Kastner, Ziegler:

A Comparison of Jackknife Estimators of Variance for GEE2

Sonderforschungsbereich 386, Paper 167 (1999)

Online unter: http://epub.ub.uni-muenchen.de/

A Comparison of Jackknife Estimators of Variance for GEE2

C. Kastner * A. Ziegler **

August 31, 1999

Abstract

Marginal regression modeling with generalised estimating equations became very popular in the last decade. While the mean structure is of primary interest in first-order generalised estimating equations (GEE1), second-order generalised estimating equations (GEE2) allow the estimation of both the mean and the association structure. It has repeatedly been shown that the usual robust variance estimator for the GEE1 is conservative, especially in small samples. As an alternative, the jackknife estimator of variance can be used. In this discussion paper, we extend the different jackknife estimators of variance to GEE2 models. The variance estimators are compared in a simulation study. While there is only little difference in the variance estimates of the mean structure across simulated models, the results differ substantially with respect to the association structure. The fully iterated jackknife estimator seems to be the most appropriate when focusing on the GEE2.

Keywords: Generalised Estimating Equations, Marginal Models, Jackknife Estimators

1 Introduction

Marginal regression modeling with generalised estimating equations became very popular in the last decade. The mean structure is of primary interest in first-order generalised estimating equations (GEE1). There, the association structure is treated as nuisance. In family studies, however, the association structure is of primary interest. The mean structure is required to adjust the association structure for covariates. In this situation, second-order generalised estimating equations (GEE2) might be applied which simultaneously analyse the mean and the association structure. An overview on these different models can be found e.g. in Ziegler, Kastner and Blettner (1998).

The major advantage of GEE compared with likelihood approaches is that higher order moments need not be correctly specified. The parameters of the

*Institute of Statistics, LMU München, Ludwigstr. 33, 80539 München, Germany, kchris@stat.uni-muenchen.de

**Medical Centre for Methodology and Health Research, Institute of Medical Biometry and Epidemiology, Philipps–University of Marburg, Bunsenstr. 3, 35033 Marburg, Germany, ziegler@mailer.uni-marburg.de

mean structure for GEE1 models and of the mean and the association structure for GEE2 models can still be estimated strongly consistently. To correct for possible misspecification, White (1982) proposed the robust variance estimator of variance, which is also termed sandwich matrix. As an alternative, the jackknife estimator of variance can be applied (Lipsitz, Laird and Harrington, 1990; Ziegler, 1997). Since the fully iterated Jackknife estimator is computational very demanding, several approximations have been proposed (Lipsitz et al., 1990; Ziegler, 1997). In Monte-Carlo simulations, the class of jackknife estimators was shown to be superior compared with the usual robust variance estimator when analysing GEE1 models, especially for small samples.

The jackknife estimators of variance have, however, not been applied to GEE2 models. Therefore, we extend them to GEE2 models in this paper. We compare the different jackknife estimators of variance and the usual robust variance estimator in a Monte-Carlo simulation study.

This paper is organised as follows. The theory of the GEE2 and the employed notation is presented in section 2. In section 3 the design of a the simulation study is described. In this section, its results are described in detail and also discussed.

2 Generalized estimating equations for mean and association structures

Let y_{it} be the response of observation t, $t = 1, \ldots, T$, from cluster i, $i, 1, \ldots, n$. For each y_{it} a vector of covariates x_{it} is available, which possibly contains an intercept. The data are summarised to $y_i = (y_{i1}, \ldots, y_{iT})'$ and $X_i = (x'_{i1}, \ldots, x'_{iT})'$. The pairs (y_i, X_i) are assumed to be independent. In our simulation study we focus on continuous responses. Therefore, we use the identity link function to connect the conditional mean of y_{it} given X_i and the $p \times 1$ parameter vector β of the mean structure:

$$\mu_{it} = \mathrm{E}(y_{it}|X_i) = \mathrm{E}(y_{it}|x_{it}) = x'_{it}\beta \qquad (1)$$

We furthermore assume that the correlation coefficient is a function of the $q \times 1$ parameter vector α of the association structure but independent of the mean structure parameter β . We choose the area tangens hyperbolicus as association link function so that for $t \neq t'$

$$\rho_{itt'} = \mathrm{Corr}(y_{it}, y_{it'}|X_i) = \frac{\exp\{k(\tilde{x_{it}}, \tilde{x_{it'}})'\alpha\} - 1}{\exp\{k(\tilde{x_{it}}, \tilde{x_{it'}})'\alpha\} + 1} \qquad (2)$$

Equation (2) guarantees that the correlation coefficient does not exceed 1 in absolute values. k is a function that correctly describes the relationship between the explanatory variables for the association structure $\tilde{x}_{it}$ and $\tilde{x}_{it'}$ and the correlation coefficient (Lipsitz et al., 1990). The GEE proposed by Prentice (1988) are given by

$$\mathrm{s}(\hat{\theta}) = s\begin{pmatrix} \hat{\beta} \\ \hat{\alpha} \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} D_i & 0 \\ 0 & E_i \end{pmatrix}' \begin{pmatrix} V_i & 0 \\ 0 & W_i \end{pmatrix}^{-1} \begin{pmatrix} y_i - \mu_i \\ z_i - \rho_i \end{pmatrix} = 0 \qquad (3)$$

where μ_i is the $T \times 1$ vector of the μ_{it} and ρ_i is the $T(T-1)/2 \times 1$ vector of the $\rho_{itt'}$. z_i is the corresponding vector of the product of the standardised residuals

$z_{itt'} = (y_{it} - \mu_{it})(y_{it'} - \mu_{it'})/\sigma_{it}\sigma_{it'}$ with $\sigma_{it} = \text{Var}(y_{it}|X_i)$. $D_i = \partial\mu_i/\partial\beta'$ and $E_i = \partial\rho/\partial\alpha'$ are the first derivatives w.r.t. β and α, respectively, while V_i and W_i are the conditional working covariance matrices of y_i and z_i given X_i, respectively. Usually, W_i is chosen as the working matrix for applications (Ziegler et al., 1998) so that W_i is a $T(T-1)/2$ dimensional identity matrix.

The GEE (3) for β and α may be solved separately by an alternating modified Fisher scoring algorithm because they can be separated in two independent estimating equations. Equation (3) can be derived from the generalised method of moments (Ziegler, 1995). Thus, $\hat{\theta} = (\hat{\beta}', \hat{\alpha}')'$ is a strongly consistent estimator of $\theta = (\beta', \alpha')'$ under suitable regularity conditions (Hansen, 1982), if eqs. (1) and (2) are correctly specified. Furthermore, $\hat{\beta}$ and $\hat{\alpha}$ are jointly asymptotic normal. The robust variance matrix, also termed Huber or sandwich variance matrix, is given by

$$\text{Cov}(\hat{\theta}) = \begin{pmatrix} A_{11} & 0 \\ -A_{21} & A_{22} \end{pmatrix}^{-1} \begin{pmatrix} B_{11} & B_{12} \\ B_{12}' & B_{22} \end{pmatrix} \begin{pmatrix} A_{11} & 0 \\ -A_{21} & A_{22} \end{pmatrix}^{-1}$$

with the matrices $A_{11} = \sum D_i' V_i^{-1} D_i$, $A_{21} = \sum E_i' W_i^{-1} \frac{\partial z_i}{\partial \beta}$, $A_{22} = \sum E_i' W_i^{-1} E_i$, $B_{11} = \sum D_i' V_i^{-1} \text{Cov}(y_i) V_i^{-1} D_i$, $B_{12} = \sum D_i' V_i^{-1} \text{Cov}(y_i, z_i) W_i^{-1} E_i$ and $B_{22} = \sum E_i' W_i^{-1} \text{Cov}(z_i) W_i^{-1} E_i$.

In the framework of GEE1, Paik (1988) recommended to use jackknife estimators of variance instead of the robust variance matrix in small samples because the robust variance matrix yielded biased estimates. Lipsitz and colleagues (1990; 1994) showed for the GEE1 that the unweighted deletion-1 jackknife estimator of variance

$$\left(\frac{n-p}{n}\right) \sum_{i=1}^{n} \left(\hat{\beta}_{-i} - \hat{\beta}\right) \left(\hat{\beta}_{-i} - \hat{\beta}\right)'$$

is asymptotically equivalent to the corresponding robust variance matrix. This property can be easily extended to the GEE2 of eq. (3). Here, $\frac{n-p}{n}$ is replaced by $\frac{n-(p+q)}{n}$. Furthermore, the jackknife now involves both β and α. Deletion-1 jackknife estimators are usually obtained by a modified Fisher scoring with starting value $\hat{\theta} = (\hat{\beta}', \hat{\alpha}')'$, where each family is successively omitted in a loop. Instead of the fully iterated (FIJ) jackknife estimator, a 'one-step' approximation (1-SJ) might be used by stopping the algorithm after one Fisher scoring step (Lipsitz et al., 1990). For GEE1, the 'one-step' approximation gave better coverage probabilities than the fully iterated jackknife estimator in Monte-Carlo simulations (Lipsitz et al., 1990). The jackknife estimator of variance can also be approximated without successively leaving out each cluster during the calculations as shown by Ziegler (1997) for GEE1. This generally increases the computation speed. The approximation of the jackknife estimator of variance (AJS) for Prentice's GEE2 is given for $T \geq 2$ by

$$\frac{n-(p+q)}{n} A\,C\,A$$

with

$$A = \begin{pmatrix} A_{11} & 0 \\ -A_{21} & A_{22} \end{pmatrix}^{-1}$$

and

$$C \quad = \quad \left[\sum_{i=1}^{n} \left\{ F_i' K_i \tilde{V}_i^{-1/2} \begin{pmatrix} y_i - \mu_i \\ z_i - \rho_i \end{pmatrix} \begin{pmatrix} y_i - \mu_i \\ z_i - \rho_i \end{pmatrix}' \tilde{V}_i^{-1/2} K_i F_i' \right\} \right]$$

Here, F_i is the $(T + T(T-1)/2) \times (p+q)$ block diagonal matrix of $(V_i^{-1/2} D_i, W_i^{-1/2} E_i)$. Analogously, $\tilde{V}_i^{-1/2}$ is the block diagonal matrix of $(V_i^{-1/2}, W_i^{-1/2})$, and, finally, $K_i = I_{(T+T(T-1)/2 \times (T+T(T-1)/2} + (I_{(T+T(T-1)/2 \times (T+T(T-1)/2} - F_i A^{-1} F_i')^{-1} F_i A^{-1} F_i'$. For $T = 1$, the estimation can be performed by letting $z_i - \rho_i = 0$ and a adding a row and column of 0 to F_i, so that

$$F_i = \begin{pmatrix} F_i & 0 \\ 0 & 0 \end{pmatrix}$$

3 Simulation study

In order to compare the properties of the three jackknife estimators with the usual robust estimator of variance, we perform a simulation study using a continuous response variable, the identity link function and the area tangens hyperbolicus association link function.

The jackknife was shown to be superior to the classic robust variance for small sample sizes (Paik, 1988; Lipsitz et al., 1990). Thus, 50 clusters (families) of size 3 were simulated with 1,000 replicates for each model. The simulation proceeds as follows. First, the design matrix X is generated for each cluster. Second, the response vector y is simulated for each cluster using a multivariate normal distribution. The estimation is done by MAREG (Kastner, Fieger and Heumann, 1997).

Mancl and Leroux (1996) have shown that the efficiency of GEE estimates is quite sensitive to the between- and within-cluster variation of the explanatory variables. Thus, we choose eight different models that specifically focus on this aspect in our simulations. They all include one non-random binary and one non-random continuous explanatory variable for the mean structure. The covariates are subject to variation as they are chosen to be either cluster-constant or non mean-balanced cluster-specific.

The binary variable is dummy-coded. The binary variable is one in 40 of the 50 clusters for the cluster-constant model. For the generation of a within-cluster varying binary variable, we set the number of ones to 40, 30 and 10 for the three observations within a cluster. This results in the patterns displayed in table 1.

pattern	frequency
1-1-1	10
1-1-0	20
1-0-0	10
0-0-0	10

Table 1: Frequency table for pattern of within-cluster varying binary variable

The non-random continuous variable is generated from the frequency distribution of grouped data. Therefore, the lower and upper bound and the frequency has to be specified. Within a class, the continuous values are equidistant. The distribution for the cluster-constant case is given in table 2. To generate a within-cluster varying continuous variable, we use the three different frequency distributions which are given in table 2.

Cluster-constant		Within-cluster varying			
Interval	freq.	Interval obs.1	Interval obs.2	Interval obs.3	freq.
[0; 1]	5	[0; 1]	[0; 2]	[1; 3]	5
[1; 3]	10	[1; 3]	[2; 5]	[3; 4]	10
[3; 6]	20	[3; 6]	[5; 6]	[4; 8]	20
[6; 8]	10	[6; 8]	[6; 8]	[8; 9]	10
[8; 9]	5	[8; 9]	[8; 9]	[9; 9]	5

Table 2: Frequency table of continuous variable

This results in four different configurations for the parameters of the mean structure. In any case the theoretical parameters for the explanatory variables are $\theta_0 = 1$ (regression constant), $\theta_1 = 3$ (binary variable) and $\theta_2 = -0.2$ (continuous variable). Thus, the effects itself are cluster-constant.

In addition to the mean structure, the association structure is also subject to variation. We use the exchangeable and the unstructured association structure in our simulations. Parameters for the unspecified correlation structure are chosen to be $\theta_3 = 0.5$, $\theta_4 = 1$ and $\theta_5 = 0.7$, resulting in correlation coefficients of 0.245, 0.462 and 0.336 for the pairs 1-2, 1-3 and 2-3, respectively. The parameter value for the exchangeable correlation structure is set to $\theta_3 = 0.7$. The variance is set to $\sigma_{it}^2 = 1$ for all three observations within a cluster. Therefore, the correlation matrix is identical to the covariance matrix.

Given the design matrix X, the parameters of the mean structure and the covariance matrix, the response vector for each cluster is simulated using a multivariate normal distribution (Fieger, Heumann, Kastner and Watzka, 1997). The pseudo random numbers were generated using DRAND48, which is supplied by SunOS (1995, man Pages(3C)) as a C-library function.

The results from the Monte-Carlo simulations are shown in tables 3 to 10. The tables display the mean parameter estimate and the standard error of the mean from the 1,000 replicates in addition to the theoretical parameter values and the different estimates of the standard error.

Obviously, with either of the four approaches for estimating or approximating the robust variance matrix, the standard error of the mean is well approximated for the parameter estimates of the mean structure, i.e. θ_0, θ_1 and θ_2. This is in line with the findings of Lipsitz and colleagues (Lipsitz et al., 1990; Lipsitz, Dear and Zhao, 1994) and Ziegler (1997). However, the results differ substantially with respect to the association structure. The standard errors using the usual robust variance matrix according to Prentice (1988) are far too large, resulting in conservative tests. On the other side, the 1-SJ is too liberal for all eight models. The AJS generally is conservative for the simulated models. The best approximation to the true standard error of the mean is obtained with the FIJ. Therefore, we recommend to use the FIJ in small samples. For a look at the first glance, both the 1-SJ and the AJS seem to be appropriate.

param.	Theoret. value	Mean param. estimate	Std.err. of the mean	Standard error			
				Prentice	AJS	FIJ	1-SJ
θ_0	1	1.007	0.233	0.227	0.235	0.234	0.234
θ_1	3	2.995	0.214	0.203	0.209	0.207	0.207
θ_2	-0.2	-0.201	0.043	0.040	0.042	0.042	0.042
θ_3	0.7	0.622	0.198	0.391	0.236	0.198	0.187

Table 3: Simulation results with mean-unbalanced cluster constant covariates and exchangeable correlation structure. AJS: approximation of the jackknife estimator of variance, FIJ: fully iterated jackknife estimator of variance, 1-SJ: one-step approximation of the jackknife estimator of variance

param.	Theoret. value	Mean param. estimate	Std.err. of the mean	Standard error			
				Prentice	AJS	FIJ	1-SJ
θ_0	1.003	0.236	0.225	0.230	0.231	0.228	
θ_1	3.000	0.215	0.202	0.207	0.205	0.202	
θ_2	-0.200	0.041	0.040	0.041	0.041	0.041	
θ_3	0.422	0.282	0.405	0.286	0.277	0.265	
θ_4	0.924	0.317	0.622	0.366	0.304	0.291	
θ_5	0.617	0.300	0.478	0.312	0.285	0.273	

Table 4: Simulation results with mean-unbalanced cluster constant covariates and unspecified correlation structure. AJS: approximation of the jackknife estimator of variance, FIJ: fully iterated jackknife estimator of variance, 1-SJ: one-step approximation of the jackknife estimator of variance

param.	Theoret. value	Mean param. estimate	Std.err. of the mean	Standard error			
				Prentice	AJS	FIJ	1-SJ
θ_0	0.996	0.254	0.246	0.255	0.255	0.253	
θ_1	3.001	0.215	0.204	0.210	0.207	0.207	
θ_2	-0.198	0.040	0.039	0.041	0.041	0.041	
θ_3	0.630	0.200	0.387	0.240	0.201	0.189	

Table 5: Simulation results with mean-unbalanced cluster constant binary and cluster varying continuous covariate and exchangeable correlation structure. AJS: approximation of the jackknife estimator of variance, FIJ: fully iterated jackknife estimator of variance, 1-SJ: one-step approximation of the jackknife estimator of variance

param.	Theoret. value	Mean param. estimate	Std.err. of the mean	Standard error			
				Prentice	AJS	FIJ	1-SJ
θ_0	1.005	0.266	0.245	0.250	0.254	0.247	
θ_1	3.001	0.215	0.204	0.207	0.207	0.203	
θ_2	-0.200	0.043	0.039	0.040	0.041	0.039	
θ_3	0.435	0.297	0.392	0.287	0.275	0.264	
θ_4	0.946	0.314	0.607	0.369	0.305	0.292	
θ_5	0.662	0.294	0.460	0.317	0.289	0.278	

Table 6: Simulation results with mean-unbalanced cluster constant binary and cluster variing continuous covariate and unspecified correlation structure. AJS: approximation of the jackknife estimator of variance, FIJ: fully iterated jackknife estimator of variance, 1-SJ: one-step approximation of the jackknife estimator of variance

param.	Theoret. value	Mean param. estimate	Std.err. of the mean	Standard error			
				Prentice	AJS	FIJ	1-SJ
θ_0	1	0.988	0.228	0.217	0.222	0.221	0.221
θ_1	3	3.004	0.161	0.160	0.160	0.163	0.160
θ_2	-0.2	-0.199	0.043	0.041	0.042	0.042	0.042
θ_3	0.7	0.651	0.200	0.346	0.238	0.201	0.190

Table 7: Simulation results with mean-unbalanced cluster variing binary and cluster constant continuous covariate and exchangeable correlation structure. AJS: approximation of the jackknife estimator of variance, FIJ: fully iterated jackknife estimator of variance, 1-SJ: one-step approximation of the jackknife estimator of variance

param.	Theoret. value	Mean param. estimate	Std.err. of the mean	Standard error			
				Prentice	AJS	FIJ	1-SJ
θ_0	0.995	0.222	0.215	0.217	0.219	0.215	
θ_1	2.999	0.155	0.146	0.144	0.148	0.144	
θ_2	-0.199	0.043	0.040	0.041	0.041	0.041	
θ_3	0.438	0.279	0.394	0.285	0.275	0.266	
θ_4	0.973	0.341	0.515	0.377	0.310	0.302	
θ_5	0.656	0.287	0.393	0.314	0.287	0.278	

Table 8: Simulation results with mean-unbalanced cluster variing binary and cluster constant continuous covariate and unspecified correlation structure. AJS: approximation of the jackknife estimator of variance, FIJ: fully iterated jackknife estimator of variance, 1-SJ: one-step approximation of the jackknife estimator of variance

param.	Theoret. value	Mean param. estimate	Std.err. of the mean	Standard error			
				Prentice	AJS	FIJ	1-SJ
θ_0	1	1.000	0.245	0.243	0.249	0.250	0.248
θ_1	3	3.000	0.163	0.162	0.164	0.165	0.162
θ_2	-0.2	-0.200	0.040	0.040	0.041	0.041	0.041
θ_3	0.7	0.646	0.204	0.333	0.241	0.202	0.192

Table 9: Simulation results with mean-unbalanced cluster variing covariates and exchangeable correlation structure. AJS: approximation of the jackknife estimator of variance, FIJ: fully interated jackknife estimator of variance, 1-SJ: one-step approximation of the jackknife estimator of variance

param.	Theoret. value	Mean param. estimate	Std.err. of the mean	Standard error			
				Prentice	AJS	FIJ	1-SJ
θ_0	1	0.998	0.244	0.239	0.241	0.246	0.240
θ_1	3	2.997	0.152	0.147	0.147	0.150	0.146
θ_2	-0.2	-0.200	0.040	0.039	0.040	0.041	0.039
θ_3	0.5	0.460	0.280	0.376	0.286	0.274	0.265
θ_4	1	0.974	0.319	0.494	0.376	0.309	0.301
θ_5	0.7	0.681	0.305	0.383	0.322	0.292	0.283

Table 10: Simulation results with mean-unbalanced cluster variing covariates and unspecified correlation structure. AJS: approximation of the jackknife estimator of variance, FIJ: fully iterated jackknife estimator of variance, 1-SJ: one-step approximation of the jackknife estimator of variance

Acknowledgment

The work of C. Kastner is supported by the Deutsche Forschungsgemeinschaft.

References

Fieger, A., Heumann, C., Kastner, C. and Watzka, K. (1997). Generische Bibliothek zur Linearen Algebra und zur Simulation in C++, *SFB386 – Discussion Paper 63*, Ludwig-Maximilians-Universität München.

Hansen, L. (1982). Large sample properties of generalized methods of moments estimators, *Econometrica* **50**: 1029–1055.

Kastner, C., Fieger, A. and Heumann, C. (1997). MAREG and WinMAREG—a tool for marginal regression models, *Computational Statistics and Data Analysis* **24**: 235–241.

Lipsitz, S. R., Dear, K. B. and Zhao, L. P. (1994). Jackknife estimators of variance for parameter estimates from estimating equations with applications to clustered survival data, *Biometrics* **50**: 842–846.

Lipsitz, S. R., 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, Part A—Theory and Methods* **19**: 821–845.

Mancl, L. A. and Leroux, B. G. (1996). Efficiency of regression estimates for clustered data, *Biometrics* **52**: 500–511.

Paik, M. C. (1988). Repeated measurement analysis for nonnormal data in small samples, *Communications in Statistics, Part B—Simulation and Computation* **17**: 1155–1171.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation, *Biometrics* **44**: 1033–1048.

SunOS (1995). *SunOS Reference Manual*, Mountain View.

White, H. (1982). Maximum likelihood estimation of misspecified models, *Econometrica* **50**: 1–25.

Ziegler, A. (1995). The different parameterizations of the GEE1 and the GEE2, *in* G. U. H. Seeber, B. Francis, R. Hatzinger and G. Steckel-Berger (eds), *Statistical Modelling*, Vol. 104 of *Lecture Notes in Statistics*, Springer, New York.

Ziegler, A. (1997). Practical considerations of the jackknife estimator of variance for generalized estimationg equations, *Statistical Papers* **38**: 363–369.

Ziegler, A., Kastner, C. and Blettner, M. (1998). The generalised estimating equations: An annotated bibliography, *Biometrical Journal* **40**: 115–139.