackknife estimators, the estimated test sizes were frequently greater than 0.05 and greater than 0.064 for cases with 10 subjects. The observed fractions of Wald-type test statistics rejecting individual and joint hypotheses based on chi-square and F-distributions are shown in Table 3 for a 0.05 nominal level. The estimated sizes with the F-distribution were in most cases closer to 0.05 than with the asymptotic chi-square distribution and were further from 0.05 than with the chi-square distribution only when the estimated size with the F-distribution was less than 0.05. Using the F-distribution in conjunction with the bias-corrected covariance estimator gave test sizes nearly always between 0.036 and 0.064 for both equal and unequal cluster sizes. These results indicate that the bias-corrected estimator used in conjunction with the F-distribution will produce tests with the proper size. Also, computing the critical value based on an F-distribution rather than a chi-square distribution resulted in only a minimal loss of power, usually 0.05 or less and always less than 0.10 (results not shown). The degrees-of-freedom–adjusted estimator used in conjunction with the F-distribution also gave test sizes nearly always within 0.036 to 0.064 when the cluster sizes were equal, but test sizes were frequently greater than 0.064 when the cluster sizes were unequal. The estimated test sizes for the jackknife estimator used in conjunction with the F-distribution were similar to the bias-corrected estimator when the number of subjects was 20 or more, but the estimated test sizes were often inflated for cases with 10 subjects.

7. Discussion

We evaluated the properties of GEE and several alternative covariance estimators for use in situations involving a small number of subjects. The GEE robust estimator tended to underestimate the variance, and the size of Wald-type tests could be substantially greater than the nominal level. The degree of inflation in the test size was usually small if the number of subjects was 20 or more and cluster sizes were equal. However, the degree of inflation in the test size could be substantial if cluster sizes were not equal, even with 40 subjects. These results indicate that, even for sample sizes involving a moderate number of subjects, using the GEE robust estimator can produce tests that are too liberal if the cluster sizes are unequal. The estimated tests sizes with the model-based estimator using the true correlation structure were usually closer to the nominal level compared with the robust estimator, but still the estimated test sizes were often inflated. These results indicate that even careful modeling of the correlation structure is not sufficient to maintain the nominal test size when the number of independent clusters is small.

The bias-corrected, jackknife, and degrees-of-freedom–adjusted covariance estimators generally performed better than the GEE robust estimator with respect to maintaining the nominal test size. However, the test sizes still tended to be inflated for all covariance estimators for cases involving 20 or fewer subjects and using the asymptotic chi-square distribution to compute the critical values. Using an F-distribution

to compute the critical values produced tests with sizes substantially closer to the nominal level. The jackknife covariance estimator produced test sizes similar to the bias-corrected covariance estimator except for cases with 10 subjects. For cases involving 10 subjects, the test sizes with the jackknife estimator were substantially larger than the test sizes of the bias-corrected and degrees-of-freedom–adjusted covariance estimators. The poor performance of the jackknife estimator was likely due to the convergence problems with estimation of the delete-one estimates caused by zero or near-zero cell counts. This problem has been observed with other resampling methods in small sample sizes for binary responses (Sherman and le Cessie, 1997). The degrees-of-freedom–adjusted estimator used in conjunction with the F-distribution gave tests of proper size when the cluster sizes were equal, but test sizes were frequently inflated when the cluster sizes were unequal. These results indicate that the degrees-of-freedom correction used in conjunction with the F-distribution may provide adequate protection against the bias of the GEE robust estimator in small samples only when the cluster sizes are approximately equal. The bias-corrected covariance estimator, which reduces the bias of the residual estimator of the unknown response covariance matrix, generally gave test sizes closest to the nominal level, and when used in conjunction with the F-distribution, the estimated test sizes were nearly always close to the nominal level (0.036–0.064), even when the number of subjects was 10 and for tests of joint hypotheses. These results indicate that the bias-corrected estimator used in conjunction with the F-distribution will produce tests with the proper size for all the conditions evaluated. Also, computing the critical value based on an F-distribution resulted in only a minimal loss of power.

A limitation of the current simulation study is that we did not study the effects of varying the magnitude of the parameter estimates or response probabilities. However, we can infer from the simulation results of Gunsolley et al. (1995), who considered response probabilities from 0.025 to 0.30, that the underestimation of the GEE robust estimator would increase as the response probability approaches zero or one. A topic of further interest would be the effect of the covariate distribution on the GEE robust estimator, particularly for a covariate with a skewed distribution. The hat-matrix expression for the bias of the residual estimator (5) indicates that the bias of the GEE robust estimator may also be large when the covariate distribution is skewed (Sherman and le Cessie, 1997). We considered only the case of a binary response, but the alternative covariance estimators can easily be applied to other response types. We can expect the relative performance of the alternative covariance estimators for other response types to be similar to what we observed for a binary response except that the jackknife estimator would not have the same convergence problems due to zero or near-zero cell counts a for continuous response.

A GEE SAS macro implementing the bias-corrected covariance estimator can be obtained by contacting the author at lman@biostat.washington.edu.

Acknowledgements

This work was supported in part by National Institute of Dental and Craniofacial Research grants T32 DE07227 and P30 DE09743.

Résumé

Nous proposons dans cet article une alternative à l'estimateur sandwich des covariances des équations d'estimation généralisées. Quand le nombre d'unités indépendantes est faible, la taille des tests qui utilisent l'estimateur sandwich peut être augmentée. Dans ce cas, le recours pour estimer les covariances aux méthodes de rééchantillonnage, comme le bootstrap ou le jackknife, a été suggéré; mais des difficultés surviennent quand la réponse est binaire et que le rééchantillonnage conduit à des totaux nuls ou quasi-nuls. Nous proposons une estimation des covariances corrigée du biais qui évite ce problème. Nous avons comparé l'estimateur corrigé du biais à l'estimateur sandwich et à celui du jackknife pour 16 à 64 réponses binaires par unité et des échantillons de 10 à 40 unités. L'estimateur corrigé du biais conduisait à des tests de taille proche de la valeur nominale, même avec 10 unités et des nombres inégaux de réponses par unité; l'estimateur sandwich et celui du jackknife conduisaient en revanche à des tests dont la taille pouvait atteindre le double ou le triple de la valeur nominale. Les méthodes sont appliquées aux données d'un essai thérapeutique concernant le traitement de la perte osseuse périodontique.

References

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* Philadelphia: Society for Industrial and Applied Mathematics.

Emrich, L. J. and Piedmonte, M. R. (1991). A method for generating high-dimensional multivariate binary variates. *The American Statistician* **45,** 302–304.

Emrich, L. J. and Piedmonte, M. R. (1992). On some small sample properties of generalized estimating equation estimates for multivariate dichotomous outcomes. *Journal of Statistical Computation and Simulation* **41,** 19–29.

Feng, Z., McLerran, D., and Grizzle, J. (1996). A comparison of statistical methods for clustered data analysis with Gaussian error. *Statistics in Medicine* **15,** 1793–1806.

Gunsolley, J. C., Getchell, C., and Chinchilli, V. A. (1995). Small sample characteristics of generalized estimating equations. *Communications in Statistics: Simulation* **24,** 869–878.

Keener, R. W., Kmenta, J., and Weber, N. C. (1991). Estimation of the covariance matrix of the least-squares regression coefficients when the disturbance covariance matrix is of unknown form. *Econometric Theory* **7,** 22–45.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73,** 13–22.

Lipsitz, S. R., Fitzmaurice, G. M., Orav, E. J., and Laird, N. M. (1994). Performance of generalized estimating equations in practical situations. *Biometrics* **50,** 270–278.

Lipsitz, S. T., Laird, N. M., and Harrington, D. P. (1990). Using the jackknife to estimate the variance of regression estimators from repeated measures studies. *Communications in Statistics: Theory and Methods* **19,** 821–845.

MacKinnon, J. G. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* **29,** 305–325.

Mancl, L. A. and Leroux, B. G. (1998). Evaluation of between- and within-subject effects for the analysis of clustered data. *Proceedings of the Section on Statistical Education, American Statistical Association,* 57–62.

Neuhaus, J. M. and Kalbfleisch, J. D. (1998). Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* **54,** 638–645.

Paik, M. C. (1988). Repeated measurement analysis for nonnormal data in small samples. *Communications in Statistics: Simulations* **17,** 1155–1171.

Preisser, J. S. and Qaqish, B. F. (1996). Deletion diagnostics for generalized estimating equations. *Biometrics* **83,** 551–562.

Qu, Y., Piedmonte, M. R., and Williams, G. W. (1994). Small sample validity of latent variable models for correlated binary data. *Communications in Statistics: Simulations* **23,** 243–269.

Sharples, K. and Breslow, N. (1992). Regression analysis of correlated binary data: Some small sample results for the estimating equation approach. *Journal of Statistical Computation and Simulation* **42,** 1–20.

Sherman, M. and le Cessie, S. (1997). A comparison between bootstrap methods and generalized estimating equations for correlated outcomes in generalized linear model. *Communications in Statistics: Simulation* **26,** 901–925.

White, H. (1980). A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity. *Econometrica* **48,** 817–838.

Wu, C. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *The Annals of Statistics* **14,** 1261–1350.

Received September 1999. *Revised May* 2000.
Accepted May 2000.