Biometrical Journal **46** (2004) 5, 540–553 DOI: 10.1002/bimj.200110052

Modelling Heterogeneous Dispersion in Marginal Models for Longitudinal Proportional Data

Peter X.-K. Song[*,1], **Zhenguo Qiu**[2], and **Ming Tan**[3]

[1] Department of Mathematics and Statistics, York University, Toronto, ON, Canada M3J 1P3
[2] B.C. Research Institute for Children's and Women's Health, Vancouver, BC, Canada V5Z 4H4
[3] Greenebaum Cancer Center N9E17, University of Maryland, Baltimore, MD 21201, USA

Received 24 July 2001, revised 22 October 2003, accepted 15 June 2004

Summary

Continuous proportional data is common in biomedical research, e.g., the pre-post therapy percent change in certain physiological and molecular variables such as glomerular filtration rate, certain gene expression level, or telomere length. As shown in (Song and Tan, 2000) such data requires methods beyond the common generalised linear models. However, the original marginal simplex model of (Song and Tan, 2000) for such longitudinal continuous proportional data assumes a constant dispersion parameter. This assumption of dispersion homogeneity is imposed mainly for mathematical convenience and may be violated in some situations. For example, the dispersion may vary in terms of drug treatment cohorts or follow-up times. This paper extends their original model so that the heterogeneity of the dispersion parameter can be assessed and accounted for in order to conduct a proper statistical inference for the model parameters. A simulation study is given to demonstrate that statistical inference can be seriously affected by mistakenly assuming a varying dispersion parameter to be constant in the application of the available GEEs method. In addition, residual analysis is developed for checking various assumptions made in the modelling process, e.g., assumptions on error distribution. The methods are illustrated with the same eye surgery data in (Song and Tan, 2000) for ease of comparison.

Key words: Continuous proportions; Generalised linear models; GEEs; Longitudinal data; Residual analysis; Simplex distribution; Varying dispersion.

1 Introduction

The concept of dispersion parameter is a familiar one in generalised linear models (GLMs). The dispersion parameter of a normal distribution is simply its variance; and the dispersion parameter of Poisson distributions is always equal to 1, which is the ratio of variance to mean and where the over-dispersion occurs when such a ratio is larger than 1. Dispersion models (Jørgensen, 1997), as an extension of the GLMs, include dispersion parameters describing the distributional shape, which is beyond what the location or mean parameter alone can describe.

The simplex distribution of Barndorff-Nielsen and Jørgensen (1991) for the error term represents a special dispersion model, and is useful for modelling continuous proportional data. Based on this distribution, Song and Tan (2000) developed a marginal model for longitudinal continuous proportional data, which was used to analyse an eye surgery data. Similar to Liang and Zeger's marginal models e.g. Diggle et al. (2002), Song and Tan (2000) assumed a constant dispersion in their model and their focus is on modelling the trend component. A technical advantage by setting a constant dispersion parameter is that, as shown in Liang and Zeger's GEE1 (1986) approach, regression coefficients can be separately estimated from the dispersion parameter. This is because the GEE1 can factorise a constant dispersion out the estimating equation.

* Corresponding author: e-mail: song@mathstat.yorku.ca

However, in practice the assumption of homogeneous dispersion may be questionable. For example, the magnitude of dispersion may vary across drug treatment cohorts due to different rates of disease progression or over different follow-up times due to different environmental exposures. It is clear that the marginal pattern of a population depends not only on its averaged trend but also on its dispersion characteristics, as described by the dispersion models. Therefore, incorporating varying dispersion in the modelling process allows us to assess the heterogeneity of dispersion and to develop a simultaneous inference for the entire marginal models concerning both trend and dispersion components. Such an access to the profile of the dispersion parameter is important, as shown in our simulation studies in Section 4, mistakenly assuming a varying dispersion to be constant in the application of GEE1 method could cause some serious problems in statistical inference. For example, the asymptotic normality theory for the estimators may no longer be valid, and this theory is crucial to test for statistical significance for the effects of some covariates of interest. In addition, a proper estimation for the dispersion parameter is appealing, for example, in residual analysis, where a standardisation for residuals is usually taken to stablise their variances. The computation of standardised residuals always asks for an appropriate estimate of the dispersion parameter.

In this paper, we propose a new marginal model that consists of three components to be modeled: the population-averaged effects, the dispersion pattern, and the correlation. In the context of longitudinal data analysis, the first version of generalised estimating equation approach, known as of GEE1, is proposed by Liang and Zeger (1986) and later extended by Prentice and Zhao (1991) to include a set of estimating equations on correlation parameters, referred to GEE2 in the literature. As a matter of fact, estimating a varying dispersion parameter can be easily incorporated with the GEE2 using the mean-variance relationship of the classical GLMs or the exponential dispersion family distributions. However, the mean-variance relationship is no longer valid for the simplex distribution, because it is not an exponential dispersion family distribution. Therefore, in this paper, we suggest to add another set of estimating equations to deal with the dispersion component through a certain moment property different from the mean-variance relationship. The resulting estimating equations extend the currently popular GEE2, although it is still called as GEE2 in the present paper.

Modelling dispersion parameter has been considered by many authors for different models in the literature. Among others, Smyth (1989) discussed generalised linear models with varying dispersion for cross-sectional data, and Artes and Jørgensen (2000) proposed a model for the index parameter of dispersion models, attempting to attack this problem with an underlying application closely related to von Mises distribution for longitudinal circular data. We found their method did not work well for the simplex distribution. Paik (1992) proposed an estimation procedure that extends Liang and Zeger's GEEs by allowing observations from distributions with different dispersion parameters. However, Paik's procedure is not applicable for the simplex distribution, because there is no closed form expression for the variance of the distribution. By utilising a certain moment property of the simplex distribution, we come up with a different solution from those given by Artes and Jørgensen (2000) and Paik (1992). In fact, because of different perspectives and models, our estimating equation for the dispersion parameter of the simplex distribution is simpler and numerically more efficient than theirs.

The rest of the paper is organised as follows. Section 2 presents dispersion marginal models with varying dispersion. An extended GEE2 is presented in Section 3, and Section 4 gives a simulation study that demonstrates the importance of modelling the dispersion parameter to conduct a proper statistical analysis in the presence of heterogeneous dispersion. Section 5 discusses model diagnostics through residual analysis. The proposed methods are applied to re-analyse the eye surgery data in Section 6. Finally we conclude with some remarks.

2 Marginal Models

To develop marginal simplex models for longitudinal continuous proportional data with varying dispersion, first let y_{ij}, $j = 1, \ldots, n_i$ be the sequence of observed repeated measurements on the ith of m subjects, and t_{ij}, $j = 1, \ldots, n_i$, be the sequence of corresponding times on which the measurements are taken on each subject. Associated with each y_{ij} are the values, x_{ijk}, $k = 1, \ldots, p$, of p covariates or expla-

natory variables. We assume that y_{ij} are realisations of random variables Y_{ij} which follow simplex distributions $Y_{ij} \sim S^{-}(\mu_{ij}, \sigma_{ij}^2)$, where $\mu_{ij} \in (0,1)$ are the mean parameters and $\sigma_{ij}^2 > 0$ are the dispersion parameters, and both may be specified as functions of covariates. The density function of the simplex distribution is, suppressing indices, given by

$$p(y;\mu,\sigma^2) = \left[2\pi\sigma^2\{y(1-y)\}^3\right]^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} d(y;\mu)\right\}, \quad y \in (0,1)\,,$$

where d is the unit deviance,

$$d(y;\mu) = \frac{(y-\mu)^2}{y(1-y)\,\mu^2(1-\mu)^2},$$

and its unit variance function is $v(\mu) = \mu^3(1-\mu)^3$. See (Jørgensen, 1997) for more details.

Let

$$\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^\top, \; \boldsymbol{x}_{ij} = (1, x_{ij1}, \ldots, x_{ijp})^\top.$$

We assume that $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ are independent.

A marginal simplex model consists of three components given as follows. The first component is a model to describe the population-averaged effects, where the mean parameter μ_{ij} depends on the time-varying covariates $\boldsymbol{x}_{ij}$ via a generalised linear model of the form

$$h(\mu_{ij}) = \boldsymbol{x}_{ij}^\top \boldsymbol{\beta} \tag{1}$$

where h is a known link function and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^\top$ is the regression coefficients to be estimated. The link function is usually chosen to be the logit link function that maps the unitary interval to $(-\infty, \infty)$.

The second component is a model to describe the pattern of dispersion parameter σ_{ij}^2 as a function of covariates z_{ij} (maybe a subset of $\boldsymbol{x}_{ij}$), given by

$$g(\sigma_{ij}^2) = \boldsymbol{z}_{ij}^\top \boldsymbol{\gamma} \tag{2}$$

where g is a known link function and $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_r)^\top$ with γ_0 corresponding to the intercept term. To express the dispersion as of a multiplicative form, the logarithm link function is used to obtain a log-linear model and hence

$$\sigma_{ij}^2 = \exp\left(\boldsymbol{z}_{ij}^\top \boldsymbol{\gamma}\right) = \prod_{k=0}^{r} \left(\mathrm{e}^{\gamma_k}\right)^{z_{ijk}} = \mathrm{e}^{\gamma_0} \prod_{k=1}^{r} \left(\mathrm{e}^{\gamma_k}\right)^{z_{ijk}}.$$

The third component is for modelling correlation structure. The correlation between Y_{ij} and Y_{ik} is a function of the location parameters and perhaps of additional parameters, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)^\top$, namely,

$$\mathrm{corr}\,(Y_{ij}, Y_{ik}) = \rho(\mu_{ij}, \mu_{ik}, \boldsymbol{\alpha}) \tag{3}$$

where $\rho(\cdot)$ is a known function. Various types of correlation structures may be used for the ρ function. Amongst others, three commonly used in the analysis of longitudinal data are the exchangeable, AR(1) and m-dependence correlations. It is noted that the justification for a choice of a correlation structure is in general a difficult task due to little information over time available. However the Liang and Zeger's GEE1 approach for consistent parameter estimation enjoys the robustness against misspecification of correlation structure and hence has yielded popularity in longitudinal data analysis.

3 GEEs for Parameter Estimation

Denote the mean vector of subject i by $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{in_i})^\top$. Let the score vector for subject i be

$$\boldsymbol{u}_i = (u_{i1}, \ldots, u_{in_i})^\top, \quad \text{with} \quad u_{ij} = -\tfrac{1}{2} d'(y_{ij};\mu_{ij}),$$

and under the regularity conditions, $E(u_{ij}) = 0$ and therefore $E(\boldsymbol{u}_i) = \mathbf{0}$. From Song and Tan (2000), the variance of u_{ij} is given by

$$\text{var}\,(u_{ij}) = \frac{\sigma_{ij}^2}{2} E\{d''(Y_{ij};\mu_{ij})\} = \frac{3\sigma_{ij}^4}{\mu_{ij}(1-\mu_{ij})} + \frac{\sigma_{ij}^2}{v(\mu_{ij})}.$$

Following Song and Tan (2000), let $\boldsymbol{w}_i = \text{diag}\,\{v(\mu_{ij})\}\,\boldsymbol{u}_i$ be the working vector, and let $\boldsymbol{R}(\boldsymbol{\alpha})$ be an $n_i \times n_i$ working correlation matrix with a $q \times 1$ vector of correlation parameters $\boldsymbol{\alpha}$. So working covariance matrix for $\boldsymbol{w}_i$ is

$$\boldsymbol{V}_i = \text{diag}^{1/2}\big\{\text{var}\,(w_{ij})\big\}\,\boldsymbol{R}(\boldsymbol{\alpha})\text{diag}^{1/2}\big\{\text{var}\,(w_{ij})\big\}.$$

Therefore Liang and Zeger's GEE1 for the simplex margin corresponds to the estimating equation for $\boldsymbol{\beta}$ given by

$$\boldsymbol{\psi}_1(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\alpha}) = \sum_{i=1}^{m} \boldsymbol{D}_i^{\top}\boldsymbol{A}_i\boldsymbol{V}_i^{-1}\boldsymbol{w}_i = \mathbf{0}, \tag{4}$$

where $\boldsymbol{A}_i = \text{diag}\,\{\sigma_{ij}^{-2}v(\mu_{ij})\,\text{var}\,(u_{ij})\}$ and $\boldsymbol{D}_i^{\top} = \partial\boldsymbol{\mu}_i^{\top}/\partial\boldsymbol{\beta}$.

Following Prentice and Zhao (1991), the GEE2 is formed by adding an additional set of estimating equations for the correlation parameters based on the standardised score residuals, defined by

$$r_{ij} = \frac{u_{ij}}{\sqrt{\text{var}\,(u_{ij})}} = \frac{u_{ij}}{\sigma_{ij}\sqrt{\frac{1}{2}Ed''(y_{ij};\mu_{ij})}}.$$

It is easy to see that such score residuals satisfy moment properties of $E(r_{ij}) = 0$, $\text{var}\,(r_{ij}) = 1$ and

$$E(r_{ij}r_{ij'}) = \text{corr}\,(u_{ij},u_{ij'}) = \text{corr}\,(w_{ij},w_{ij'}).$$

The estimating equation for the correlation parameter $\boldsymbol{\alpha}$ then takes the form

$$\boldsymbol{\psi}_3(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\alpha}) = \sum_{i=1}^{m}\left(\frac{\partial\boldsymbol{\eta}_i^{\top}}{\partial\boldsymbol{\alpha}}\right)\boldsymbol{H}_i^{-1}(\boldsymbol{r}_i - \boldsymbol{\eta}_i) = \mathbf{0}, \tag{5}$$

where $\boldsymbol{r}_i = (r_{i1}r_{i2}, r_{i1}r_{i3}, \ldots, r_{in_i-1}r_{in_i})^{\top}$, $\boldsymbol{H}_i$ is a working covariance matrix and $\boldsymbol{\eta}_i = E(\boldsymbol{r}_i)$.

The extended GEE2 consists of the equations (4), (5), and an estimating equation for the dispersion component given as follows,

$$\boldsymbol{\psi}_2(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\alpha}) = \sum_{i=1}^{m}\left(\frac{\partial\boldsymbol{\sigma}_i^{\top}}{\partial\boldsymbol{\gamma}}\right)\Sigma_i^{-1}(\boldsymbol{d}_i - \boldsymbol{\sigma}_i) = \mathbf{0}\,, \tag{6}$$

where $\boldsymbol{d}_i = (d(y_{i1};\mu_{i1}),\ldots,d(y_{in_i};\mu_{in_i}))^{\top}$, Σ_i is a working covariance matrix, and $\boldsymbol{\sigma}_i = E(\boldsymbol{d}_i) = (\sigma_{i1}^2,\ldots,\sigma_{in_i}^2)^{\top}$. Note that here we use the squared deviance residuals, rather than the squared Pearson residuals given in Paik (1992), to form the third sets of estimating equations. The Crowder optimal matrix for Σ_i (Crowder, 1987) is in fact the $\text{cov}\,(\boldsymbol{d}_i)$ which is in general not easy to obtain. A simple choice of Σ_i is the identity matrix, leading to the method of moments estimator for $\boldsymbol{\gamma}$. Perhaps a better choice for Σ_i is a diagonal matrix with diagonal elements equal to the variances $\text{var}\,\{d(Y_{ij};\mu_{ij})\} = 2(\sigma_{ij}^2)^2$. See the appendix for the proof of this formula in detail. This indicates a gamma type of mean-variance relation, that is, the unit variance function is equal to the squared mean. With this choice, the estimating equation will effectively produce the quasi-likelihood estimator of $\boldsymbol{\gamma}$ as does the gamma regression (Wedderburn, 1974).

Let $\boldsymbol{\theta} = (\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\alpha})$ be the vector of parameters to be estimated via the extended GEE2 for which the estimates are obtained by simultaneously solving the joint equations,

$$\Upsilon(\boldsymbol{\theta}) = \Upsilon(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\alpha}) = \begin{bmatrix}\boldsymbol{\psi}_1(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\alpha})\\ \boldsymbol{\psi}_2(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\alpha})\\ \boldsymbol{\psi}_3(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\alpha})\end{bmatrix} = \mathbf{0}\,. \tag{7}$$

It is clear that the estimating equation $\Upsilon(\boldsymbol{\theta}) = \mathbf{0}$ is unbiased, namely $E\Upsilon(\boldsymbol{\theta}) = \mathbf{0}$. Hence it follows from the standard theory of estimating equations that under some mild regularity conditions, the estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\alpha}})$ is consistent and $m^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically multivariate Gaussian with zero mean and covariance matrix of the form $\lim_m m\boldsymbol{J}^{-1}(\boldsymbol{\theta})$, where $\boldsymbol{J}(\boldsymbol{\theta})$ is the Godambe information matrix given by $\boldsymbol{J}(\boldsymbol{\theta}) = \boldsymbol{S}^\top \boldsymbol{R}^{-1} \boldsymbol{S}$. Details of the sensitivity matrix $\boldsymbol{S} = E\{\partial\Upsilon(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^\top\}$ and of the variability matrix $\boldsymbol{R} = E\{\Upsilon(\boldsymbol{\theta})\Upsilon^\top(\boldsymbol{\theta})\}$ are given in the appendix.

Using the Newton-scoring algorithm, the solution $\hat{\boldsymbol{\theta}}$ for the joint equation (7) can be obtained numerically by iteratively updating the $\boldsymbol{\theta}$ values as follows,

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - \boldsymbol{S}^{-1}\Upsilon\left(\boldsymbol{\theta}^{(k)}\right).$$

4 Simulation Study

To demonstrate the importance of properly analysing the longitudinal data in the presence of heterogeneous dispersion, we conduct a simulation study where the proportional data $y_i \sim S^-(\mu_i, \sigma_i^2)$, $i = 1, \ldots, 150$, were generated independently according to the following marginal models:

$$\text{logit}\,(\mu_i) = \beta_0 + \beta_1 T_i + \beta_2 S_i\,,$$

$$\log\,(\sigma_i^2) = \gamma_0 + \gamma_1 T_i\,,$$

where covariates T and S are variables of treatment groups indicated by (−1, 0, 1), and illness severity score ranged in (0, 1, 2, 3, 4, 5, 6) that is randomly assumed to each subject by a binomial distribution $B(6, 0.5)$. For simplicity, we mainly investigated how the parameters β_j's representing the population-averaged effects would be affected by the situation of the dispersion parameter. So, we considered only the independence correlation structure, for which we were able to simulate data. We took three equally sized treatment groups, each with 50 subjects. Using the notation above, we yield $\boldsymbol{x}_i = (1, T_i, S_i - 3)^\top$ and $\boldsymbol{z}_i = (1, T_i)^\top$ in which the severity covariate was centralised by the mid-score 3. Moreover, the true values were assigned as $(\beta_0, \beta_1, \beta_2) = (0.5, -0.5, 0.5)$, $(\gamma_0, \gamma_1) = (3, 2)$. We ran the regression over 200 replications, and the corresponding summaries are listed below.

Table 1 reports the summary statistics of the parameter estimates from the extended GEE2 approach proposed in the paper with heterogeneous dispersion. These statistics include mean point estimate, 2.5th and 97.5th percentiles, empirical standard deviation and mean standard error for each of the five parameters.

When the model with the homogeneous dispersion was used to fit the simulated data, the mean estimate of $\log\,(\sigma^2)$ was 19.2 with the empirical standard deviation equal to 80.6, considerably larger than the average standard error 0.12 obtained from the sandwich asymptotic variance estimator.

Table 1 Summary statistics of the estimates, based on 200 replications generated from the heterogeneity model.

True Value	*Heterogeneous Dispersion*				*Homogeneous Dispersion*			
	Mean	(2.5%, 97.5%)	Stdev*	Stderr†	Mean	(2.5%, 97.5%)	Stdev	Stderr
β_0 (0.5)	0.4986	(0.3671, 0.6618)	0.0810	0.0792	0.4353	(−1.2075, 1.5808)	0.5709	0.1076
β_1 (−0.5)	−0.5046	(−0.6697, −0.3493)	0.0861	0.0896	−0.3120	(−1.9471, 1.9598)	0.8371	0.1316
β_2 (0.5)	0.5013	(0.3973, 0.6159)	0.0575	0.0539	0.3668	(−0.7345, 0.7901)	0.4045	0.0906
γ_0 (3.0)	2.9782	(2.7778, 3.2258)	0.1123	0.1155	–	–	–	–
γ_1 (2.0)	2.0088	(1.7008, 2.2840)	0.1491	0.1414	–	–	–	–

* Empirical standard deviation; † Mean standard error

From Table 1, we learned: (i) The point estimates $\hat{\beta}_i, i = 0, 1, 2$ in the pupulation-averaged effects model (1) from both approaches are relatively close to each other, although the model with the homogeneous dispersion produces a little larger deviation from the true values than the model with the heterogeneous dispersion. (ii) The 95% empirical confidence intervals from the two models have substantially different coverage, zero being included in the intervals given by the homogeneity model as opposed to zero being excluded in those given by the heterogeneity model, for all three β parameters. This suggests that the homogeneity model loses its power of identifying some important covariates in the presence of heterogeneous dispersion. (iii) The values of the empirical standard deviation and the standard error are very similar in the heterogeneity model, but clearly different in the homogeneity model. This indicates that the asymptotic normality theory for the estimators from the homogeneity model may be no longer valid. To visualise this, we plotted the estimated densities over the 200 estimates for each parameter in Figure 1.

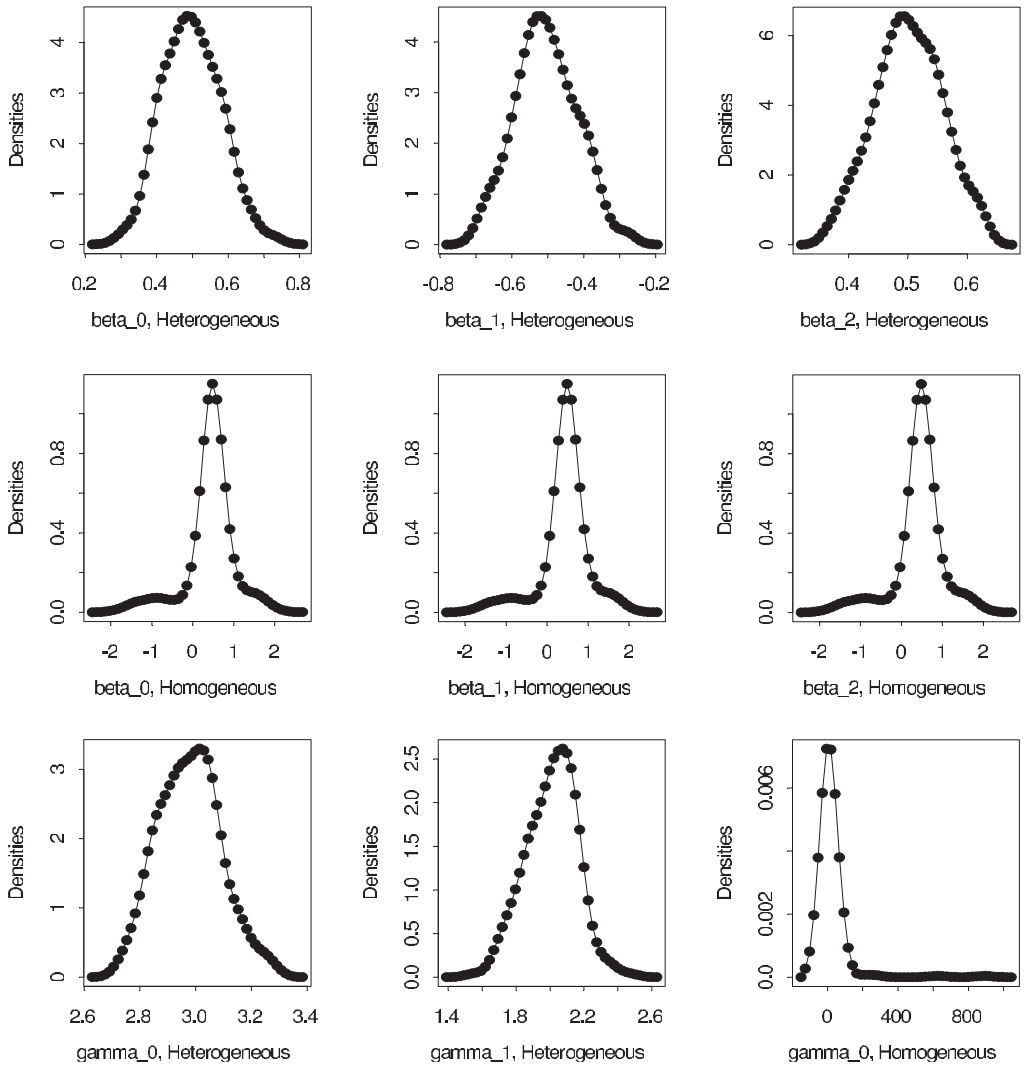

Figure 1 Estimated densities of the model parameters over 200 replications using data generated from the heterogeneity model.

Figure 1 indicates that for each parameter, the estimates from the heterogeneity model are evenly distributed along the parameter space and clearly form a bell-shaped density. In contrast, the estimates from the homogeneity model occurs more frequently on tail areas and clearly form a heavy-tailed density. The density for the estimator of γ from the homogeneity model has an extremely long tail on the right. In conclusion, the asymptotic normality for the estimators from the homogeneity model is seriously in question.

Conversely, we conducted another simulation in that the true model had the homogeneous dispersion. In particular, data were generated similarly as in the first simulation, except that now the dispersion model is constant $\log(\sigma_i^2) = \gamma_0$. The true values of β parameters are the same as above, and set $\gamma_0 = 4$, which leads to a large dispersion around 55. Table 2 gives the summary statistics over 200 replications.

Table 2 Summary statistics of the estimates, based on 200 replications generated from the homogeneity model.

True Value	*Heterogeneous Dispersion*				*Homogeneous Dispersion*			
	Mean	(2.5%, 97.5%)	Stdev*	Stderr†	Mean	(2.5%, 97.5%)	Stdev	Stderr
β_0 (0.5)	0.5017	(0.2938, 0.7088)	0.1035	0.0971	0.5014	(0.2942, 0.7105)	0.1035	0.0972
β_1 (−0.5)	−0.5078	(−0.6943, −0.2885)	0.1161	0.1171	−0.5089	(−0.6992, −0.2940)	0.1166	0.1173
β_2 (0.5)	0.5109	(0.3623, 0.6880)	0.0894	0.0842	0.5101	(0.3690, 0.6857)	0.0895	0.0843
γ_0 (4.0)	3.9624	(3.7216, 4.2071)	0.1151	0.1155	3.9695	(3.7263, 4.2149)	0.1143	0.1155
γ_1 (0.0)	−0.0099	(−0.2874, 0.2889)	0.1489	0.1414	–	–	–	–

* Empirical standard deviation; † Mean standard error

Evidentially, Table 2 indicates that the estimates from the two models are very close, the null hypothesis $H_0 : \gamma_1 = 0$ cannot be rejected at the significance level 0.05 under the heterogeneity model. As expected, the estimated densities (not shown in the paper) of the parameters are very similar between the two models, and they are all very alike to normal density curves.

In summary, when a constant dispersion assumption is in doubt, the heterogeneity model seems to be necessary and advantageous to make proper statistical inference.

5 Residual Analysis

We propose to use two types of residuals to form diagnostics for the key model assumptions: (i) marginal distributions, (ii) link functions, and (iii) the working correlation structure. The first one is the standardised score residuals r_{ij} given in (5), and the other is the regular standardised Pearson residuals, $e_{ij} = (y_{ij} - \mu_{ij})/\sqrt{\text{var}(Y_{ij})}$, where $\text{var}(Y_{ij})$ has no closed form expression as it involves the incomplete gamma function. See Jørgensen (1997) for the details.

The sample counterpart of r_{ij} or e_{ij} is obtained by replacing parameters by their corresponding estimates, denoted by $\hat{r}_{ij}$ or $\hat{e}_{ij}$, accordingly. Like most residual analyses, our residual analysis below is useful to detect strong signals associated with certain model assumption violation.

The simplex distribution assumption can be checked by the plot of $\hat{e}_{ij}$ against $\hat{\mu}_{ij}$, which aims to examine the mean-variance relation. If this assumption is true, then $\text{var}(e_{ij}) = 1$, independent of mean μ_{ij}. Therefore, points in the plot should randomly scatter around the horizontal line at zero (the expectation of residuals), with approximately 95% points in the horizontal band between -2 and 2. Any apparent departure from this would suggest either a violation of the assumption on distribution or probably a poor model fit. A series of further investigations are needed to identify which factor is responsible for such departure. This approach would become more reliable as σ^2 becomes large, because the mean-variance relation becomes dominated by $\mu(1-\mu)$, a case similar to that of a binomial distribution.

Following McCullagh and Nelder's (1989), we use the plot of the adjusted dependent variable s_{ij} against the linear predictor $\hat{\eta}_{ij}$ to check the chosen link function. In our setting, define

$$s_{ij} = h(\mu_{ij}) + \left\{ \frac{3\sigma_{ij}^4}{\mu_{ij}(1-\mu_{ij})} + \frac{\sigma_{ij}^2}{v(\mu_{ij})} \right\}^{-1/2} u(y_{ij};\mu_{ij}), \quad j = 1,\ldots,n_i; \quad i = 1,\ldots,m\,.$$

Clearly, $E(s_{ij}) = h(\mu_{ij})$ since $E(u_{ij}) = 0$, and $\mathrm{var}\,(s_{ij}) = E\{s_{ij} - h(\mu_{ij})\}^2 = 1$. If the link function is appropriate, the plot of the estimates $\hat{s}_{ij}$ against $\hat{\eta}_{ij} = \boldsymbol{x}_{ij}^\top \hat{\boldsymbol{\beta}}$ should show a straight line with approximately 95% points falling into a band with the upper and lower limits of $\hat{\eta}_{ij} \pm 2$. As in generalised linear models, this plot does not suggest the best link function for the model, but rather only gives an informal check for any strong violation of the used link.

Although it is difficult to model the true correlation structure of longitudinal data, approximately correct correlation structures allow regression coefficients to be estimated more efficiently. Thus, it is important to assess the appropriateness of the working correlation used in GEEs via residual analysis. Note that

$$\mathrm{corr}\,(r_{ij}, r_{ij'}) = \mathrm{corr}\,(w_{ij}, w_{ij'}),$$

implying that the true correlation of variable w_{ij} is equal to that of the standardised score residuals r_{ij}. Some exploratory procedures presented in Section 3.4 of Diggle et al. (2002) can be adopted for w_{ij}'s to examine the correlation of data.

6 An Example

In this section we re-analyse the ophthalmological data on the use of intraocular gas in retinal repair surgeries (Meyers et al., 1992), with a special focus on the heterogeneous dispersion. A primary analysis of the data assuming the homogeneous dispersion was done by Song and Tan (2000). Briefly, the study was to investigate the decay course of the intraocular gas in retinal repair surgeries prospectively in 31 patients. The gas was injected into the eye before surgery and patients were followed three to eight (average of 5) times over a three-month period. The response variable y_{ij} was the percent of gas left in the eye recorded as proportion (a percent). The question was if the disappearance of the gas is related to other covariates such as the concentration of the gas used. Song and Tan (2000) modelled the gas volume directly using a marginal model. With our proposed method, we are able to test if the homogeneous dispersion is true, and if not so the model allows us to identify which covariates lead to heterogeneity.

To begin, the population-averaged effects model in Song and Tan (2000) is

$$\mathrm{logit}\,(\mu_{ij}) = \beta_0 + \beta_1 \log\,(t_{ij}) + \beta_2 \log^2(t_{ij}) + \beta_3 x_{ij} \tag{8}$$

where t_{ij} is the time covariate of days after the gas injection, and x_{ij} is the covariate of gas concentration levels equal to -1, 0 and 1, corresponding to the concentration levels of 15%, 20% and 25%, respectively. To this model, the components of the estimating function $\boldsymbol{\psi}_1$ specified by (4) are given as follows.

$$\boldsymbol{D}_i^\top = X_i^\top \mathrm{diag}\,\{\mu_{ij}(1-\mu_{ij})\}\,, \qquad \boldsymbol{D}_i^\top \boldsymbol{A}_i = X_i^\top \mathrm{diag}\,\{3\sigma_{ij}^2 v(\mu_{ij}) + \mu_{ij}(1-\mu_{ij})\},$$

where X_i is of $n_i \times 3$ dimension and its jth row is $(1, \log\,(t_{ij}), \log^2(t_{ij}), x_{ij})$, and

$$\mathrm{var}\,(w_{ij}) = \sigma_{ij}^2 v(\mu_{ij})\{3\sigma_{ij}^2\mu_{ij}^2(1-\mu_{ij})^2 + 1\}.$$

Clearly the corresponding sensitivity matrix is $\boldsymbol{S}_{11} = -\sum_{i=1}^{m} \boldsymbol{D}_i^\top \boldsymbol{A}_i \boldsymbol{V}_i^{-1} \boldsymbol{A}_i \boldsymbol{D}_i$.

Also as indicated in their paper, AR(1) dependence seemed to fit the data the best, so our analysis only concerns this type of dependence, specified as of the first-order ECM model, $\mathrm{corr}\,(w_{ij}, w_{ij'}) = \exp\,(\alpha|t_{ij} - t_{ij'}|)$, for $\alpha < 0$. When $\boldsymbol{H}_i$ is chosen to be the identity matrix, the function $\boldsymbol{\psi}_3$ becomes

$$\Phi(\alpha) = \sum_{i=1}^{m} \boldsymbol{c}_i^\top (\boldsymbol{r}_i - \boldsymbol{\eta}_i) = 0$$

Table 3 Estimates, standard errors and robust z-statistics from the heterogeneous dispersion model for the eye surgery data.

Parameter	β_0	β_1	β_2	β_3	γ_0	γ_1	γ_2	α
Estimate	2.7445	−0.0223	−0.3144	0.4114	6.1551	−0.4583	−0.4938	−1.8484
Stderr*	0.2107	0.3367	0.0855	0.2122	0.1988	0.0803	0.1427	0.3881
z-statistic	13.0256	−0.0663	−3.6771	1.9393	30.9613	−5.7073	−3.4604	−4.7627

*Standard Error

where $\boldsymbol{c}_i = [|t_{i1} - t_{i2}| \exp(-\alpha|t_{i1} - t_{i2}|), \ldots, |t_{in_i-1} - t_{in_i}| \exp(-\alpha|t_{in_i-1} - t_{in_i}|)]^\top$ and the corresponding sensitivity matrix $S_{33} = -\sum_{i=1}^{m} \boldsymbol{c}_i^\top \boldsymbol{c}_i$.

The model that addresses the heterogeneity in two covariates of time and gas concentration level takes the following form

$$\log\left(\sigma_{ij}^2\right) = \gamma_0 + \gamma_1 \log(t_{ij}) + \gamma_2 x_{ij}. \tag{9}$$

We ran the Newton-scoring algorithm given in Section 3 and found estimates and standard errors that are listed in Table 3.

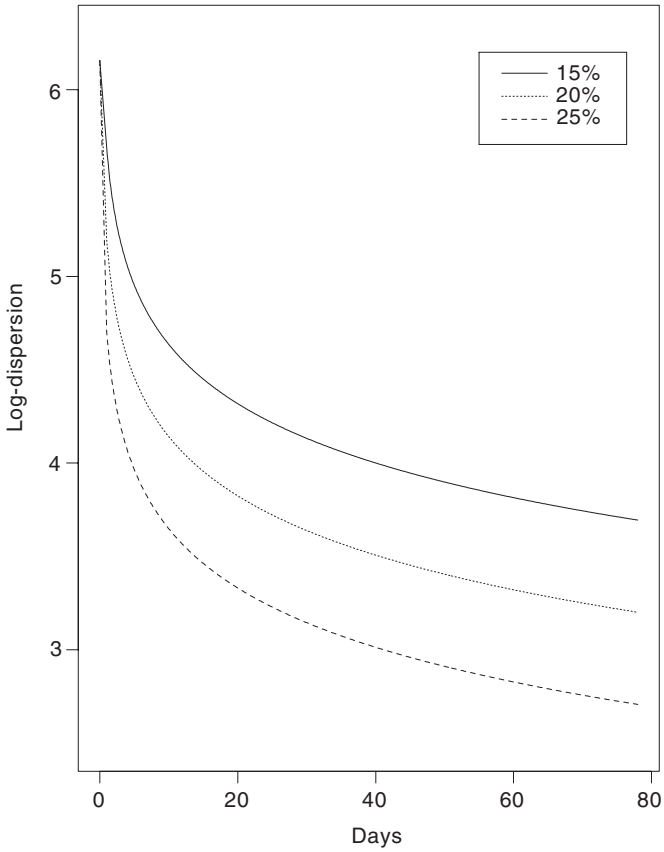

Figure 2 Fitted curves for the pattern of heterogeneous dispersion over time across three treatment levels.

Clearly, both covariates of time and treatment are significant factors attributed to the heterogeneous dispersion in model (9). Figure 2 displays the fitted curves for the pattern of dispersion profile over time across three different gas concentration levels.

Based on the model with the time-varying dispersion, our findings for other parameters are very similar to those in Song and Tan (2000). Similar to Song and Tan (2000), we found that the quadratic time term $\log^2(t_{ij})$ is significant, that the linear time $\log(t_{ij})$ is not significant, and that the gas concentration covariate is marginally insignificant, at the significance level 0.05. Also, The estimated lag-1 autocorrelation $\hat{\rho} = e^{\hat{\alpha}} = 0.1575(0.0611)$ and its z-statistic is 2.5769, suggesting that ρ is significantly different from zero.

In contrast to the simulation study, here we did not see dramatic differences between the results from the heterogeneity and homogeneity models. We gave the reason as follows. In the simulation study we chose the intercept and slope parameters to be comparable, respectively 3 and 2, so that a change on the covariate would greatly affect the size of dispersion. Therefore, the results from the homogeneity and heterogeneity models were evidently different. However, in the data analysis the intercept dominates the contribution to the dispersion over the two slope coefficients, implying that the overall dispersion remains mostly very large, and therefore no big differences appeared in the results from the two types of models.

Now we consider the residual analysis for the above model with time-varying dispersion. Panel A of Figure 3 shows the scatter-plot of the estimated standardised Pearson residuals $\hat{e}_{ij}$'s against the fitted mean values $\hat{\mu}_{ij}$, to check the distribution assumption. The dashed lines at 2 and -2 represent the asymptotic 95% upper and lower limits, respectively. The residuals seem to behave reasonably well as expected, only three of them lying outside of the region. The plot seems to be in agreement with the simplex marginal distribution.

Panel B of Figure 3 provides a rough check of the logit link function used in the proposed model, showing the scatter-plot of the estimated adjusted dependent variables $\hat{s}_{ij}$ against the estimated logit

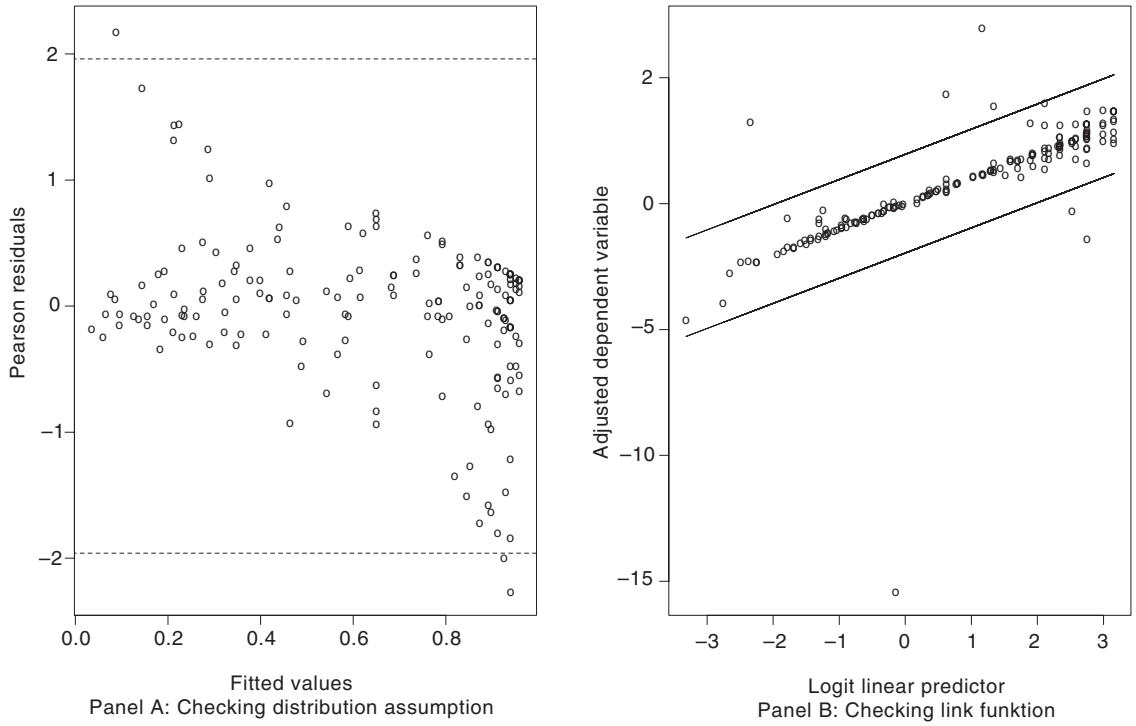

Figure 3 Diagnostic plots in the eye surgery data analysis.

linear predictor $\hat{\eta}_{ij}$. The two solid lines stand for the asymptotic 95% confident bands within which almost 96% points are contained. This clearly supports the logit link function assumption.

Checking the working correlation seems to be nontrivial, since the data are measured at irregular time points and the residuals available at a given time are sparse. So we feel that the proposed method for checking the working correlation may not be reliable here. Alternatively, Diggle's variogram plot (Diggle, 1990) may be used here to reach an appropriate conclusion. However this is not the focus of the paper, and hence the details are omitted.

7 Concluding remarks

In this paper we developed an approach to modelling the heterogeneous dispersion parameter, relaxing the usual assumption of constant dispersion in Liang and Zeger's marginal models. An extended version of GEEs was proposed to estimate the parameters in the model for dispersion. Through the analysis of the eye surgery data, we found that the dispersion can be a function over follow-up time as well as treatment arm, and that the shape of marginal distributions is time-varying in addition to the time-varying locations. This proposed method improves the modelling of longitudinal data and provides a tool for better understanding the marginal profiles of the longitudinal continuous proportional data.

The extended GEEs in this paper was developed under the assumption of no missing values in data. Since missing values often occur in longitudinal studies in practice, it would be of great interest to further extend the proposed GEEs to conduct data analysis with missing values. It is known that GEEs produce consistent estimators for the model parameters when missing values are completely random and ignored in the analysis. However, when data contain random missing values or informative missing values, the consistency for the GEEs estimators is generally no longer valid. Resolving this issue has been an active research topic in the longitudinal data analysis. For example, Robins et al. (1995) suggested the inverse probability weighted GEEs that produce consistent estimates if the drop-out process is properly modelled. Another approach suggested by Paik (1997) is to impute the missing values by the conditional expectation given the observed data. More references can be found in Diggle et al. (2002), Verbeke and Molenberghs (2000), or Ziegler et al. (1998).

Acknowledgements The authors are grateful to the two referees for their valuable suggestions and comments that led to an improvement of the paper. The authors thank Dr. Sanford Meyers for making his data available for inclusion as an example. The first two authors' research was supported by the National Science and Engineering Research Council of Canada.

Appendix

A Godambe information matrix

This section gives the components of Godambe information matrix needed for computing the estimated standard errors for estimates and hence for constructing Wald test statistics.

The sensitivity matrix is a 3×3 block matrix,

$$\boldsymbol{S} = E\left\{\frac{\partial \Upsilon(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\top}}\right\} = \begin{pmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} & \boldsymbol{S}_{13} \\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22} & \boldsymbol{S}_{23} \\ \boldsymbol{S}_{31} & \boldsymbol{S}_{32} & \boldsymbol{S}_{33} \end{pmatrix},$$

where clearly $\boldsymbol{S}_{12} = \boldsymbol{0}$, $\boldsymbol{S}_{13} = \boldsymbol{0}$, and $\boldsymbol{S}_{23} = \boldsymbol{0}$. Also the block $\boldsymbol{S}_{21} = \boldsymbol{0}$ because $Eu_{ij} = 0$. So in general the $\boldsymbol{S}$ matrix takes the form

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{11} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_{22} & \boldsymbol{0} \\ \boldsymbol{S}_{31} & \boldsymbol{S}_{32} & \boldsymbol{S}_{33} \end{pmatrix},$$

and its inverse matrix is

$$\boldsymbol{S}^{-1} = \begin{pmatrix} \boldsymbol{S}_{11}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{S}_{22}^{-1} & \mathbf{0} \\ -\boldsymbol{S}_{33}^{-1}\boldsymbol{S}_{31}\boldsymbol{S}_{11}^{-1} & -\boldsymbol{S}_{33}^{-1}\boldsymbol{S}_{32}\boldsymbol{S}_{22}^{-1} & \boldsymbol{S}_{33}^{-1} \end{pmatrix},$$

provided that all diagonal blocks are invertible. When the distribution of $r_{ij}r_{ij'}$ is independent of the mean and dispersion parameters, both $\boldsymbol{S}_{31}$ and $\boldsymbol{S}_{32}$ are $\mathbf{0}$. Therefore the matrix $\boldsymbol{S}$ becomes a block-diagonal matrix with

$$\boldsymbol{S}_{11} = -\sum_{i=1}^{m} \boldsymbol{D}_i^{\top}\boldsymbol{A}_i\boldsymbol{V}_i^{-1}\boldsymbol{A}_i\boldsymbol{D}_i,$$

$$\boldsymbol{S}_{22} = -\sum_{i=1}^{m} \left(\frac{\partial \boldsymbol{\sigma}_i^{\top}}{\partial \boldsymbol{\gamma}}\right)\Sigma_i^{-1}\left(\frac{\partial \boldsymbol{\sigma}_i}{\partial \boldsymbol{\gamma}^{\top}}\right)$$

and

$$\boldsymbol{S}_{33} = -\sum_{i=1}^{m} \left(\frac{\partial \boldsymbol{\eta}_i^{\top}}{\partial \boldsymbol{\alpha}}\right)\boldsymbol{H}_i^{-1}\left(\frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\alpha}^{\top}}\right).$$

The variability matrix $\boldsymbol{R}$ is also a 3×3 block matrix,

$$\boldsymbol{V} = E\{\Upsilon(\boldsymbol{\theta})\Upsilon^{\top}(\boldsymbol{\theta})\} = \begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} & \boldsymbol{V}_{13} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} & \boldsymbol{V}_{23} \\ \boldsymbol{V}_{31} & \boldsymbol{V}_{32} & \boldsymbol{V}_{33} \end{pmatrix}.$$

The nine blocks are detailed as follows.

$$\boldsymbol{V}_{11} = E\{\boldsymbol{\psi}_1\boldsymbol{\psi}_1^{\top}\} = \sum_{i=1}^{m} \boldsymbol{D}_i^{\top}\boldsymbol{A}_i\boldsymbol{V}_i^{-1}\mathrm{cov}\,(\boldsymbol{w}_i)\boldsymbol{V}_i^{-1}\boldsymbol{A}_i\boldsymbol{D}_i,$$

$$\boldsymbol{V}_{12} = E\{\boldsymbol{\psi}_1\boldsymbol{\psi}_2^{\top}\} = \sum_{i=1}^{m} \boldsymbol{D}_i^{\top}\boldsymbol{A}_i\boldsymbol{V}_i^{-1}\mathrm{cov}\,(\boldsymbol{w}_i, \boldsymbol{d}_i)\,\Sigma_i^{-1}\left(\frac{\partial \boldsymbol{\sigma}_i}{\partial \boldsymbol{\gamma}^{\top}}\right),$$

$$\boldsymbol{V}_{13} = E\{\boldsymbol{\psi}_1\boldsymbol{\psi}_3^{\top}\} = \sum_{i=1}^{m} \boldsymbol{D}_i^{\top}\boldsymbol{A}_i\boldsymbol{V}_i^{-1}\mathrm{cov}\,(\boldsymbol{w}_i, \boldsymbol{r}_i)\,\boldsymbol{H}_i^{-1}\left(\frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\alpha}^{\top}}\right),$$

$$\boldsymbol{V}_{22} = E\{\boldsymbol{\psi}_2\boldsymbol{\psi}_2^{\top}\} = \sum_{i=1}^{m} \left(\frac{\partial \boldsymbol{\sigma}_i^{\top}}{\partial \boldsymbol{\gamma}}\right)\Sigma_i^{-1}\mathrm{cov}\,(\boldsymbol{d}_i)\,\Sigma_i^{-1}\left(\frac{\partial \boldsymbol{\sigma}_i}{\partial \boldsymbol{\gamma}^{\top}}\right),$$

$$\boldsymbol{V}_{23} = E\{\boldsymbol{\psi}_2\boldsymbol{\psi}_3^{\top}\} = \sum_{i=1}^{m} \left(\frac{\partial \boldsymbol{\sigma}_i^{\top}}{\partial \boldsymbol{\gamma}}\right)\Sigma_i^{-1}\mathrm{cov}\,(\boldsymbol{d}_i, \boldsymbol{r}_i)\,\boldsymbol{H}_i^{-1}\left(\frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\alpha}^{\top}}\right),$$

$$\boldsymbol{V}_{33} = E\{\boldsymbol{\psi}_3\boldsymbol{\psi}_3^{\top}\} = \sum_{i=1}^{m} \left(\frac{\partial \boldsymbol{\eta}_i^{\top}}{\partial \boldsymbol{\alpha}}\right)\boldsymbol{H}_i^{-1}\mathrm{cov}\,(\boldsymbol{r}_i)\,\boldsymbol{H}_i^{-1}\left(\frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\alpha}^{\top}}\right).$$

Because of symmetry, $\boldsymbol{V}_{21} = \boldsymbol{V}_{12}^{\top}$, $\boldsymbol{V}_{31} = \boldsymbol{V}_{13}^{\top}$, and $\boldsymbol{V}_{32} = \boldsymbol{V}_{23}^{\top}$.

It is noted that $\mathrm{cov}\,(\boldsymbol{w}_i) = \mathrm{diag}\,\{v(\mu_{ij})\}\,\mathrm{cov}\,(\boldsymbol{u}_i)\,\mathrm{diag}\,\{v(\mu_{ij})\}$, and an estimate of $\mathrm{cov}\,(\boldsymbol{u}_i)$ is obtained by plugging the estimates $\hat{\mu}_{ij}$ and replacing $\mathrm{cov}\,(\boldsymbol{u}_i)$ by $\hat{\boldsymbol{u}}_i\hat{\boldsymbol{u}}_i^{\top}$ in the expression. The same approach is applied to estimate the remaining blocks of $\boldsymbol{V}$.

B Proof of mean-variance relation

This section presents the proof for the formula $\mathrm{var}\{d(Y_{ij};\mu_{ij})\} = 2(\sigma_{ij}^2)^2$, $Y_{ij} \sim S^-(\mu_{ij}, \sigma_{ij}^2)$. From the appendix of Song and Tan (2000), suppressing coordinates, $E\{d(Y;\mu)\} = \sigma^2$, and hence it is sufficient to show that $E\{d^2(Y;\mu)\} = 3(\sigma^2)^2$.

A simple algebra leads to

$$
\begin{aligned}
E\{d^2(Y;\mu)\} &= \int_0^1 d^2(y;\mu)\, p(y;\mu,\sigma^2)\, \mathrm{d}y \\
&= \sqrt{\frac{\lambda}{2\pi}}\frac{(1+\xi)^4}{\xi^4}\int_0^\infty \{x^{\frac{3}{2}} + (1-4\xi)\, x^{\frac{1}{2}} + 2\xi(3\xi-2)\, x^{-\frac{1}{2}} \\
&\quad + 2\xi^2(3-2\xi)\, x^{-\frac{3}{2}} + \xi^3(\xi-4)\, x^{-\frac{5}{2}} + \xi^4 x^{-\frac{7}{2}}\} f(x;\xi,\lambda)\, \mathrm{d}x,
\end{aligned}
$$

where $\lambda = 1/\sigma^2$ and

$$
f(x;\xi,\lambda) = \exp\left\{-\frac{\lambda}{2}\frac{(1+\xi)^2}{\xi^2}\frac{(x-\xi)^2}{x}\right\}.
$$

Using formulas (5.41)–(5.43) of (Jørgensen, 1997), we obtain

$$
\int_0^\infty x^{\frac{3}{2}} f(x;\xi,\lambda)\, \mathrm{d}x = \left(\frac{2\pi}{\lambda}\right)^{\frac{1}{2}} \frac{\lambda^2\xi^3(1+\xi)^4 + 3\lambda\xi^4(1+\xi)^2 + 3\xi^5}{\lambda^2(1+\xi)^5}
$$

and

$$
\int_0^\infty x^{-\frac{7}{2}} f(x;\xi,\lambda)\, \mathrm{d}x = \left(\frac{2\pi}{\lambda}\right)^{\frac{1}{2}} \frac{\lambda^2(1+\xi)^4 + 3\lambda\xi(1+\xi)^2 + 3\xi^2}{\lambda^2\xi^2(1+\xi)^5}.
$$

Plugging these results and those from Song and Tan (2000), we get $E\{d^2(Y;\mu)\} = 3(\sigma^2)^2$.

References

Artes, R. and Jørgensen, B. (2000). Longitudinal data estimating equations for dispersion models. *Scandinavian Journal of Statistics* **27**, 321–334.

Barndorff-Nielsen, O. E. and Jørgensen, B. (1991). Some parametric models on the simplex. *Journal of Multivariate Analysis* **39**, 106–116.

Crowder, M. (1987). On linear and quadratic estimating functions. *Biometrika* **74**, 591 – 597.

Diggle, P. J. (1990). *Time Series: A Biostatistical Introduction.* Oxford University Press, Oxford.

Diggle, P. J. Heagerty, P., Liang, K.-Y. and Zeger, S. L. (2002). *The Analysis of Longitudinal Data, 2nd ed.* Oxford University Press, Oxford.

Jørgensen, B. (1997). *The Theory of Dispersion Models.* Chapman Hall, London.

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

McCullagh, P. and Nelder, J. A. (1989). *Generalised Linear Models, 2nd ed.* Chapman and Hall, London.

Meyers, S. M., Ambler, J. S., Tan, M., Werner, J. C. and Huang, S. S. (1992). Variation of perfluorpropane disappearance after vitrectomy. *Retina* **12**, 359–363.

Paik, M. C. (1992). Parametric variance function estimation for nonnormal repeated measurement data. *Biometrics* **48**, 19–30.

Paik, M. C. (1997). The generalized estimating equation approach when data are not missing completely at random. *Journal of American Statistical Association* **92**, 1320–1329.

Prentice, R. L. and Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* **47**, 825–839.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **40**, 1025–1035.

Smyth, G. K. (1989). Generalised linear models with varying dispersion. *Journal of the Royal Statistical Society, Series B* **51**, 47–60.

Song, P. X.-K. and Tan, M. (2000). Marginal models for longitudinal continuous proportional data. *Biometrics* **56**, 496–502.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* **61**, 439–447.

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