Generalized estimating equations: a hybrid approach for mean parameters in multivariate regression models.

Christoph Lange
Department of Biostatistics, Harvard School of Public Health,
655 Huntington Avenue, Boston, MA 02115, USA

John C. Whittaker
Dept Epidemiology and Public Health,
Imperial College School of Medicine,
St Mary's Campus, Norfolk Place, London W2 1PG, UK

Alex J. MacGregor
Twin Research & Genetic epidemiology Unit
St Thomas' Hospital, London SE1 7EH, UK

January 7, 2003

Abstract

We propose an extension of the generalized estimating equation approach to multivariate regression models (Liang & Zeger, 1986) which allows the estimation of dispersion and association parameters in the covariance matrix partly using estimating equations as in Prentice & Zhao (1991), and partly by the direct use of consistent estimators. The advantages of this hybrid approach over that of Prentice & Zhao (1991) are a reduction in the number of fourth moment assumptions that must be made, and the consequent reduction in numerical complexity. We show that the type of estimation used for covariance parameters does not affect the asymptotic efficiency of the mean parameter estimates. The advantages of the hybrid model are illustrated by a simulation study. This work was motivated by problems in statistical genetics, and we illustrate our approach using a twin study examining association between the osteocalcin receptor and various osteoporisis related traits.
Keywords: GEE, GEE2, association mapping,

1 Introduction

There is great interest in statistical genetics in the investigation of association between marker loci, that is genes of known location and observable genotype, and traits of interest, particularly those related to human disease (see eg Clayton (2001), Zhao (2000), Schulze & McMahon (2002)), but also in plant and animal breeding (eg Jansen and Stam (1994), Zeng (1994), Knott & Haley (2000), Lange & Whittaker 2001). Analysis of univariate responses is usually based on the appropriate generalized linear model, but it is common to observe correlated multivariate responses, which cannot be handled easily unless multivariate normality is reasonable since for multivariate non-normal traits the likelihood function is in general

difficult to specify. This suggests that the generalized estimating equation (GEE) approach developed by Liang & Zeger (1986) for longitudinal data is an ideal tool for these analyses. However, direct application of the Liang & Zeger approach is problematic, since it is often unrealistic to assume the same dispersion parameters and link functions for all components of the response. We could use the extension of GEE to multivariate data due to Prentice & Zhao (1991), but here other difficulties arise. Firstly, the approach is computationally slow. Secondly, while in the original GEE-approach for longitudinal data analysis the researcher had to specify only the first two moment assumptions and to choose consistent estimators for the single dispersion parameter and the association parameters of the correlation matrix, the problem of estimating several dispersion parameters and more complex association structures involved in multivariate data analysis necessitates the introduction of an additional set of estimating equations for the covariance parameters (Prentice & Zhao, 1991). These estimating equations require fourth moment assumptions. Fourth moment assumptions may be feasible in some applications (Prentice & Zhao, 1991), particularly for special cases e.g. correlated binary data (Lipsitz, Laird & Harrington, 1991), but in many situations the researcher may find it difficult to make reasonable assumptions, especially when multivariate data is given with different distributions in each dimension. Furthermore, the use of sophisticated assumptions for second and higher moments is controversial: for example, McCullagh (1992) argued that many data sets would be too small to allow proper estimation of the parameters in such sophisticated assumptions, and that more extensive data sets would be liable to be plagued by outliers, to which estimates would be very sensitive.

In this paper we propose a new approach which allows the estimation of an arbitrary subset of the covariance-parameters directly by consistent estimators, thus removing the need to estimate all covariance-parameters through the estimating equation. This substantially reduces the numerical complexity of the estimation procedure and, more importantly, allows the researcher to choose when to make fourth moment assumptions. As we will show, the standard theory for the asymptotic distribution of GEE-estimators remains valid for this new, 'hybrid' approach. It follows that the asymptotic efficiency of the mean-parameter estimates is not influenced by the type of estimation used for the covariance parameters and so estimating the mean parameters by the hybrid approach is asymptotically as efficient as estimating them by a pure GEE-approach. However, the asymptotic distribution of the covariance-parameter estimates obtained by the consistent estimators is not provided by the GEE-theory. Thus, covariance parameters may not be estimated as efficiently as is possible using GEE with correctly specified fourth moment assumptions.

Taking this to its logical conclusion, it is of course possible to dispense with the fourth moment assumptions entirely. As we will show, such an approach is much less numerically complex than the Prentice & Zhao (1991) approach, and will generally be appropriate where the primary interest is in the mean rather than the covariance parameters. However, when complex covariance structures are assumed and consistent moment-based estimators are hard to derive, the advantages of our approach become redundant and the approaches by Shults & Chaganty (1998) and by Chaganty & Shults (1998) might be considered.

We begin by establishing notation and motivating further our new approach. The estimating equation for our extended model is given, and after a discussion of the relationship with existing work we derive the large sample properties of the resulting estimates. The numerical complexity of our approach is then considered. Application of the proposed methodology is illustrated with a simulation study and by analysis of a study into association between the osteocalcin receptor and the bone disease osteoporisis. We note that, while motivated by problems in statistical genetics, the hybrid GEE-method has potential applications in many areas of statistics.

2 Notation

To establish notation, assume that we observe m outcome variables on n subjects and that the outcome variables from different subjects are independent. The jth outcome variable on the ith subject is denoted by Y_{ij}. Let the vector of random variables $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im})^T$ be the outcome vector and X_i the $m \times p$ matrix of covariate values for the ith subject. For the jth outcome variable the first moment assumption is written as

$$\mu_{ij} \quad = \quad E\left(Y_{ij} \,|X_i\right) \quad = \quad h_j\left(P_j \; X_i \; \beta\right) \tag{2.1}$$

where $P_j \in \mathbb{R}^{1\times m}$ is the projection matrix on the jth dimension, h_j is the link function for the jth outcome variable and β a vector of regression coefficients. For simplicity of exposition, we do not allow the number of observations per subject to vary between subjects, though our results could be extend to this more general situation. The specification of β in multivariate model is illustrated in data simulation section and in the data analysis. Further illustrations can be found in Lange & Whittaker (2001) and Lange et al (2002).

We consider as a motivating example for our second moment assumptions a study examining the relationship between genotype at the osteocalcin microsatellite marker D1S3737 and a number of traits related to the metabolic bone disease osteoporosis in 1366 female dizygotic twins (Andrew *et al*, 2001), analysed in more detail in section 8. Note that we expect correlation between the osteoporosis traits, both within individuals and, because of shared polygenic and environmental factors, between members of the same twin pair, so that an 'individual' here is a twin pair. Our key response variable is bone mineral density (BMD) measured at four sites in the spine. It is difficult to specify even second moment assumptions for these traits, and third or fourth order assumptions would be highly speculative. Furthermore, it is the mean parameters that are of primary interest. We therefore choose a simple correlation structure for these traits, with the corresponding association parameters estimated using moment based estimators; for instance in section 8 we assume an exchangeable correlation structure for both BMD measurements on a single individual and between BMD at the same site in members of the same twin pair.

For other responses, reasonable fourth moment assumptions may be possible, and hence association parameters could be estimated using GEE as in Prentice & Zhao (1991) if desired. In the twins data, for instance, heel ultrasound measurements BUA (broadband ultrasound attenuation) and VOS (velocity of ultrasound), summarised as binary variables where 1 indicates exceedence of the thresholds 76.0 dB/MHz and 1660 m/s respectively, could be modelled as in Lipsitz, Laird & Harrington (1991). However, as above, it will be difficult to make the moment assumptions required to estimate parameters describing the covariance of BMD measurements with these ultrasound measurements, so we would choose to estimate these parameters using moment based estimators.

This example suggests the following model. We split the outcome variables for each subject into $\left(Y_{i1}, \ldots, Y_{im_{(con)}}\right)$, $m_{(con)} < m$, with covariance matrix $\mathbf{V}_{i(11)}$, within which covariance parameters will be estimated via moment based estimators, and $\left(Y_{i(m_{(con)}+1)}, \ldots, Y_{im}\right)$ with covariance matrix $\mathbf{V}_{i(22)}$ within which covariance parameters will be estimated via GEE as in Prentice & Zhao (1991). Parameters describing the covariance structure between $\left(Y_{i1}, \ldots, Y_{im_{(con)}}\right)$ and $\left(Y_{i(m_{(con)}+1)}, \ldots, Y_{im}\right)$ will be estimated using moment based estimators. We indicate covariance parameters estimated by GEE and moment based estimation by $\boldsymbol{\alpha}^{(\boldsymbol{GEE})}$ and $\boldsymbol{\alpha}^{(\boldsymbol{con})}$ respectively. This would suggest a second moment

assumption of the form

$$\begin{aligned} \boldsymbol{Var}\left(\mathbf{Y}_{\boldsymbol{i}}\left|\mathbf{X}_{\boldsymbol{i}}\right.\right) &= \mathbf{V}_{\boldsymbol{i}}\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}, \boldsymbol{\alpha}^{(\boldsymbol{con})}\right) \\ &= \begin{pmatrix} \mathbf{V}_{\boldsymbol{i(11)}}\left(\boldsymbol{\alpha}^{(\boldsymbol{con})}\right) & \mathbf{V}_{\boldsymbol{i(12)}}\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}, \boldsymbol{\alpha}^{(\boldsymbol{con})}\right) \\ \mathbf{V}_{\boldsymbol{i(12)}}\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}, \boldsymbol{\alpha}^{(\boldsymbol{con})}\right)^{\boldsymbol{T}} & \mathbf{V}_{\boldsymbol{i(22)}}\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}, \boldsymbol{\alpha}^{(\boldsymbol{con})}\right) \end{pmatrix} \end{aligned}$$

In fact we generalize this structure in two ways. Firstly, we divide $\alpha^{(con)}$ into two subsets $(\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}, \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})})$, so that we can allow the estimators of $\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}$ to depend on estimates of $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}$. Thus we model the variance of $\left(Y_{i1}, \ldots, Y_{im_{(con)}}\right)$ as

$$\mathbf{V}_{i(11)}\left(\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}, \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right) = \left(\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}\right)^{\frac{1}{2}} \boldsymbol{A}_i^{\frac{1}{2}} R(\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}) \boldsymbol{A}_i^{\frac{1}{2}} \left(\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}\right)^{\frac{1}{2}} \in \mathbb{R}^{m_{(con)} \times m_{(con)}} \quad (2.2)$$

where for $j = 1, \ldots, m_{(con)}$, $V_j(.)$ is the variance function and ϕ_j the dispersion parameter for the jth outcome variable, $\boldsymbol{A}_i = diag(V_1(\mu_{i1}), \ldots, V_{m_{(con)}}(\mu_{im_{(con)}}))$ is the diagonal matrix of the variance functions, $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})} = diag(\phi_1, \ldots, \phi_{m_{(con)}})$ the diagonal matrix of dispersion parameters and $R(\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})})$ a $m_{(con)} \times m_{(con)}$ "working" correlation matrix dependent on the parameter vector $\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}$. Typically, $R(\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})})$ might be assumed to have a simple structure, e.g. exchangeable or unstructured. Note that the estimation of $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}$ and $\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}$ is facilitated by allowing these parameters to be estimated in two stages: first $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}$ is estimated, then the Pearson residuals are calculated and used to estimate $\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}$.

Secondly, we allow $\mathbf{V}_{i(12)}, \mathbf{V}_{i(21)}$ and $\mathbf{V}_{i(22)}$ to depend both on parameters estimated via GEE and via moment based estimation. Putting all this together, $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}$ contains the covariance parameters whose consistent estimates depend on $\mathbf{Y}$ and the GEE estimates of mean parameters, while $\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}$ includes the parameters whose consistent estimates depend additionally on the estimates of $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}$. Writing $\boldsymbol{\alpha}^{(\boldsymbol{GEE})}$ for association parameters estimated via GEE, we can then formally define the second moment of the hybrid approach by the "working" variance matrix $\mathbf{V}_i\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}, \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}, \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right)$

$$\begin{aligned} Var\left(\mathbf{Y}_i\left|\mathbf{X}_i\right.\right) &= \mathbf{V}_i\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}, \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}, \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right) \\ &= \begin{pmatrix} \mathbf{V}_{i(11)}\left(\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}, \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right) & \mathbf{V}_{i(12)}\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}, \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right) \\ \mathbf{V}_{i(12)}\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}, \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}, \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right)^T & \mathbf{V}_{i(22)}\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}, \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}, \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right) \end{pmatrix} \end{aligned} \quad (2.3)$$

3 Estimating equation

We now define the generalized estimating equation for the extended multivariate GEE-model specified by moment assumptions (2.1) and (2.3). We write the elements of the "working" variance matrix $\mathbf{V}_i\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}, \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}, \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right) = (\sigma_{ist})_{s,t=1,\ldots,m}$ that are dependent on the association parameter $\boldsymbol{\alpha}^{(\boldsymbol{GEE})}$ as a column vector, so that

$$\boldsymbol{\sigma}_i^T\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}\right) = \{\ldots, \sigma_{ist}, \ldots\} \in \mathbb{R}^l$$

with $\frac{d\,\sigma_{ist}}{d\,\boldsymbol{\alpha}^{(\boldsymbol{GEE})}} \neq \mathbf{0} \in \mathbb{R}^{dim\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}\right)}$ and $l \leq \frac{1}{2}\left\{m(m+1) - m_{(con)}(m_{(con)}+1)\right\}$.
Similarly, the elements of the "working" variance matrix $\mathbf{V}_i\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}, \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}, \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right)$ that depend on

$\boldsymbol{\alpha_1^{(con)}}$ or/and $\boldsymbol{\alpha_2^{(con)}}$, but not on $\boldsymbol{\alpha^{(GEE)}}$ are written as the column vector

$$\boldsymbol{\sigma}_i^T\left(\boldsymbol{\alpha_1^{(con)}},\boldsymbol{\alpha_2^{(con)}}\right) \;=\; \{\ldots,\sigma_{is't'},\ldots\}\in\mathbb{R}^{m(m+1)/2-l}$$

with $\frac{d\,\sigma_{is't'}}{d\left\{\boldsymbol{\alpha_1^{(con)}},\boldsymbol{\alpha_2^{(con)}}\right\}^T}\neq\mathbf{0}\in\mathbb{R}^{\left[dim\;\;\boldsymbol{\alpha_1^{(con)}}\;\;+dim\;\;\boldsymbol{\alpha_2^{(con)}}\;\right]}$ and $\frac{d\,\sigma_{is't'}}{d\,\boldsymbol{\alpha^{(GEE)}}}=\mathbf{0}\in\mathbb{R}^{dim\left(\boldsymbol{\alpha^{(GEE)}}\right)}$.
The corresponding vectors of empirical covariances are written as $\mathbf{s}_i^T\left(\boldsymbol{\alpha^{(GEE)}}\right) \;=\; \{\ldots,s_{ist},\ldots\}\in\mathbb{R}^l$ and $\mathbf{s}_i^T\left(\boldsymbol{\alpha_1^{(con)}},\boldsymbol{\alpha_2^{(con)}}\right) \;=\; \{\ldots,s_{is't'},\ldots\}\in\mathbb{R}^{m(m+1)/2-l}$respectively. We assume that the "working" variance matrix of the empirical covariances $\mathbf{s}_i\left(\boldsymbol{\alpha^{(GEE)}}\right)$ and $\mathbf{s}_i\left(\boldsymbol{\alpha_1^{(con)}},\boldsymbol{\alpha_2^{(con)}}\right)$ has block-diagonal structure with the off-diagonal matrix equal to $\mathbf{0}$, i.e.

$$\begin{aligned} &Var\left[\left\{\mathbf{s}_i\left(\boldsymbol{\alpha^{(GEE)}}\right),\mathbf{s}_i\left(\boldsymbol{\alpha_1^{(con)}},\boldsymbol{\alpha_2^{(con)}}\right)\right\}^T\right]= \qquad (3.4)\\ &= \begin{pmatrix} Var\left\{\mathbf{s}_i\left(\boldsymbol{\alpha^{(GEE)}}\right)\right\} & \mathbf{0}\\ \mathbf{0} & Var\left\{\mathbf{s}_i\left(\boldsymbol{\alpha_1^{(con)}},\boldsymbol{\alpha_2^{(con)}}\right)\right\}\end{pmatrix}\end{aligned}$$

This partitioning of the "working" variance matrix for the empirical covariances $\mathbf{s}_i\left(\boldsymbol{\alpha^{(GEE)}}\right)$ and $\mathbf{s}_i\left(\boldsymbol{\alpha_1^{(con)}},\boldsymbol{\alpha_2^{(con)}}\right)$ into two independent blocks is natural, because we have assumed that the association parameters $\boldsymbol{\alpha_1^{(con)}}$ and $\boldsymbol{\alpha_2^{(con)}}$ represent parameters that are either of low interest or for which it was difficult to state a second moment assumption. Any non-zero choice for the off-diagonal matrix would contradict these assumptions. An important consequence of the partitioning of the "working" variance matrix in equation (3.4) is a substantial reduction of the dimension of the estimating equation. While in the multivariate approach of Prentice & Zhao (1991) all residuals, $\mathbf{y}_i-\boldsymbol{\mu}_i$, $\mathbf{s}_i\left(\boldsymbol{\alpha^{(GEE)}}\right)-\boldsymbol{\sigma}_i\left(\boldsymbol{\alpha^{(GEE)}}\right)$ and $\mathbf{s}_i\left(\boldsymbol{\alpha_1^{(con)}},\boldsymbol{\alpha_2^{(con)}}\right)-\boldsymbol{\sigma}_i\left(\boldsymbol{\alpha_1^{(con)}},\boldsymbol{\alpha_2^{(con)}}\right)$, appear in the estimating equation, here the estimating equation will contain only the residuals that depend on $\boldsymbol{\beta}$ and $\boldsymbol{\alpha^{(GEE)}}$: $\mathbf{y}_i-\boldsymbol{\mu}_i$ and $\mathbf{s}_i\left(\boldsymbol{\alpha^{(GEE)}}\right)-\boldsymbol{\sigma}_i\left(\boldsymbol{\alpha^{(GEE)}}\right)$. The implications of this for the numerical complexity of the estimation procedure will be discussed in the section 6.

Including parameters describing the association structure of the "working" variance matrix for the empirical covariances $\mathbf{s}_i\left(\boldsymbol{\alpha^{(GEE)}}\right)$ in $\boldsymbol{\alpha_1^{(con)}}$ and $\boldsymbol{\alpha_2^{(con)}}$, writing $\mathbf{W}_i\left(\boldsymbol{\alpha_1^{(con)}},\boldsymbol{\alpha_2^{(con)}}\right)=Var\left\{\mathbf{s}_i\left(\boldsymbol{\alpha^{(GEE)}}\right)\right\}$ and putting

$$\begin{aligned}\boldsymbol{D}_i &= \begin{pmatrix} d\boldsymbol{\mu}_i/d\boldsymbol{\beta} & \mathbf{0}\\ \mathbf{0} & d\boldsymbol{\sigma}_i\left(\boldsymbol{\alpha^{(GEE)}}\right)/d\boldsymbol{\alpha^{(GEE)}}\end{pmatrix}\\ \tilde{\mathbf{V}}_i &= \begin{pmatrix} \mathbf{V}_i\left(\boldsymbol{\alpha^{(GEE)}},\boldsymbol{\alpha_1^{(con)}},\boldsymbol{\alpha_2^{(con)}}\right) & \mathbf{0}\\ \mathbf{0} & \mathbf{W}_i\left(\boldsymbol{\alpha_1^{(con)}},\boldsymbol{\alpha_2^{(con)}}\right)\end{pmatrix}\\ \mathbf{f}_i &= \begin{pmatrix} \mathbf{y}_i & - & \boldsymbol{\mu}_i\\ \mathbf{s}_i\left(\boldsymbol{\alpha^{(GEE)}}\right) & - & \boldsymbol{\sigma}_i\left(\boldsymbol{\alpha^{(GEE)}}\right)\end{pmatrix}\end{aligned}$$

we now define the generalized estimating equation to be

$$\sum_{i=1}^{n}\mathbf{U}_i\left(\boldsymbol{\beta},\boldsymbol{\alpha^{(GEE)}},\boldsymbol{\alpha_1^{(con)}},\boldsymbol{\alpha_2^{(con)}}\right) \;=\; \mathbf{0} \qquad (3.5)$$

where $\mathbf{U}_i\left(\boldsymbol{\beta}, \boldsymbol{\alpha}^{(GEE)}, \boldsymbol{\alpha}_1^{(con)}, \boldsymbol{\alpha}_2^{(con)}\right) = \boldsymbol{D}_i^T\ \tilde{\mathbf{V}}_i^{-1}\ \mathbf{f}_i$. We assume that $\hat{\boldsymbol{\alpha}}_1^{(con)}(Y, \boldsymbol{\beta})$ a $n^{\frac{1}{2}}$-consistent estimator of $\boldsymbol{\alpha}_1^{(con)}$ when $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}^{(GEE)}$ are known and that $\hat{\boldsymbol{\alpha}}_2^{(con)}\left(Y, \boldsymbol{\beta}, \boldsymbol{\alpha}^{(GEE)}, \boldsymbol{\alpha}_1^{(con)}\right)$ is a $n^{\frac{1}{2}}$-consistent estimator of $\boldsymbol{\alpha}_2^{(con)}$ when $\boldsymbol{\beta}$, $\boldsymbol{\alpha}^{(GEE)}$ and $\boldsymbol{\alpha}_1^{(con)}$ are known. Then the GEE-estimator $\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{G}}, \hat{\boldsymbol{\alpha}}_{\boldsymbol{G}}^{(GEE)}\right)^T$ is defined as the solution of

$$\mathbf{U}_i\left[\boldsymbol{\beta}, \boldsymbol{\alpha}^{(GEE)}, \hat{\boldsymbol{\alpha}}_1^{(con)}\left(\boldsymbol{\beta}, \boldsymbol{\alpha}^{(GEE)}\right), \hat{\boldsymbol{\alpha}}_2^{(con)}\left\{\boldsymbol{\beta}, \boldsymbol{\alpha}^{(GEE)}, \hat{\boldsymbol{\alpha}}_1^{(con)}(\boldsymbol{\beta})\right\}\right] = \mathbf{0}. \tag{3.6}$$

4 Relationship to previous work

Since the parameters in the "working" variance matrix $\mathbf{V}_i$ are estimated partly by additional dimensions in the estimating equation (3.6) and partly by two stages of moment based estimation, this estimating equation can be viewed as a hybrid between the original GEE for longitudinal data analysis by Liang & Zeger (1986) and its multivariate extension by Prentice & Zhao (1991). However, it is also an extension and generalization of both approaches.

In Prentice & Zhao (1991) the vector of association parameters of the "working" variance matrix is estimated by consistent estimators in a single step. In the hybrid approach we estimate $\boldsymbol{\alpha}_1^{(con)}$ and $\boldsymbol{\alpha}_2^{(con)}$ in two stages by consistent moment-based estimators. In the first stage we estimate the parameter vector $\boldsymbol{\alpha}_1^{(con)}$, where for example we may have $\boldsymbol{\alpha}_1^{(con)} = \left(\phi_1, \ldots, \phi_{m_{(con)}}\right)$. In the second step we use the estimates of $\boldsymbol{\alpha}_1^{(con)}$ to obtain consistent estimates for $\boldsymbol{\alpha}_2^{(con)}$.

This is in correspondence with the original GEE approach by Liang & Zeger (1986). However, since Liang & Zeger (1986) focused on longitudinal data analysis, they considered variance structures with only one single scalar dispersion parameter, ie $\boldsymbol{\alpha}_1^{(con)} = \phi$ and $Var(Y) = \phi V(\mu)$. Hence the scalar dispersion parameter vanishes in the estimating equation. Therefore, in Liang & Zeger (1986), the estimating equation and consequently the GEE-estimator do not depend directly on the consistent estimation of the scalar dispersion parameter $\boldsymbol{\alpha}_1^{(con)} = \phi$.

In contrast, the current approach allows the entire matrix $\mathbf{V}_i$ to depend on a vector of association parameters $\boldsymbol{\alpha}_1^{(con)}$. As a consequence, provided $\boldsymbol{\alpha}_1^{(con)}$ is of dimension 2 or more, it does not vanish in the estimating equation (3.6) and thus the estimating equation (3.6) and the corresponding GEE-estimator depend directly on the consistent estimate of $\boldsymbol{\alpha}_1^{(con)}$. The proof of the asymptotic properties of the GEE-estimator by Liang & Zeger (1986), and consequently its multivariate extension by Prentice & Zhao (1991), do not cover this dependence of the estimating equations on two groups of consistent estimates, with the estimates of the second group dependent on the estimates of the first. It is therefore necessary to derive the asymptotic properties of the GEE-estimator defined as the solution of the estimating equation (3.6). This will be done in Theorem (5.1).

Note that this two step estimation greatly facilitates the application of the estimating equation approach to multivariate nonnormal data. For example, take $\boldsymbol{\alpha}_1^{(con)}$ to be the dispersion parameters $\left(\phi_1, \ldots, \phi_{m_{(con)}}\right)$ and suppose $\boldsymbol{\alpha}_2^{(con)}$ contains all the relevent correlation parameters. There are many situations where we might find it difficult to derive directly consistent moment based estimators for all relevant association parameters, but it is easy to estimate the dispersion parameters, $\left(\phi_1, \ldots, \phi_{m_{(con)}}\right)$, which allows us to compute the Pearson residuals and then estimate $\boldsymbol{\alpha}_2^{(con)}$ by consistent estimators. For relatively simple association structures, eg exchangeable, unstructured, block structures, etc., potential estimators for $\boldsymbol{\alpha}_2^{(con)}$ are discussed in Liang & Zeger (1986). When time series are analyzed and autoregressive correlation structures are assumed consistent estimators for $\boldsymbol{\alpha}_2^{(con)}$ can be found in Brockwell

& Davis (1991). For more sophisticated correlation structures Zeger (1988) illustrated how methods of moments can used to derive consistent estimators for $\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}$.

Of course, we could avoid the additional dependence on $\hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}$ noted above by including the association parameters $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}$ with $\boldsymbol{\alpha}^{(\boldsymbol{GEE})}$ and estimating both $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}$ with $\boldsymbol{\alpha}^{(\boldsymbol{GEE})}$ via the estimating equation. However, there are good reasons for not doing this. Firstly, if we include $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}$ in the estimating equation, we have to specify all fourth moment assumptions, and all empirical variances and covariances will have to be included in the estimating equation. A reduction of the dimension of the estimating equation and consequently of its numerical complexity will therefore not be possible. Secondly, we may prefer moment based estimates for $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}$ to the estimating equation estimates. For example, when we have i.i.d. normally distributed data and we estimate $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})} = \phi$ via the estimating equation, the GEE-estimate for ϕ is given by $\frac{1}{nm}\sum (y_{ij} - \mu_{ij})^2$; we might prefer to estimate ϕ by $\frac{1}{nm-p}\sum (y_{ij} - \mu_{ij})^2$.

When the primary interest of the data analysis is the estimation of the mean parameters, Prentice & Zhao (1991) proposed to simplify their approach by assuming a diagonal structure for $\mathbf{W}_i$. This simplification can be seen as special case of the hybrid model with $dim\left\{(\mathbf{s}_i\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}\right)\right\} = o(m)$. and therefore the above results on the asymptotic efficiency and numerical complexity of the hybrid model apply also to the simplified approach by Prentice & Zhao (1991).

5 Large sample properties

Here we derive the large sample property of the hybrid GEE-estimates defined by equation (3.6).

Theorem 5.1. *Under mild regularity conditions and given that:*

- $\hat{\boldsymbol{\alpha}}_{\mathbf{2}}^{(\boldsymbol{con})}$ *is* $n^{\frac{1}{2}}$*-consistent given* $\boldsymbol{\beta}$, $\boldsymbol{\alpha}^{(\boldsymbol{GEE})}$ *and* $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}$
- $\hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}$ *is* $n^{\frac{1}{2}}$*-consistent given* $\boldsymbol{\beta}$ *and* $\boldsymbol{\alpha}^{(\boldsymbol{GEE})}$
- $\left\|\boldsymbol{\partial}\hat{\boldsymbol{\alpha}}_{\mathbf{2}}^{(\boldsymbol{con})}\left(\boldsymbol{\beta}, \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}\right) / \boldsymbol{\partial}\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}\right\| \leq H(Y, \boldsymbol{\beta})$ *which is* $O_{m_{(con)}}(1)$,

then $n^{\frac{1}{2}}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{G}} - \boldsymbol{\beta}\right)$ *and* $n^{\frac{1}{2}}\left(\hat{\boldsymbol{\alpha}}_{\boldsymbol{G}}^{(\boldsymbol{GEE})} - \boldsymbol{\alpha}^{(\boldsymbol{GEE})}\right)$ *are asymptotically multivariate Gaussian with mean zero and covariance matrix* $\mathbf{V}^{\hat{\boldsymbol{\beta}}_{\boldsymbol{G}}}$ *and* $\mathbf{V}^{\hat{\boldsymbol{\alpha}}_{\boldsymbol{G}}}$ *given by*

$$\begin{aligned}
\mathbf{V}^{\hat{\boldsymbol{\beta}}_{\boldsymbol{G}}} &= \lim_{n\to\infty} n \left[\left(\sum_{i=1}^{n} \boldsymbol{D}_{i(11)}^T \mathbf{V}_i^{-1} \boldsymbol{D}_{i(11)}\right)^{-1}\right. \qquad (5.7)\\
&\times \left.\left\{\sum_{i=1}^{n} \boldsymbol{D}_{i(11)}^T \mathbf{V}_i^{-1} Cov\left(\mathbf{Y}_i\right) \mathbf{V}_i^{-1} \boldsymbol{D}_{i(11)}\right\}\left(\sum_{i=1}^{n} \boldsymbol{D}_{i(11)}^T \mathbf{V}_i^{-1} \boldsymbol{D}_{i(11)}\right)^{-1}\right]\\
\mathbf{V}^{\hat{\boldsymbol{\alpha}}_{\boldsymbol{G}}} &= \lim_{n\to\infty} n \left[\left(\sum_{i=1}^{n} \boldsymbol{D}_{i(22)}^T \mathbf{W}_i^{-1} \boldsymbol{D}_{i(22)}\right)^{-1}\right. \qquad (5.8)\\
&\times \left.\left\{\sum_{i=1}^{n} \boldsymbol{D}_{i(22)}^T \mathbf{W}_i^{-1} Cov\left(\mathbf{s}_i\right) \mathbf{W}_i^{-1} \boldsymbol{D}_{i(22)}\right\}\left(\sum_{i=1}^{n} \boldsymbol{D}_{i(22)}^T \mathbf{W}_i^{-1} \boldsymbol{D}_{i(22)}\right)^{-1}\right]
\end{aligned}$$

with $\boldsymbol{D}_{i(11)} = d\boldsymbol{\mu}_i/d\boldsymbol{\beta}$ *and* $\boldsymbol{D}_{i(22)} = d\boldsymbol{\sigma}_i/d\boldsymbol{\alpha}^{(\boldsymbol{GEE})}$

The proof is based on Liang & Zeger (1986) and Prentice & Zhao (1991) and is sketched in Appendix I. As in Liang & Zeger (1986) and Prentice & Zhao (1991), the consistency and the asymptotic normality of $\hat{\boldsymbol{\beta}}_{\boldsymbol{G}}$ and $\mathbf{V}^{\hat{\boldsymbol{\beta}}_{\boldsymbol{G}}}$ depend only on the correct specification of the mean and not on the correct specification of the association structure in equation (2.3). Therefore the estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{G}}$ remains robust against miss-specification of the variance structure and its standard errors $\mathbf{V}^{\hat{\boldsymbol{\beta}}_{\boldsymbol{G}}}$ can be estimated consistently by (5.8) regardless of the correct specification of the variance assumption, $\mathbf{V}_i = Var(\mathbf{Y_i}|\mathbf{X}_i)$. Corresponding results hold for $\hat{\boldsymbol{\alpha}}_{\boldsymbol{G}}^{(\boldsymbol{GEE})}$.

Consequently the asymptotic distribution of the GEE-estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{G}}$ depends neither on the precision of the GEE-estimate for the covariance parameter vector $\boldsymbol{\alpha}^{(\boldsymbol{GEE})}$ nor on the precision of the consistent moment-based estimators for the covariance parameters $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}$ and $\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}$. The asymptotic efficiency of the GEE-estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{G}}$ is not affected by whether GEE-estimation or moment estimation is used to estimate the parameters in the variance matrix $\mathbf{V}_i$. Thus when our main interest is in the mean parameter $\boldsymbol{\beta}$, so that the variance matrix is primarily specified to increase the efficiency of estimation of $\boldsymbol{\beta}$, consistent moment based estimators should be used in preference to GEE estimation where possible as this minimizes the number of fourth moment assumptions required. Use of moment based estimators has the additional advantage of decreasing the computational complexity considerably, as we show below.

As in Liang & Zeger (1986) and Prentice & Zhao (1991) we compute $\hat{\boldsymbol{\beta}}_{\boldsymbol{G}}$ and $\hat{\boldsymbol{\alpha}}_{\boldsymbol{G}}^{(\boldsymbol{GEE})}$ by iterating between a modified Fisher scoring for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}^{(\boldsymbol{GEE})}$ and moment estimation of $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}$ and $\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}$. Given the current estimate for $\hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}$ and $\hat{\boldsymbol{\alpha}}_{\mathbf{2}}^{(\boldsymbol{con})}$ in the kth. iteration step we obtain the next estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}^{(\boldsymbol{GEE})}$ by the following iteration step:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{k+1} \;=\; \hat{\boldsymbol{\beta}}_k \quad &+ \quad \left\{\sum_{i=1}^{n} \mathbf{D}_{i(11)}^T\left(\hat{\boldsymbol{\beta}}_k\right)\mathbf{V}_i^{-1}\left(\hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\alpha}}^{(\boldsymbol{GEE})}{}_i\right)\boldsymbol{D}_{i(11)}\left(\hat{\boldsymbol{\beta}}_k\right)\right\}^{-1} \\ &\times \quad \left[\sum_{i=1}^{n} \mathbf{D}_{i(11)}^T\left(\hat{\boldsymbol{\beta}}_k\right)\mathbf{V}_i^{-1}\left(\hat{\boldsymbol{\beta}}_k, \hat{\boldsymbol{\alpha}}^{(\boldsymbol{GEE})}{}_i\right)\left\{\mathbf{Y}_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}_k)\right\}\right]\end{aligned} \tag{5.9}$$

$$\begin{aligned}\hat{\boldsymbol{\alpha}}^{(\boldsymbol{GEE})}{}_{k+1} \;=\; \hat{\boldsymbol{\alpha}}^{(\boldsymbol{GEE})}{}_k \quad &+ \quad \left\{\sum_{i=1}^{n} \mathbf{D}_{i(22)}^T\left(\hat{\boldsymbol{\theta}}_k\right)\mathbf{W}_i^{-1}\left(\hat{\boldsymbol{\theta}}_k\right)\boldsymbol{D}_{i(22)}\left(\hat{\boldsymbol{\theta}}_k\right)\right\}^{-1} \\ &\times \quad \left[\sum_{i=1}^{n} \mathbf{D}_{i(22)}^T\left(\hat{\boldsymbol{\theta}}_k\right)\mathbf{W}_i^{-1}\left(\hat{\boldsymbol{\theta}}_k\right)\left\{\mathbf{s}_i\left(\hat{\boldsymbol{\theta}}_k\right) - \boldsymbol{\sigma}_i(\hat{\boldsymbol{\theta}}_k)\right\}\right]\end{aligned} \tag{5.10}$$

where $\mathbf{V}_i\left(\boldsymbol{\beta}, \boldsymbol{\alpha}^{(\boldsymbol{GEE})}\right) = \mathbf{V}_i\left[\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta}), \hat{\boldsymbol{\alpha}}_{\mathbf{2}}^{(\boldsymbol{con})}\left\{\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta})\right\}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta})\right]$,
$\mathbf{W}_i(\boldsymbol{\theta}) = \mathbf{W}_i\left[\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta}), \hat{\boldsymbol{\alpha}}_{\mathbf{2}}^{(\boldsymbol{con})}\left\{\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta}_i)\right\}\right]$ and $\boldsymbol{\theta}^T = \left\{\boldsymbol{\beta}^T, \left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}\right)^T\right\}$

Here we have, in comparison with the standard GEE-approach, an additional direct dependence of $\hat{\boldsymbol{\beta}}_{k+1}$ on the parameter estimates $\hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}$ and an additional indirect dependence of $\hat{\boldsymbol{\alpha}}^{(\boldsymbol{GEE})}{}_{k+1}$ on $\hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}$.

We can take the above approach to its limit by estimating all covariance parameters using moment based estimators. Then we need only to make qualitative second moment assumptions, e.g. to assume an unstructured, exchangeable or band-correlation covariance matrix. This gives the following "ad-hoc"

model for nonnormal, multivariate data sets. Given the observation vector $\mathbf{y_i} = (y_{i1}, \ldots, y_{im})$ the mean assumption is

$$\mu_{ij} \quad = \quad E\left(Y_{ij}\,|\mathbf{X}_i\right) \;=\; h_i\left(P_i\; X_i\;\beta\right) \tag{5.11}$$

and the second moment assumption

$$\begin{aligned} Var\left(\mathbf{Y}_i\,|\mathbf{X}_i\right) \quad &= \quad \mathbf{V}_i\left(\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}, \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right) \\ &= \quad \boldsymbol{\Phi}^{\frac{1}{2}}\;\boldsymbol{A}_i^{\frac{1}{2}}\;R(\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})})\;\boldsymbol{A}_i^{\frac{1}{2}}\;\boldsymbol{\Phi}^{\frac{1}{2}} \;\in\; \mathbb{R}^{m_{(con)}\times m_{(con)}} \end{aligned} \tag{5.12}$$

where $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})} = \boldsymbol{\Phi} = \left(\phi_1, \ldots, \phi_{m_{(con)}}\right)$, $R(\boldsymbol{\alpha})$ is a "simple" correlation matrix, for which $\boldsymbol{\alpha}$ can be estimated directly by consistent estimators using the scaled residuals. The assumption of a "simple" correlation matrix should not be too restrictive, since for many types of correlation matrixes these consistent estimators are known. Further note that the model proposed in the equations (5.11) and (5.12) is not covered by the longitudinal GEE-model proposed by Liang & Zeger (1986), because of the additional dependence of the estimating equation on the consistent estimator of $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})} = \boldsymbol{\Phi}$.

6 Numerical Complexity

We have presented a hybrid GEE-approach for multivariate data which allows us to estimate an arbitrary part of the covariance-parameters using an additional estimating equation for the empirical covariances, while estimating the remaining covariance parameters using consistent estimators. Theorem (5.1) shows that the asymptotic distribution of the GEE-estimates for the mean-parameters is unaffected by the method used to estimate the covariance parameters. However, the more covariance parameters are estimated using consistent estimators the fewer fourth moment estimators are required, and the lower the numerical complexity of the estimation procedure is, as the following theorem shows.

Theorem 6.1 (Relative Numerical Complexity). *With*

$$dim\left\{\mathbf{s}_i\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}\right)\right\} \;=\; o\left(m^c\right) \qquad where \qquad c \;=\; 0, 1, 2$$

and denoting the numerical complexity of iteration step (5.9) and (5.10) for the standard GEE-approach for multivariate data (where all association parameters are estimated by the estimating equation) by $\mathcal{C}_{GEE}$ and for the hybrid-approach by $\mathcal{C}_{hybrid}$, then the relative numerical complexity comparing the standard multivariate GEE-approach with the hybrid-approach is given by

$$\mathcal{C}_{GEE}/\mathcal{C}_{hybrid} \;=\; O\left\{m^{min(6-3c,3)}\right\} \tag{6.13}$$

If additionally $dim\left\{\mathbf{s}_i\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}\right)\right\} \;=\; \kappa m^2 + O\left(m\right)$ with $\kappa \in (0, 0.5)$ and $dim\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}\right) \;=\; O\left(m\right)$ then

$$\lim_{m\to\infty}\;\left(\mathcal{C}_{GEE}/\mathcal{C}_{hybrid}\right) \quad = \quad \left(2\kappa\right)^{-3} \tag{6.14}$$

The proof is given in appendix II. To demonstrate the practical implications of theorem (6.1) we consider two hypothetical examples. First assume that we observe for each individual M time series of count data with a constant length of 4 observations. The association structure within each time series is assumed to

be the same (e.g. AR(2)), but the parameter vector of the association structure may have different values for each series. The association structure between the time series is modeled by band-correlation matrixes. While the association structure within time series is one of the points of interest, the association structure between the time series is of no intrinsic interest and is only considered to improve the efficiency of the GEE-estimates for the mean-parameters. We will therefore use consistent estimators for the relevant covariance parameters in the hybrid approach. Then with $m = 4M$ and $dim\left\{\mathbf{s}_i\left(\alpha(GEE)\right)\right\} = O\left(m\right)$ the relative numerical complexity $\mathcal{C}_{GEE}/\mathcal{C}_{hybrid}$ increases with $O\left(m^3\right)$. When we analyse data sets with many observations per individual or when we have to consider a large number of models this substantially lower numerical complexity of the hybrid approach can be an important advantage.

Consider now the reversed situation. We observe for each individual 4 time series with length M. Again the association structure within all time series is considered to be the same (e.g. AR(2)), but the parameter vector of the association structure may have different values for each time series and is to be estimated by the estimating equation, while the between-association structure has nuisance character. With $m = 4M$ it easy to see that $dim\left(\alpha(GEE)\right) = O\left(m\right)$ and $dim\left\{\mathbf{s}_i\left(\alpha(GEE)\right)\right\} = m^2/8 + O\left(m\right)$. Therefore the relative numerical complexity $\mathcal{C}_{GEE}/\mathcal{C}_{hybrid}$ is approximately $(2/8)^{-3} = 64$ for sufficiently large m. Here the difference between the two approaches is not as dramatic as for the previous example, but is still substantial.

In practice we have found moment based estimation of covariance parameters to have much better numerical properties than estimation by including these parameters in the GEE. Great care needs to be taken with the Prentice & Zhao (1991) approach, and in our experience this leads to actual differences in computing time exceeding that predicted by our theoretical results. These theoretical results are therefore best taken as representing lower bounds for the differences between the approaches.

Finally, note that the model in which all covariance parameters are estimated using moment based estimators which we discussed above has numerical complexity equivalent to the original GEE-approach for longitudinal data by Liang & Zeger (1986) and $O(n^{-3})$-times the complexity of the GEE-model for multivariate data proposed by Prentice & Zhao (1991).

7 Simulation experiment

Quasi-likelihood regression models for count data with dependent observations have previously been discussed by Zeger (1988). Here we consider a similar situation related to the twin study for osteoporisis introduced above and discussed further in the next section. We assume that we observe for each twin two count variables where the counts for the first twin of the ith family are denoted by Y_{i1} and Y_{i2} and for the second twin of the ith family by Y_{i3} and Y_{i4}. Further, both twins are exposed to correlated environmental effects, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2})$ with

$$E(\epsilon_{ij}) = 1, Var(\epsilon_{i1}) = \sigma^2, Var(\epsilon_{i2}) = \sigma^2 \text{ and } Cov(\epsilon_{i1}, \epsilon_{21}) = \sigma^2\, r \tag{7.15}$$

with $|r| < 1$. Conditional on the environmental effect vector $\boldsymbol{\epsilon}_i$, we assume that the environmental effect is same for both phenotypes of one twin and so the mean and variance of $\mathbf{Y}_i = Y_{i1}, \ldots, Y_{i4}$ are given by

$$E(Y_{ij}\,|\boldsymbol{\epsilon}_i\,) \;=\; exp\left(\mathbf{x}_{ij}^T\,\boldsymbol{\beta}_j\right)\left(\;\epsilon_{i1}\;\mathbf{1}_{\{j\leq 2\}} \;+\; \epsilon_{i2}\;\mathbf{1}_{\{j>2\}}\right) \;=\; u_{ij}, \quad \text{and} \quad Var(Y_{ij}\,|\boldsymbol{\epsilon}_i\,) \;=\; u_{ij} \tag{7.16}$$

The marginal moments are then

$$\mu_{ij} = E\left(Y_{ij}\right) = exp\left(\mathbf{x}_{ij}^T\,\boldsymbol{\beta}_j\right), \quad Var\left(Y_{ij}\right) = \mu_{ij} + \sigma^2\mu_{ij}^2 \tag{7.17}$$

$$j \neq j': \quad Cov\left(Y_{ij}, Y_{ij'}\right) \;=\; \mu_{ij}\mu_{ij'}\sigma^2\left(\mathbf{1}_{\{(j-2.5)(j'-2.5)>0\}} + r\mathbf{1}_{\{(j-2.5)(j'-2.5)<0\}}\right) \tag{7.18}$$

where $\mathbf{x}_{ij}$ is the vector of predictor variables for the jth trait in the ith offspring. Note that the indicator variable $\mathbf{1}_{\{(j-2.5)(j'-2.5)>0\}}$ is 1 when j and j' correspond to the same twin, ie $j, j' = 1, 2$ or $j, j' = 3, 4$. $r\mathbf{1}_{\{(j-2.5)(j'-2.5)<0\}}$ describes therefore the correlation between twins. Since σ^2 does not vanish in the estimating equations (variance structure (7.17)), the original GEE-approach by Liang & Zeger (1986) can only be applied when the "working" covariance matrix is assumed to be

$$j \neq j': \quad Cov\left(Y_{ij}, Y_{ij'}\right) \;=\; \mu_{ij}\mu_{ij'}\sigma^2\mathbf{1}_{\{(j-2.5)(j'-2.5)>0\}}, \tag{7.19}$$

that is when the association between the pairs of counts is ignored. In contrast, both the multivariate approach by Prentice & Zhao and the hybrid approach proposed here can cope with the "working" covariance matrix (7.18). Note that in the notation of the hybrid approach introduced above, σ^2 corresponds to $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}$ and r to $\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}$. As in Zeger (1988), σ^2 and r can be estimated in each updating step by

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{4}\sum_{j=1}^{4}\frac{\sum\limits_i \left(y_{ij}-\hat{\mu}_{ij}\right)^2 - \hat{\mu}_{ij}}{\sum\limits_i \hat{\mu}_{ij}^2} \\
\hat{r} &= \frac{1}{4}\sum_{(j,j')\in\left\{\begin{smallmatrix}(1,3),(1,4),\\(2,3),(1,4)\end{smallmatrix}\right\}}\frac{\sum\limits_i (y_{ij}-\hat{\mu}_{ij})(y_{ij'}-\hat{\mu}_{ij'})}{\sum\limits_i \hat{\mu}_{ij}\hat{\mu}_{ij'}}
\end{aligned}$$

We now compare the efficiency of the above discussed GEE-approaches by simulation. Although the approach by Prentice & Zhao requires fourth moment assumptions for the above discussed model for count data and is about $O(2^4) = O(16)$-times more numerically complex than the hybrid approach, it has the same asymptotic efficiency as the hybrid approach. We therefore consider only the efficiency of the original Liang & Zeger GEE-approach and the hybrid approach. The vector of counts of model (7.16) are simulated by generating the underlying and unobservable bivariate random variable $(\epsilon_{i1}, \epsilon_{i2})$ using a multivariate normal distribution with (7.15) and then using

$$E\left(Y_{ij}\,|\epsilon_{ij}\right) = \epsilon_{ij}\; exp\left(x'_{ij}\beta_j\right) = u_{ij} \quad var\left(Y_t\,|\epsilon_t\right) \;=\; u_t \tag{7.20}$$

to generate the target count process via a Poisson distribution with $\lambda = u_{ij}$. The predictor vector $x_{ij} = (1, x_{ij1}, x_{ij2})$ contains an intercept and two uncorrelated predictor variables, x_{ij1} and x_{ij2}, generated by normal distributions with mean 0 and standard error 1. The mean parameters are given by $\beta_1 = (6, 0.05, 0.05)$, $\beta_2 = (6, 0.05, -0.05)$, $\beta_3 = (6, -0.05, 0.05)$ and $\beta_4 = (6, -0.05, -0.05)$. The simulation experiment was conducted for correlation $r = 0.0, 0.3, 0.6, 0.9$ and sample sizes $50, 75, 100, 200, 500, 1000$. For each setup 2000 replicates were simulated. Table (1) shows the relative efficiencies of the hybrid approach and the Liang & Zeger GEE-approach.

As we would expect, Table (1) shows that for moderate correlation (0.0-0.3) only a minor increase in efficiency can be observed when the hybrid approach is used instead of the original approach by Liang & Zeger (1986). However, for correlation greater than 0.3 the relative efficiency of the hybrid approach increases noticeably. It is also important to note that the increase in efficiency is higher for smaller sample sizes than for large samples.

8 Data analysis: Osteoporosis study

We consider the data on spine BMD relating to the same analytical problem described in Andrew et al *et al* (2001). We denote the responses for the ith twin pair by $Y_{i1}, Y_{i2}, \ldots Y_{i8}$ where $Y_{i1}, \ldots, Y_{i4}$ are BMD at the sites one to four in the first twin and $Y_{i5}, \ldots, Y_{i8}$ are the corresponding measurements in the second twin. Ordering of the twins is of course arbitrary. Our objective is to investigate association between this vector of traits and genotype at the D1S3737 marker, here reduced for the sake of simplicity to the single explanatory variable x_{ij}, which counts the number of transmitted "9"-alleles, ie $x_{ij} = 0, 1, 2$. We analyse only the post-menopausal group, since it is here that any genotypic effects would be expected to become apparent.

Standard practice would be to perform univariate analyses of the BMD response, as in Andrew *et al* (2001) or to treat these data as multivariate normal, probably after marginal transformation. However, even after transformationwe wish to avoid the normality assumption, so a robust multivariate approach via GEE is preferred here. The BMD are densities, hence we selected an inverse link-function for the mean. Then the first moment assumption for the first twin is given by

$$E\left(\frac{1}{Y_{ij}}\right) \quad = \quad \beta_{0j} + \beta_j \, x_{i1}, \quad j = 1, \ldots, 4$$

and for the second twin by

$$E\left(\frac{1}{Y_{i(j+4)}}\right) \quad = \quad \beta_{0j} + \beta_j \, x_{i2}, \quad j = 1, \ldots, 4$$

with Y_{ij} as defined above. Since it is not sensible to make sophisticated assumptions about the correlation structure between the four phenotypes, and any fourth moment assumptions would be highly dubious, we analyzed the data set by using the "ad-hoc" model, equation (5.11) and (5.12).

For the covariance structure we assume 4 different dispersion parameters for each BMD, ie $\phi_j = \phi_{j+4}, j = 1, .., 4$, and an exchangeable correlation matrix within each twin, where the "within-subject" correlation parameter is denoted by α_w. Further, the correlation between members of the same twin pair is also assumed to be exchangeable, with correlation parameter α_b. So our "working" variance-covariance assumption is given by

$$\begin{aligned} Var(Y_{ij}) &= \phi_j \, \mu_{ij}^2 \\ Cov(Y_{ij}, Y_{ij'}) &= \mu_{ij}\mu_{ij'} \left(\alpha_w 1_{(j-4.5)(j'-4.5)>0} + \alpha_b 1_{(j-4.5)(j'-4.5)<0}\right) \end{aligned}$$

The variance parameters are estimated in a two step procedure. First, for each BMD a dispersion is fitted by using the residuals, then the dispersion parameters are used to compute the Pearson residuals and based on them the correlation parameters α_w and α_b. That is, each updating step $\phi_j, j = 1, .., 4$, α_w and α_b are estimated by

$$\begin{aligned} \hat{\phi}_j &= \frac{1}{2n} \sum_i \left\{ r_{ij}^2 + r_{i(j+4)}^2 \right\}, \\ \hat{\alpha}_w &= \frac{1}{12n} \sum_i \sum_{j=1}^{4} \sum_{j'=j+1}^{4} \left\{ r_{ij} r_{ij'} + r_{i(j+4)} r_{i(j'+4)} \right\} \\ \hat{\alpha}_b &= \frac{1}{16n} \sum_i \sum_{j=1}^{4} \sum_{j'=5}^{8} r_{ij} r_{ij'}, \end{aligned}$$

where $r_{ij} = (y_{ij} - \hat{\mu}_{ij})/\sqrt{V(\hat{\mu}_{ij})}$ is the Pearson residual. The results of this analysis are shown in Table 2. It is important to note that the "within-subject" correlation α_w is estimated to be 0.87. The simulation study in the previous section showed that the efficiency gain of the hybrid approach over the original GEE-approach is of practical relevance for such correlations.

A alternative GEE approach for family data is discussed in Lange et al (2002).

9 DISCUSSION

We have extended the work of Liang & Zeger (1986) and Prentice & Zhao (1991) as follows. Instead of a single link-function h and a single dispersion parameter ϕ as in Liang & Zeger (1986)), we have allowed each outcome variable to have a different link functions h_j (mean equation (2.1)) and over-dispersion parameter, e.g. $\boldsymbol{\alpha}_1^{(con)} = (\phi_1, \ldots, \phi_{m^{(con)}})$ (variance equation (2.3)). In contrast to Prentice & Zhao (1991), where the parameters of the "working" variance matrix were estimated by additional dimensions in the generalized estimating equation, we allow some of the association parameters to be estimated using moment based estimators via a two stage process where the second moment based estimator, $\hat{\boldsymbol{\alpha}}_2^{(con)}$, may depend on the estimates of the first, $\hat{\boldsymbol{\alpha}}_1^{(con)}$. These modifications make it possible to model multivariate non-Gaussian data easily. The standard GEE asymptotic results are shown to hold for this new approach. The key advantage of the hybrid approach is that we can choose the parameters for which we are willing to make the fourth moment assumptions required for full GEE estimation. Covariance parameters for which we are unable or unwilling to make these assumptions can be estimated using consistent moment based estimators without any loss in the asymptotic efficiency with which the mean parameters are estimated. This leads to a considerable reduction in computational complexity and to increased numerical stability: our experience suggests that optimal computational stability is obtained by estimating as many covariance parameters as possible via moment based estimators rather than including these parameters in the generalized estimating equation.

The ability of the hybrid approach to model more sophisticated working variance structures than is allowed by the original Liang & Zeger approach, without the need to make additional fourth moment assumptions, is illustrated by the simulation study. This can lead to a noticeable increase of efficiency, especially when the sample size is small. The hybrid approach may thus allow the problems of low efficiency of GEE estimation often reported for small sample sizes to be avoided.

Software: The Splus function implementing the 'ad-hoc' method and the data set are available on the web-page of the *Statistical Modelling*.

10 ACKNOWLEDGMENTS

We wish to thank Professor Geert Molenberghs, Professor David Balding, Dr Mike Denham and Dr. John Reeves for their helpful comments on an earlier version of the paper, and the Twin Research and Genetic epidemiology Unit, St Thomas' Hospital, London for providing the osteocalcin data. Perceptive and constructive comments from two referees and an associate editor were very helpful in preparing this version of the paper. This research was supported in parts by grant MH59532 and in parts by grant HL66383 of The National Institutes of Health.

11 Appendix: Proof of Theorem (5.1)

As above suppose that $\mu_i = \mu_i(\boldsymbol{\beta})$ and $\boldsymbol{\sigma}_i\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}\right) = \boldsymbol{\sigma}_i\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}, \boldsymbol{\beta}\right)$. Let $\boldsymbol{\theta}^T = \left\{\boldsymbol{\beta}^T, \left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}\right)^T\right\}$ and let $\hat{\boldsymbol{\theta}}^T = \left\{\hat{\boldsymbol{\beta}}^T, \left(\hat{\boldsymbol{\alpha}}^{(\boldsymbol{GEE})}\right)^T\right\}$ solve the estimating equation of the form (3.6) with weight matrix $\tilde{\mathbf{V}}_i = \tilde{\mathbf{V}}_i\left[\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta}), \hat{\boldsymbol{\alpha}}_{\mathbf{2}}^{(\boldsymbol{con})}\left\{\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta})\right\}\right]$. Write $\boldsymbol{\alpha}^*(\boldsymbol{\theta}) = \hat{\boldsymbol{\alpha}}_{\mathbf{2}}^{(\boldsymbol{con})}\left\{\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta})\right\}$ and under some regularity conditions $n^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)$ can be approximated

$$\text{by} \quad \left[\sum_{i=1}^{n} -\frac{\delta}{\delta\boldsymbol{\theta}}\mathbf{U}_i\left\{\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta}), \boldsymbol{\alpha}^*(\boldsymbol{\theta})\right\}/n\right]^{-1}\left[\sum_{i=1}^{n}\mathbf{U}_i\left\{\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta}), \boldsymbol{\alpha}^*(\boldsymbol{\theta})\right\}/n^{\frac{1}{2}}\right]$$

$$\begin{aligned}
\text{where} \quad \frac{\delta}{\delta\boldsymbol{\theta}}\mathbf{U}_i\left\{\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta}), \boldsymbol{\alpha}^*(\boldsymbol{\theta})\right\} &= \frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{U}_i\left\{\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta}), \boldsymbol{\alpha}^*(\boldsymbol{\theta})\right\} \\
&+ \frac{\partial}{\partial\boldsymbol{\alpha}^*}\mathbf{U}_i\left\{\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta}), \boldsymbol{\alpha}^*(\boldsymbol{\theta})\right\}\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\alpha}^*(\boldsymbol{\theta})\right\} \\
&+ \frac{\partial}{\partial\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}}\mathbf{U}_i\left\{\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta}), \boldsymbol{\alpha}^*(\boldsymbol{\theta})\right\}\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta})\right\} \\
&= \boldsymbol{\mathcal{A}}_i + \mathcal{B}_i\mathcal{C} + \boldsymbol{\mathcal{D}}_i\,\mathcal{E}
\end{aligned}$$

Let $\boldsymbol{\theta}$ be fixed; Taylor expansion gives

$$\begin{aligned}
\frac{\sum_{i=1}^{n}\mathbf{U}_i\left\{\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta}), \boldsymbol{\alpha}^*(\boldsymbol{\theta})\right\}}{n^{\frac{1}{2}}} &= \frac{\sum_{i=1}^{n}\mathbf{U}_i\left(\boldsymbol{\theta}, \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}, \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right)}{n^{\frac{1}{2}}} \\
&+ \frac{\sum_{i=1}^{n}\frac{\boldsymbol{\partial}}{\boldsymbol{\partial}\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}}\mathbf{U}_i\left(\boldsymbol{\theta}, \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}, \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right)}{n}n^{\frac{1}{2}}\left\{\boldsymbol{\alpha}^*(\boldsymbol{\theta}) - \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right\} \\
&+ \frac{\sum_{i=1}^{n}\frac{\boldsymbol{\partial}}{\boldsymbol{\partial}\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}}\mathbf{U}_i\left(\boldsymbol{\theta}, \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}, \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right)}{n}n^{\frac{1}{2}}\left\{\hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta}) - \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}\right\} + o(\mathbf{1}) \\
&= \mathcal{A}^* + \mathcal{B}^*\,\mathcal{C}^* + \mathcal{D}^*\mathcal{E}^* + o(\mathbf{1})
\end{aligned}$$

Now, $\mathcal{B}^* = o(\mathbf{1})$ and $\mathcal{D}_i^* = o(\mathbf{1})$, since $\partial\mathbf{U}_i\left(\boldsymbol{\theta}, \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}, \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right)/\partial\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}$ and $\partial\mathbf{U}_i\left(\boldsymbol{\theta}, \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}, \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right)/\partial\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}$ are linear function of $\mathbf{f}_i$'s whose means are zero, and the conditions of the theorem give

$$\begin{aligned}
\mathcal{C}^* &= n^{\frac{1}{2}}\left[\hat{\boldsymbol{\alpha}}_{\mathbf{2}}^{(\boldsymbol{con})}\left\{\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta})\right\} - \hat{\boldsymbol{\alpha}}_{\mathbf{2}}^{(\boldsymbol{con})}\left(\boldsymbol{\theta}, \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}\right) + \hat{\boldsymbol{\alpha}}_{\mathbf{2}}^{(\boldsymbol{con})}\left(\boldsymbol{\theta}, \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}\right) - \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right] \\
&= n^{\frac{1}{2}}\left[\frac{\partial\hat{\boldsymbol{\alpha}}_{\mathbf{2}}^{(\boldsymbol{con})}}{\partial\left(\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}\right)}\left(\boldsymbol{\theta}, \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}\right)\left\{\hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})}(\boldsymbol{\theta}) - \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}\right\} + \hat{\boldsymbol{\alpha}}_{\mathbf{2}}^{(\boldsymbol{con})}\left(\boldsymbol{\theta}, \boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}\right) - \boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}\right] + o(1) \\
&= O(\mathbf{1})
\end{aligned}$$

Consequently with $\mathcal{E}_i^* = O(\mathbf{1})$, the expression $\sum_{i=1}^{n} \mathbf{U}_i \left\{ \boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}_{\mathbf{1}}^{(\boldsymbol{con})} (\boldsymbol{\theta}), \boldsymbol{\alpha}^{*} (\boldsymbol{\theta}) \right\} / n^{\frac{1}{2}}$ is asymptotically equivalent to $\mathcal{A}^*$, whose asymptotic distribution is multivariate Gaussian with zero mean and covariance matrix

$$\lim_{n \to \infty} \left(\sum_{i=1}^{n} \boldsymbol{D}_i^T \, \tilde{\mathbf{V}}_i^{-1} cov(\mathbf{Y_i}) \, \tilde{\mathbf{V}}_i^{-1} \boldsymbol{D}_i / n \right) \tag{11.14}$$

Finally, it is easy to see that $\sum_{i=1}^{n} \mathcal{B}_i = o(1)$, $\sum_{i=1}^{n} \boldsymbol{\mathcal{D}}_i = o(1)$, $\mathcal{C} = O(1)$, $\mathcal{E} = O(1)$ and that $\sum_{i=1}^{n} \boldsymbol{\mathcal{A}}_i / n$ converges as $n \to \infty$ to $- \sum_{i=1}^{n} \boldsymbol{D}_i^T \tilde{\mathbf{V}}_i^{-1} \boldsymbol{D}_i / n$. This completes the proof.

12 Proof of Theorem (6.1)

Equation (6.13): Note that, since multiplications are much more computationally demanding operations than additions, we will consider here only multiplications (Deuflhard & Hohmann (1993)). Including additions would not cause any major changes to the proof given below, but would complicate the notation. We denote the numerical complexity of a matrix operation by $\mathcal{C}(.)$. $\boldsymbol{\alpha}^{(\boldsymbol{GEE})}$, $\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})}$ and $\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})}$ are the parameters describing the covariance matrix $\mathbf{V}_i$ and it holds that

$$dim \left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})} \right) \; = \; O\left(m^a\right) \quad \text{for} \quad a \; = \; 0, 1, 2$$

$$dim \left[\left\{ \left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})} \right)^T, \left(\boldsymbol{\alpha}_{\mathbf{1}}^{(\boldsymbol{con})} \right)^T, \left(\boldsymbol{\alpha}_{\mathbf{2}}^{(\boldsymbol{con})} \right)^T \right\}^T \right] \; = \; o\left(m^b\right) \quad \text{for} \quad b \; = \; 0, 1, 2$$

While inequality $a \leq b$ follows directly from the definition of a and b, the inequality $a \leq c$ is implicated by the assumption that the inverse of $\mathbf{D}_{i(22)}^T \mathbf{W}_i^{-1} \boldsymbol{D}_{i(22)}$ does exist (iteration step (5.10)) and therefore the rank of $\boldsymbol{D}_{i(22)}$ has to be $dim \left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})} \right)$. The three most numerically complex operations in iteration step (5.10) are the computation of the inverse of $\mathbf{W}_i$ and the two matrix multiplication's, $\mathbf{D}_{i(22)}^T \times \mathbf{W}_i^{-1}$ and $\left(\mathbf{D}_{i(22)}^T \mathbf{W}_i^{-1} \right) \times \boldsymbol{D}_{i(22)}$. Suppose now that we conduct iteration step (5.10) for the hybrid approach. Since $\mathbf{W}_i$ is a $dim \left\{ s \left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})} \right) \right\} \times dim \left\{ s \left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})} \right) \right\}$ matrix, the number of multiplication's involved in the calculation of its inverse increases with $O\left(\left[dim \left\{ s \left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})} \right) \right\} \right]^3 \right)$ (Press et al., 1991). Note the cubic order of numerical complexity does not depend on the numerical procedure that is used to compute the inverse of a matrix. Even, when instead of the Gauss procedure the Cholesky decomposition, which exploits the symmetry of the matrix, is used for the computation of the inverse, the complexity order of computing the inverse still remains 3rd. order (Hämmerlin & Hoffmann (1992)). So the numerical complexity of the inverse operation for $\mathbf{W}_i$ in the iteration step (5.10) can be written as $\mathcal{C}\left(\mathbf{W}_i^{-\mathbf{1}} \right) = O\left(m^{3c} \right)$. It is easy to see that for the two matrix products the numerical complexity in iteration step (5.10) is $\mathcal{C}\left(\mathbf{D}_{i(22)}^T \times \mathbf{W}_i^{-1} \right) = O\left(m^{a+2c} \right)$ and $\mathcal{C}\left\{ \left(\mathbf{D}_{i(22)}^T \mathbf{W}_i^{-1} \right) \times \boldsymbol{D}_{i(22)} \right\} = O\left(m^{2a+c} \right)$, so that the total numerical complexity of iteration step (5.10) is given by

$$\mathcal{C}(\text{"iteration step (5.10)"}) \; = \; O\left\{ n \; max\left(m^{3c}, m^{a+2c}, m^{2a+c} \right) \right\} \; = \; O\left(n \; m^{3c} \right)$$

Since, assuming that $dim(\boldsymbol{\beta})$ does not depend on n, the numerical complexity of iteration step (5.9) increases with $O(n\ m^3)$ and the rate of growth for the number of the multiplications necessary to calculate all possible second moments $y_{ij}y_{ij'}$ for the consistent moment-based estimators is only $O(n\ m^2)$, the total numerical complexity of the hybrid-approach can be expressed by

$$\mathcal{C}_{hybrid} \ = \ O\left\{n\ m^{max(3c,3)}\right\}$$

The calculation of the total complexity of the standard GEE-approach for multivariate data is done analogously. Since all parameters of the variance matrix $\mathbf{V}_i$ are now estimated by the estimating equations, the "working" variance matrix $\mathbf{W}_i$ of all empirical covariances is a $(m(m+1)/2) \times (m(m+1)/2)$ matrix. So the numerical complexities for inverting matrix $\mathbf{W}_i$ and two previously considered matrix-multiplication's in iteration step (5.10) are $\mathcal{C}\left(\mathbf{W}_i^{-\mathbf{1}}\right) \ = \ O\left(m^6\right)$, $\mathcal{C}\left(\mathbf{D}_{i(22)}^T \ \times \ \mathbf{W}_i^{-1}\right) \ = \ O\left(m^{b+4}\right)$ and $\mathcal{C}\left\{\left(\mathbf{D}_{i(22)}^T\mathbf{W}_i^{-1}\right) \times \boldsymbol{D}_{i(22)}\right\} \ = \ O\left(m^{2b+2}\right)$, giving a total numerical complexity for iteration step (5.10) of

$$\mathcal{C}\left(\text{"iteration step (5.10)"}\right) \ = \ O\left\{n\ max\left(m^6, m^{b+4}, m^{2a+2}\right)\right\} \ = \ O\left(n\ m^6\right)$$

and the total numerical complexity of the multivariate GEE-approach can be expressed by

$$\mathcal{C}_{GEE} \ = \ O\left\{n\ m^6\right\}$$

Therefore the relative numerical complexity is

$$\mathcal{C}_{GEE}/\mathcal{C}_{hybrid} \quad = \quad O\left\{m^{min(6-3c,3)}\right\}$$

Equation (6.14): Since $dim\left(\boldsymbol{\alpha}^{(\boldsymbol{GEE})}\right)$ is $O(m)$, the numerical complexity of iteration step (5.10) and also of the total estimation procedure is determined by the inversion of the matrix $\mathbf{W}_i$. So the total complexities of the standard multivariate GEE-approach and the hybrid approach can be expressed by

$$\begin{aligned} C_{GEE} \ &= \ \omega \ \left(m^2/2\right)^3 \ + \ O\left(m^5\right) \\ C_{hybrid} \ &= \ \omega \ \left(\kappa\ m^2\right)^3 \ + \ O\left(m^5\right) \end{aligned}$$

with $\omega \in \mathbb{R}_{>0}$ Thus equation (6.14) follows directly.

Table 1. *Average relative efficiency between the hybird approach and the original GEE-approach by Liang & Zeger (1986)*

Sample Size	*Correlation r* 0.0	0.3	0.6	0.9
50	*1.01*	*1.03*	*1.11*	*1.30*
75	*1.00*	*1.02*	*1.08*	*1.27*
100	*1.01*	*1.01*	*1.07*	*1.19*
200	*1.00*	*1.03*	*1.05*	*1.23*
500	*1.02*	*1.01*	*1.05*	*1.11*
1000	*1.00*	*1.01*	*1.04*	*1.12*
inf	*1.00*	*1.01*	*1.04*	*1.10*

Table 2. *Multivariate "ad-hoc" GEE-analysis of osteoporisis realted traits*

parameter	*estimate*	*std. error*	*p-value*	ϕ_j	α
β_1	*0.042*	*0.019*	*0.029*	*0.19*	$\alpha_w = 0.88$
β_2	*0.037*	*0.016*	*0.019*	*0.18*	$\alpha_b = 0.27$
β_3	*0.036*	*0.015*	*0.013*	*0.17*	
β_4	*0.032*	*0.014*	*0.023*	*0.17*	

References

[1] T. Andrew, P. Reed Y.T. Mak, A.J. McGregor, and T.D. Spector. Association of heel ultrasound measurements, bone mineral density and bone turnover markers with the osteocalcin gene in female dizygotic twins. *submitted*, 2002.

[2] D. Balding, M. Bishop, and C. Cannings. *Handbook of Statistical Genetics*. Wiley Edition, New York, 2001.

[3] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer Verlag, 1991.

[4] N.R. Chaganty and J. Shults. On the eliminating the asymptotic bias in the quasi-least squares estimate of the correlation parameter. *Journal of Statistical Planning and Inference*, 76:145–161, 1999.

[5] D. Clayton. *Population association.*, chapter 19, pages 519–540. In Balding et al. [2], 2001.

[6] P. Deuflhard and A. Hohmann. *Numerische Mathematik I, Eine algorithmisch orientierte Einfuehrung*. Walter de Gruyter, 1993.

[7] R.C. Jansen. High resolution of quantitative traits into multiple loci via interval mapping. *Genetics*, 136:1447–1455, 1994.

[8] SA Knott and CS Haley. Multitrait least squares for quantitative trait loci detection. *Genetics*, 158:899–911, 2000.

[9] C. Lange, E. Silverman, X. Xu, S. Weiss, and N.M. Laird. A multivariate transmission disequilibrium test. *Biostatistics*, in press, 2002.

[10] C. Lange and J.C. Whittaker. A generalized estimating equation approach to mapping of quantitative trait loci (qtl). *Genetics*, 159:1325–1337, 2001.

[11] Y. Lee and J.A. Nelder. Hierarchical generalized linear models. *Journal of the Royal Statistical Society*, 58(4):619–678, 1996.

[12] K.-Y. Liang and S.L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

[13] K.-Y. Liang, S.L. Zeger, and B. Qaqish. Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society, B*, 54(1):3–40, 1992.

[14] S.R. Lipsitz, N.M. Laird, and D.P. Harrington. Generalized estimating equations for correlated binary data - using the odds ration as a measure of association. *Biometrika*, 78(1):153–160, 1991.

[15] P. McCullagh. Quasi-likelihood functions. *The Annals of Statistics*, 11:59–67, 1983.

[16] P. McCullagh. *Discussion*, pages 3–40. Volume 54 of Liang [13], 1992.

[17] R. Prentice and L. Zhao. Estimating equations for parameters in means and covariance of multivariate discrete and continuous response. *Biometrics*, 47:825–839, 1991.

[18] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes In C*. Cambridge University Press, 1991.

[19] A. Rotnitzky and N.P. Jewell. Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika*, 77(3):485–497, 1990.

[20] T.G. Schulze and F.J. McMahon. Genetic association mapping at the crossroads: Which test and why? overview and practical guidelines. *American Journal of Medical Genetics*, 114:1–11, 2002.

[21] J. Shults and N.R. Chaganty. Analysis of serially correlated data using quasi-least squares. *Biometrics*, 54:1622–1630, 1999.

[22] R.W.M. Wedderburn. Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika*, 61(3):439, 1974.

[23] S.L. Zeger. A regression model for time series of counts. *Biometrika*, 75(4):621–629, 1988.

[24] Z.-B. Zeng. Precision mapping of quantitative trait loci. *Genetics*, 136:1457–1468, 1994.

[25] H. Zhao. Family-based association studies. *Stat Methods in Med Res*, 9:563–587, 2000.

[26] L. Zhao, R. Prentice, and S. Self. Multivariate mean parameter estimation by using a partly exponential model. *Journal of the Royal Statistical Society, Series B*, 54:805–811, 1992.