UCLA

Department of Statistics Papers

Title

Orthogonality and Transformations in Variance Components Models

Permalink

<https://escholarship.org/uc/item/2q19j75x>

Authors

P. J. Solomon
Jeremy M.G. Taylor

Publication Date

2011-10-25

Orthogonality and transformations in variance components models

P.J. Solomon¹ and J.M.G. Taylor²

Department of Statistics¹, University of Adelaide,
Adelaide, SA 5005, Australia

psolomon@maths.adelaide.edu.au

Department of Biostatistics², UCLA School of Public Health,
Los Angeles, CA, 90095, USA

jeremy@sunlab.ph.ucla.edu

December 17, 1997

Abstract

In this paper we consider variance components and other models for repeated measures in which a general transformation is applied to the response variable. Using Cox and Reid's concept of parameter orthogonality (1987, *JRSS B* **49**, 1-18) and some approximations to the information matrix we show that the intraclass correlation coefficient in the one-way model is robust to the choice of transformation. This robustness result generalises to a vector of parameters determining the correlation structure, to more complex variance components models, to multivariate normal models, to some longitudinal models and models involving linear regression functions. The results suggest a natural way to parametrise the covariance structure in repeated measures models is in terms of the variance and the correlation determined by separate sets of parameters.

Key Words: intraclass correlation, longitudinal data, parameter orthogonality, power transformations, repeated measures, variance components.

1 Introduction

Components of variance models are important in many fields of application. The simplest variance component model for repeated measures is the normal theory

⁰Author listing is alphabetical.

one-way model $Y_{ij} = \mu + a_i + e_{ij}$, $j = 1, \dots, n_i; i = 1, \dots, I$, where $a_i \sim N(0, \sigma_a^2)$ independently of $e_{ij} \sim N(0, \sigma_e^2)$. The three parameters in this model are μ, σ_e^2 and σ_a^2 , although an alternative parametrisation for the two variance parameters could be in terms of the intraclass correlation $\rho = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2)$ and the sum of the variance components $\theta = \sigma_a^2 + \sigma_e^2$. The intraclass correlation is often of substantial interest in itself, for example in genetics it measures the proportion of an effect that could be attributed to heritability.

The concept of intraclass correlation generalises in more complex variance components models to ratios of sums of variances. In the simplest case described above it is clear that estimates of μ, σ_a^2 and σ_e^2 would change substantially if Y was replaced by a transformation of Y , for example $\log Y$, or $Y^{1/2}$ or $\log(Y + c)$ for some constant c . By contrast, the intraclass correlation is a dimensionless parameter, and we expect therefore that it is less affected by transformation. Another way to think about this is that robust estimation of the intraclass correlation is likely to be possible because it is a quantity which is inherent to the data rather than the specific choice of model. For the Box-Cox family of transformation models $(Y^\psi - 1)/\psi$, Solomon (1985) conjectured that the intraclass correlation was orthogonal to ψ . This was formally proved in Taylor *et al* (1996) for the case $\psi = 0$, who also showed it to be approximately true for other values of ψ .

In important work by Cox and Reid (1987) the following framework is considered: there is a set of parameters $\psi, \phi_1, \dots, \phi_K$ and the aim is to reparametrise the ϕ s to be asymptotically orthogonal to ψ . A new set of parameters $\psi, \lambda_1, \dots, \lambda_K$ is created, such that ψ is orthogonal to each λ . There are a number of reasons why one might do this: it might lead to better inference concerning ψ or the λ s, it might improve numerical methods of estimation, it might enable nuisance parameters to be properly handled, or it might simply lead to better insight into the structure of the problem.

Cox and Reid (1987) specify a set of differential equations which need to be solved to give such an orthogonal parametrisation. Let $i_{\alpha\beta}$ denote an element of the expected information matrix $-E\{\partial^2 \log(L)/\partial\alpha\partial\beta\}$. Then the differential equations in Cox and Reid's notation are

$$\sum_{j=1}^K i_{\phi_j \phi_k} \frac{\partial \phi_j}{\partial \psi} = -i_{\psi \phi_k}$$

for $k = 1, \dots, K$. In general one only needs to find a solution to these differential equations. Such a solution will involve arbitrary constants of integration, which are

formulated as convenient functions of the λ s. In one of their examples Cox and Reid focus on power transformations, in particular the model $Y_i^\psi = \mu + e_i$, $e_i \sim N(0, \sigma_e^2)$. They obtain an approximation to the differential equations which can be solved. This leads to the reparametrisation of (σ_e^2, μ) to (λ_0, λ_1) so that the model is represented in the form $Y_i^\psi \sim N(\lambda_1^\psi, \psi^2 \lambda_1^{(2\psi-2)} \lambda_0)$. With this parametrisation, λ_1 and λ_0 are approximately orthogonal to ψ . We note that they have the dimensions of Y and Y^2 respectively, and are approximately the mean and variance of Y . With this parametrisation, estimation of λ_1 and λ_0 will be robust with respect to the choice of ψ .

In this paper we study the implications of Cox and Reid's notion of parameter orthogonality for robust interpretation of parameters in components of variance models. We do this by extending the class of models considered by Cox and Reid in two directions: we consider more general components of variance models for repeated measures, and further we consider general transformations of the form $h(Y, \psi)$, where h is a known, sufficiently differentiable function of two variables and invertible with respect to Y . Also, the derivative of h with respect to ψ should be nonzero. The power transformation family is a special case of this model. More generally, for any two transformations g_1 and g_2 we could take $h(Y, \psi) = \psi g_1(Y) + (1 - \psi)g_2(Y)$ or $h(Y, \psi) = g_1(Y)^\psi g_2(Y)^{1-\psi}$. We think of the specification of the model using $h(Y, \psi)$, with h unspecified, as a means to consider any transformation of Y which might be of interest for example $\log(Y)$, $\exp(Y)$ or $\tan^{-1}(Y)$.

The paper is organised as follows: in section 2 we obtain the orthogonalising differential equations for the general transformation model with one-way structure, orthogonalising with respect to the transformation parameter ψ . We show that these equations are satisfied by the intraclass correlation, and obtain a set of orthogonal parameters which are functions of the variance components. We then extend the results to more general variance component models with multiple crossed and nested effects, as well as to multivariate normal models and longitudinal models. In section 3 we consider the case where the mean μ is replaced by a linear regression function. In section 4 we determine a parameter which is orthogonal to the intraclass correlation in the one-way model. A brief discussion is given in section 5.

2 Parameters orthogonal to transformation

2.1 The one-way model

Let Y_{ij} denote the j th observation on subject i , and let Y_i denote the vector of observations for subject i . Assume

$$h(Y_{ij}, \psi) = \mu + a_i + e_{ij}$$

with $i = 1, \dots, I$, $j = 1, \dots, n_i$ and $N = \sum n_i$. We re-express the covariance parameters as θ and ρ , where $\theta = \sigma_e^2 + \sigma_a^2$ and $\rho = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2)$. We will assume $\sigma_a^2 > 0$, $\sigma_e^2 > 0$ and $n_i > 1$ for some i . Let Ω_i denote the covariance matrix of the set of observations for subject i so that $\Omega_i = \theta M$, where M is an $n_i \times n_i$ correlation matrix with ones on the diagonal and ρ everywhere else.

Let $h_{10} = \partial h / \partial Y$, $h_{01} = \partial h / \partial \psi$, $h_{11} = \partial^2 h / \partial \psi \partial Y$ and h^{-1} be the inverse of h with respect to its first component. For ease of notation we suppress the dependence of h^{-1} on ψ . The log likelihood $\log L$ is given by

$$-\frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_i \log \det(\Omega_i) - \frac{1}{2} \sum_i (h_i - \mu 1_i)' \Omega_i^{-1} (h_i - \mu 1_i) + \sum_i \sum_j \log h_{10}(Y_{ij}, \psi)$$

where h_i is the vector of length n_i of $h(Y_i, \psi)$, 1_i is a vector of n_i ones and the last term in the sum is from the Jacobian. It is well-known that $i_{\mu\mu} = \sum_i 1_i' \Omega_i^{-1} 1_i$ and that $i_{\mu\theta} = i_{\mu\rho} = 0$.

The general form for the expected information with respect to variance parameters ϕ_j and ϕ_k is

$$i_{\phi_j \phi_k} = \frac{1}{2} \sum_i \text{tr} \left(\Omega_i^{-1} \frac{\partial \Omega_i}{\partial \phi_j} \Omega_i^{-1} \frac{\partial \Omega_i}{\partial \phi_k} \right).$$

Throughout the algebraic development in this article we will make use of the following two well-known results. First that $\partial \Omega^{-1} / \partial \phi = -\Omega^{-1} (\partial \Omega / \partial \phi) \Omega^{-1}$, and second that for multivariate normal Z , with $Z \sim N(0, \Omega)$ and symmetric matrix A then $E[Z'AZ] = \text{tr}(\Omega A)$.

The form of Ω_i for the one-way model is such that $\partial \Omega_i / \partial \theta = \Omega_i / \theta$, which leads to considerable simplification, giving

$$i_{\theta\theta} = \frac{N}{2\theta^2}, \tag{1}$$

$$i_{\theta\rho} = \frac{1}{2\theta} \sum_i \frac{\rho(1-n_i)n_i}{(1-\rho)(1-\rho+\rho n_i)} \quad (2)$$

and

$$i_{\rho\rho} = \frac{1}{2} \sum_i \frac{(n_i-1)n_i(1+(n_i-1)\rho^2)}{(1-\rho)^2(1-\rho+\rho n_i)^2}.$$

To simplify the notation, we write $i_{\theta\rho} = \frac{1}{2\theta}f(\rho)$ and $i_{\rho\rho} = \frac{1}{2}g(\rho)$.

Some approximations are required to obtain the expected information terms involving ψ . We have

$$\frac{\partial \log L}{\partial \psi} = - \sum_i h_{01}(Y_i, \psi)' \Omega_i^{-1} (h_i - \mu 1_i) + \sum_i \sum_j \frac{h_{11}(Y_{ij}, \psi)}{h_{10}(Y_{ij}, \psi)}.$$

This leads to

$$i_{\psi\theta} = -\frac{1}{\theta} E\left\{ \sum_i h_{01}(Y_i, \psi)' \Omega_i^{-1} (h_i - \mu 1_i) \right\},$$

which equals

$$-\frac{1}{\theta} E\left\{ \sum_i \sum_j \frac{h_{11}(Y_{ij}, \psi)}{h_{10}(Y_{ij}, \psi)} \right\}, \quad (3)$$

because the expected score $E\{\partial \log(L)/\partial \psi\}$ equals zero. Expanding h_{11} and h_{10} using first-order Taylor series about $h^{-1}(\mu)$, approximately the mean of Y when θ is small, we approximate (3) by

$$-\frac{1}{\theta} \sum_i \sum_j \frac{h_{11}(h^{-1}(\mu), \psi)}{h_{10}(h^{-1}(\mu), \psi)}$$

which we write as $-(N/\theta)(h_{11}/h_{10})$.

To obtain an expression for $i_{\psi\rho}$ we again approximate $h_{01}(Y_{ij}, \psi)$ using a Taylor series expansion to give

$$h_{01}(h^{-1}(\mu), \psi) + (a_i + e_{ij}) \frac{h_{11}(h^{-1}(\mu), \psi)}{h_{10}(h^{-1}(\mu), \psi)}.$$

This gives

$$i_{\psi\rho} \approx -\frac{h_{11}}{h_{10}} \sum_i \text{tr} \left(\Omega_i^{-1} \frac{\partial \Omega_i}{\partial \rho} \right) = -\frac{h_{11}}{h_{10}} f(\rho).$$

Similarly

$$i_{\psi\mu} = - \sum_i E\{1_i' \Omega_i^{-1} h_{01}(Y_i, \psi)\},$$

which is approximated by

$$-\sum_i 1'_i \Omega_i^{-1} 1'_i h_{01}(h^{-1}(\mu), \psi).$$

The differential equation for μ is

$$\sum_i 1'_i \Omega_i^{-1} 1_i \frac{\partial \mu}{\partial \psi} = \sum_i 1'_i \Omega_i 1_i h_{01}(h^{-1}(\mu), \psi)$$

which is solved by $\mu = h(\lambda_0, \psi)$, where λ_0 is the new parameter. The differential equations for θ and ρ are then

$$\frac{N}{2\theta^2} \frac{\partial \theta}{\partial \psi} + \frac{f(\rho)}{2\theta} \frac{\partial \rho}{\partial \psi} = \frac{N}{\theta} \frac{h_{11}}{h_{10}}$$

and

$$\frac{f(\rho)}{2\theta} \frac{\partial \theta}{\partial \psi} + \frac{g(\rho)}{2} \frac{\partial \rho}{\partial \psi} = f(\rho) \frac{h_{11}}{h_{10}}.$$

Multiplying the first equation through by $\theta f(\rho)$, multiplying the second equation by N and subtracting the resulting equations gives

$$\frac{1}{2} \frac{\partial \rho}{\partial \psi} \{f(\rho)^2 - Ng(\rho)\} = 0.$$

Either $\partial \rho / \partial \psi = 0$ or $\{f(\rho)^2 - Ng(\rho)\} = 0$ so in either case ρ is a constant with respect to ψ . In fact in this case application of Cauchy-Schwartz inequality leads to $f(\rho)^2 < Ng(\rho)$, showing that $\partial \rho / \partial \psi = 0$. Substituting this into the differential equations gives the solution $\rho = \lambda_1$ and $\theta = \lambda_2 \{h_{10}(\lambda_0, \psi)\}^2$, where λ_1 and λ_2 are arbitrary constants independent of ψ which arise from solving the equations. We can then write $\lambda_0 = h^{-1}(\mu)$, $\lambda_1 = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2)$ and $\lambda_2 = (\sigma_a^2 + \sigma_e^2) / \{h_{10}(h^{-1}(\mu), \psi)\}^2$.

The three parameters λ_1 , λ_0 and λ_2 are approximately orthogonal to ψ , and represent the intraclass correlation, and approximately the mean and variance of Y ; λ_2 is a measure of how the variance is rescaled by the derivative of h . These results demonstrate why the intraclass correlation can be robustly estimated irrespective of the choice of transformation.

Adequacy of the approximations: We have employed some approximations to linearise h or its derivatives, and anticipate that the approximations may break down when h has marked curvature or the variability of the observations is relatively large. Some such departures from the orthogonality results were confirmed by numerical work. We simulated data from the untransformed one-way model, assuming the variance components ranged from small ($\sigma_a^2 = \sigma_e^2 = 0.04$) to fairly large ($\sigma_a^2 = \sigma_e^2 = 1$) for various combinations of group sizes I and replicates n , various values of μ and balanced and unbalanced models. We considered various transformations g including Box-Cox, exp and $\log(Y + c)$, and fit the one-way model to the transformed data, obtaining estimates of λ_0 , λ_1 and λ_2 for each g .

We observed that $\lambda_0 = g^{-1}(\mu)$, $\lambda_1 = \rho$ and λ_2 are robust to transformation for small to moderate variance components, or when σ_a^2 is large relative to σ_e^2 , virtually irrespective of any lack of balance in the data. The estimates are approximately stable also when σ_e^2 is large relative to σ_a^2 . Overall, λ_0 and ρ are the parameters most stable to transformation with changes typically occurring in the second or third decimal place, with λ_2 more sensitive to marked curvature in g , such as can occur with $\exp(Y)$. When both variance components are large and g has marked curvature, especially combined with extreme imbalance in the number of replicates and very small I (≤ 5), some departures from orthogonality were observed to the order of more than 10% of the true parameter values.

In general, we observed from our empirical study that the estimates of the orthogonal parameters appear stable whenever $\text{var}(Y)g''(g^{-1}(\mu))/g'(g^{-1}(\mu))$ is less than 0.6 in magnitude, where g'' and g' represent the first and second derivatives of g . This is a combined overall measure of the spread of the observations and the curvature of g at $g^{-1}(\mu)$.

2.2 Models with general covariance structure

We now consider more complex variance components models of the form

$$h(Y_{ij}, \psi) = \mu + \epsilon_{ij},$$

where ϵ_{ij} can include many variance components which may be nested or crossed, and serial correlation can also enter. Let ϵ_i be the vector of components of ϵ_{ij} and let $\Omega_i = \text{Var}(\epsilon_i)$. We assume Ω_i takes the form $\Omega_i = \sigma^2 M_i(\rho)$, where σ^2 is a scalar equal to the variance of $h(Y_{ij}, \psi)$ and $M_i(\rho)$ is a correlation matrix determined by

a vector of parameters ρ , which in many examples would represents ratios of sums of variances. Let K be the dimension of ρ .

Following similar algebra to that in section 2.1 and exploiting the fact that $\partial\Omega_i/\partial\sigma^2 = \Omega_i/\sigma^2$ gives

$$i_{\sigma^2\sigma^2} = \frac{1}{2} \sum_i \text{tr} \left(\Omega_i^{-1} \frac{\partial\Omega_i}{\partial\sigma^2} \Omega_i^{-1} \frac{\partial\Omega_i}{\partial\sigma^2} \right)$$

which simplifies to

$$i_{\sigma^2\sigma^2} = \frac{N}{2\sigma^4}.$$

The other relevant information terms are

$$i_{\sigma^2\rho_j} = \frac{1}{2\sigma^2} \sum_i \text{tr} \left(\Omega_i^{-1} \frac{\partial\Omega_i}{\partial\rho_j} \right),$$

$$i_{\rho_j\rho_k} = \frac{1}{2} \sum_i \text{tr} \left(\Omega_i^{-1} \frac{\partial\Omega_i}{\partial\rho_j} \Omega_i^{-1} \frac{\partial\Omega_i}{\partial\rho_k} \right).$$

Proceeding as for the one-way case, some approximations based on first-order Taylor series expansions about $h^{-1}(\mu)$ give

$$i_{\psi\sigma^2} \approx -\frac{N}{\sigma^2} \frac{h_{11}(h^{-1}(\mu), \psi)}{h_{10}(h^{-1}(\mu), \psi)}$$

and

$$i_{\psi\rho_j} \approx -\frac{h_{11}(h^{-1}(\mu), \psi)}{h_{10}(h^{-1}(\mu), \psi)} \sum_i \text{tr} \left(\Omega_i^{-1} \frac{\partial\Omega_i}{\partial\rho_j} \right).$$

The differential equation involving μ is exactly the same as for the one-way model leading to the solution $\mu = h(\lambda_0, \psi)$. For the covariance parameters there are $K + 1$ differential equations to be solved namely,

$$i_{\sigma^2\sigma^2} \frac{\partial\sigma^2}{\partial\psi} + \sum_{j=1}^K i_{\sigma^2\rho_j} \frac{\partial\rho_j}{\partial\psi} = -i_{\psi\sigma^2}$$

and

$$i_{\sigma^2 \rho_k} \frac{\partial \sigma^2}{\partial \psi} + \sum_{j=1}^K i_{\rho_j \rho_k} \frac{\partial \rho_j}{\partial \psi} = -i_{\psi \rho_k}$$

for $k = 1, \dots, K$.

Although in principle it is possible to establish whether or not a unique general solution can be obtained analogous to the one-way case, for our purposes we only need establish that a solution is given when $\partial \rho_j / \partial \psi = 0$ for all j . In this case all the $K + 1$ equations reduce to

$$\frac{1}{\sigma^2} \frac{\partial \sigma^2}{\partial \psi} = 2 \frac{h_{11}(h^{-1}(\mu), \psi)}{h_{10}(h^{-1}(\mu), \psi)},$$

the solution of which is $\sigma^2 = \lambda_{K+1} \{h_{10}(\lambda_0, \psi)\}^2$. Thus the set of $K + 1$ differential equations can be solved by $\partial \rho_j / \partial \psi = 0$ for all j , giving a solution $\rho_j = \lambda_j$ $j = 1, \dots, K$. This demonstrates that parameters determining only the correlation structure of the repeated measures are orthogonal to ψ , and thus would be stable to different choices of transformation, at least to within the order of approximation used here.

Example 1: A hierarchical model with three nested random effects is a simple example of a model which falls under the above correlation framework. For instance Y might be blood pressure measurements, replicated for each individual i at each visit j in a longitudinal study. A model for such a situation is

$$h(Y_{ijk}, \psi) = \mu + a_i + b_{ij} + e_{ijk},$$

where a_i are the individual effects distributed as $N(0, \sigma_a^2)$, with σ_a^2 the between-individual component of variance, b_{ij} represent repeated measurements made at visits j for individual i and are normal with mean zero and between-visit within-individual variance component σ_b^2 ; the e_{ijk} are also normal with mean zero and between-replicate within-individual-visit variance component σ_e^2 . The random variables are assumed to be independently distributed, and we are ignoring here the possibility of serial correlation. The variance is then $\text{var}(h(Y_{ijk}, \psi)) = \sigma_a^2 + \sigma_b^2 + \sigma_e^2$, with covariances $\text{cov}(h(Y_{ijk}, \psi), h(Y_{ijk'}, \psi)) = \sigma_a^2 + \sigma_b^2$ and $\text{cov}(h(Y_{ijk}, \psi), h(Y_{ij'k'}, \psi)) = \sigma_a^2$. In this example the two ρ s which determine the correlation matrix are $(\sigma_a^2 + \sigma_b^2) / (\sigma_a^2 + \sigma_b^2 + \sigma_e^2)$ and $\sigma_a^2 / (\sigma_a^2 + \sigma_b^2 + \sigma_e^2)$. Alternatively, the two ρ s could be parametrised as $\sigma_b^2 / (\sigma_a^2 + \sigma_b^2 + \sigma_e^2)$ and $\sigma_a^2 / (\sigma_a^2 + \sigma_b^2 + \sigma_e^2)$. The orthogonality results we have obtained provide a general formal proof of empirical

results obtained by Solomon (1985) that the ρ s are robust to possibly incorrect transformation.

Example 2: For longitudinal data with measurements at times t_{ij} , Diggle (1988) suggested using the model $Y_{ij} = X_i(t_{ij})\beta + a_i + W_{ij} + e_{ij}$, where $a_i \sim N(0, \sigma_a^2)$, $e_{ij} \sim N(0, \sigma_e^2)$ and W is a stationary Gaussian stochastic process, with $\text{var}(W_{ij}) = \sigma_W^2$ and $\text{cov}(W_{ij}, W_{ik}) = \sigma_W^2 V_{jk}(\rho)$. Here V is the correlation matrix which is determined by a vector of parameters $\rho_1, \dots, \rho_{K-2}$. Usually V_{jk} would depend on the j and k only through the difference in time $|t_{ij} - t_{ik}|$, for example Diggle uses $V_{jk} = \exp(-\rho_1 |t_{ij} - t_{ik}|^{\rho_2})$. Transformations of Y can have an effect on the mean structure $X_i(t_{ij})\beta$, misspecification of which might induce spurious correlation. For the purposes of this article we ignore this and assume the model

$$h(Y_{ij}, \psi) = \mu + a_i + W_{ij} + e_{ij}.$$

Defining new parameters $\sigma^2 = \sigma_a^2 + \sigma_W^2 + \sigma_e^2$, $\rho_{K-1} = \sigma_a^2 / (\sigma_a^2 + \sigma_W^2 + \sigma_e^2)$ and $\rho_K = \sigma_W^2 / (\sigma_a^2 + \sigma_W^2 + \sigma_e^2)$, the new model falls exactly under the framework considered here with $M_{jk} = \rho_{K-1} + \rho_K V_{jk}$ for $j \neq k$ and $M_{jj} = 1$. There is a single overall variance parameter σ^2 unlinked from the K parameters which determine the correlation. Thus the ρ s are orthogonal to ψ , and in particular $\rho_1, \dots, \rho_{K-2}$, which determine the correlation of structure of W , are orthogonal to the choice of transformation. Also ρ_{K-1} and ρ_K which represent fractional parts of the total variance, which is one interpretation of intraclass correlation coefficients, are orthogonal to ψ .

There are other approaches to longitudinal data using models of a different structure. For example Taylor *et al* (1994) used the above model with W as a non-stationary stochastic process. They used either Brownian motion or an integrated Ornstein-Uhlenbeck process for W . This does not fall under the framework considered in the present paper, because the parameters which determine the correlation of Y cannot be separated from those which determine the variance.

Another popular approach for longitudinal data is a random effects model, $Y_{it} = X_i(t)\beta + a_i + b_it + e_{ij}$, with (a_i, b_i) bivariate normal. This model also does not fall under the framework considered in this paper, because the parameters determining the variance and correlation are linked and because it has non-constant variance. It would be interesting to see if there are quantities for random effects models of this type which are orthogonal to transformations; a possible candidate for this simple model would be the correlation between a_i and b_i .

2.3 Extension to multivariate normal models

We now consider the case where the mean and variance of the repeated measures can differ. This might occur in longitudinal data where the mean of Y_{ij} depends on j , the time at which the measurement is taken. Similarly the variance is also dependent on j .

Assume $h(Y_{ij}, \psi) = \mu_j + \sigma_j \epsilon_{ij}$, $j = 1, \dots, n$; $i = 1, \dots, I$ and denote the correlation matrix of the ϵ s by M , which is a function of a vector of parameters ρ with the ρ s distinct from the σ s. Denote the parameters in ρ by $\rho_1, \dots, \rho_K$ and let the covariance of Y_i be $\Omega = LML$, where L is the diagonal matrix of σ s. The matrix D_k denotes $\partial M / \partial \rho_k$ and $c_j = h_{11}(h^{-1}(\mu_j), \psi) / h_{10}(h^{-1}(\mu_j), \psi)$, with C being the $n \times n$ matrix with c_j as diagonal elements.

The orthogonalising differential equations for the μ s are very similar to those in section 2.1 and can be solved if $\partial \mu_j / \partial \psi = h_{01}(h^{-1}(\mu_j), \psi)$, which leads to a new set of orthogonal parameters $\lambda_j = h(\mu_j, \psi)$, $j = 1, \dots, n$. To orthogonalise for the σ s and ρ s requires some approximations and the evaluation of the trace of products of various matrices, analogous to the approximations in sections 2.1 and 2.2. The required terms for the differential equations are

$$\begin{aligned} \text{tr} \left(C \Omega^{-1} \frac{\partial \Omega}{\partial \sigma_j} \right) &= \frac{1}{\sigma_j} \{ c_j + (M C M^{-1})_{jj} \} \\ \text{tr} \left(C \Omega^{-1} \frac{\partial \Omega}{\partial \rho_k} \right) &= \text{tr}(C M^{-1} D_k) \\ \text{tr} \left(\Omega^{-1} \frac{\partial \Omega}{\partial \sigma_j} \Omega^{-1} \frac{\partial \Omega}{\partial \rho_k} \right) &= \frac{2}{\sigma_j} (D_k M^{-1})_{jj} \\ \text{tr} \left(\Omega^{-1} \frac{\partial \Omega}{\partial \sigma_j} \Omega^{-1} \frac{\partial \Omega}{\partial \sigma_j} \right) &= \frac{2}{\sigma_j^2} \{ 1 + (M^{-1})_{jj} \} \\ \text{tr} \left(\Omega^{-1} \frac{\partial \Omega}{\partial \sigma_j} \Omega^{-1} \frac{\partial \Omega}{\partial \sigma_k} \right) &= \frac{2}{\sigma_j \sigma_k} (M^{-1})_{jk} M_{kj} \end{aligned}$$

If the orthogonalising differential equations can be solved by $\partial \rho_k / \partial \psi = 0$, for all k , then this would demonstrate that the ρ s are orthogonal to ψ . We follow the same procedure as in section 2.2 and demonstrate that if we omit all the $\partial \rho_k / \partial \psi$ terms from the equations this leads to a consistent set of equations for the σ s which we can then solve. Following this procedure and substituting the terms for the traces above, the two sets of equations become

$$\sum_{s=1}^n \frac{1}{\sigma_s} (D_k M^{-1})_{ss} \frac{\partial \sigma_s}{\partial \psi} = \text{tr}(C M^{-1} D_k)$$

for $k = 1, \dots, K$ and

$$\frac{1}{\sigma_j^2} \{1 + (M^{-1})_{jj}\} \frac{\partial \sigma_j}{\partial \psi} + \sum_{s=1, s \neq j}^n \frac{1}{\sigma_j \sigma_s} (M^{-1})_{sj} M_{js} \frac{\partial \sigma_s}{\partial \psi} = \frac{1}{\sigma_j} \{c_j + (M C M^{-1})_{jj}\}$$

for $j = 1, \dots, n$.

All these equations are satisfied by setting $(1/\sigma_s) \partial \sigma_s / \partial \psi = c_s$ for every s . The solution of this is $\sigma_s = \eta_s h_{10}(h^{-1}(\mu_s), \psi)$, where $\eta = (\eta_1, \dots, \eta_K)$ are the new set of parameters orthogonal to ψ . Thus $\partial \rho_k / \partial \psi = 0$ is a solution which demonstrates that parameters that are unique to and determine the correlation structure for multivariate normal data are robust to the choice of transformations of the observations.

Example 3: A simple special case is the bivariate normal model assuming a common transformation for the response variables. One implication of the results from section 2.3 is that the correlation is stable to transformation, so that in practice one would expect a similar correlation in a scatterplot of the responses, irrespective of whether the axes in the scatterplot are on the original scale or some transformation of it.

2.4 Multivariate normal models with complex covariance structures

We now consider even more complex models of the form

$$h(Y_{ij}, \psi) = \mu + \epsilon_{ij}.$$

Here we combine the classes of models addressed in sections 2.2 and 2.3. For simplicity we assume the balanced case, and omit the technical details which are essentially messier versions of the previous developments.

Let ϵ_i be the n -vector of components of ϵ_{ij} and let $\Omega = \text{var}(\epsilon_i)$. We assume Ω takes the form $\Omega = L(\phi)M(\rho)L(\phi)$, where L is a diagonal matrix of standard deviations dependent on parameters $\phi_1, \dots, \phi_R$ and $M(\rho)$ is a correlation matrix determined by a vector of parameters $\rho = (\rho_1, \dots, \rho_K)$. We denote the diagonal elements of L by σ_j for $j = 1, \dots, n$.

Calculating or approximating all the required information matrix terms it can be shown that a solution to the approximate differential equations is $\partial\rho_k/\partial\psi = 0$, for all k , provided

$$\sum_{r=1}^R \frac{1}{\sigma_s} \frac{\partial\sigma_s}{\partial\phi_r} \frac{\partial\phi_r}{\partial\psi} = \frac{h_{11}}{h_{10}}$$

for every value of $s = 1, \dots, n$. If σ_s is taken to be of the form $\sum_{r=1}^R \phi_r f_r(s)$ then the equation is satisfied if

$$\frac{\partial\phi_r}{\partial\psi} = \phi_r \frac{h_{11}}{h_{10}}$$

for every r . So a solution is $\phi_r = \lambda_r h_{10} (h^{-1}(\mu), \psi)$, where λ_r is the new parameter, and the λ s and the original ρ s are orthogonal to ψ . It is also clear that the ratio of any two ϕ s are orthogonal to ψ . This is because the ratio of any two ϕ s is equal to the ratio of the corresponding λ s. A justification for $\sigma_s = \sum_{r=1}^R \phi_r f_r(s)$ is to think of ϕ_1 as an average standard deviation with $f_1 = 1$ and all the other terms represent departures from this average standard deviation so they might be small. Thus a Taylor series expansion of σ_s about $\phi = (\phi_1, 0, \dots, 0)$ would give the above form for σ_s .

An alternative and perhaps better way to set up σ_s is as $\sigma_s = \phi_1 \exp(\sum_{r=2}^R \phi_r f_r(s))$, where again ϕ_1 represents an average standard deviation. For this formulation the differential equations are solved if $\phi_1 = \lambda_1 h_{10} (h^{-1}(\mu), \psi)$ and $\partial\phi_r/\partial\psi = 0$ for $r = 2, \dots, R$; so all the other ϕ s except ϕ_1 are already orthogonal to ψ .

3 Regression models

In variance components models it is common to have fixed as well as random effects, thus it is of interest to investigate orthogonal parameters in the case when μ is replaced by $\sum_k X_{ik} \beta_k$. For simplicity we consider the model

$$h(Y_{ij}, \psi) = \sum_{k=1}^Q X_{ik} \beta_k + a_i + e_{ij},$$

where $a_i \sim N(0, \sigma_a^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$. We investigate whether the intraclass correlation ρ is still orthogonal to ψ . There is also interest in reparametrising the β s to a new set of parameters λ s which are orthogonal to ψ . Cox and Reid (1987) considered this situation for the model $Y_i^\psi = \sum_k X_{ik} \beta_k + e_i$, with $X_{i1} = 1$ and all

the other covariates X are centered. They showed that an orthogonal expression of this model is

$$Y_i^\psi \sim N(\lambda_1^\psi + \psi \lambda_1^\psi \sum_{k=2}^Q X_{ik} \lambda_k, \lambda_1^{2\psi-2} \psi^2 \lambda_0)$$

under certain conditions.

We mimic the approach of Cox and Reid to obtain an analogous expression of the model for the transformation $h(Y, \psi)$. In particular we assume that there is an intercept and that the variables are centered such that $X_{i1} = 1$ and $\sum_i X_{ij} = 0$ for $j = 2, \dots, Q$.

To simplify the notation we let μ_i denote $\sum_{k=1}^Q X_{ik} \beta_k$. Using similar approximations to those in section 2.1 we obtain

$$\begin{aligned} i_{\psi\theta} &\approx -\frac{1}{\theta} \sum_i n_i \frac{h_{11}(h^{-1}(\mu_i), \psi)}{h_{10}(h^{-1}(\mu_i), \psi)} \\ i_{\psi\rho} &\approx -\sum_i \frac{h_{11}(h^{-1}(\mu_i), \psi)}{h_{10}(h^{-1}(\mu_i), \psi)} \text{tr} \left(\Omega_i^{-1} \frac{\partial \Omega_i}{\partial \rho} \right) \\ i_{\theta\rho} &= \frac{1}{2\theta} \sum_i \text{tr} \left(\Omega_i^{-1} \frac{\partial \Omega_i}{\partial \rho} \right) \\ i_{\theta\theta} &= \frac{N}{2\theta^2} \\ i_{\rho\rho} &= \frac{1}{2} \sum_i \text{tr} \left(\Omega_i^{-1} \frac{\partial \Omega_i}{\partial \rho} \Omega_i^{-1} \frac{\partial \Omega_i}{\partial \rho} \right) \\ i_{\psi\beta_k} &\approx -\sum_i h_{01}(h^{-1}(\mu_i), \psi) 1_i' \Omega_i^{-1} 1_i' X_{ik} \\ i_{\beta_j\beta_k} &= -\sum_i X_{ij} 1_i' \Omega_i^{-1} 1_i' X_{ik} \end{aligned}$$

To proceed further we make the assumption that the effect of the covariates is small so that $f(\sum_{k=1}^Q X_{ik} \beta_k)$ can be approximated by $f(\beta_1) + \sum_{k=2}^Q X_{ik} \beta_k \partial f / \partial x(\beta_1)$ for a function f . The two differential equations involving θ and ρ are then

$$\frac{1}{2\theta} \sum_i \text{tr} \left(\Omega_i^{-1} \frac{\partial \Omega_i}{\partial \rho} \right) \frac{\partial \theta}{\partial \psi} + i_{\rho\rho} \frac{\partial \rho}{\partial \psi} = \sum_i \frac{h_{11}(h^{-1}(\mu_i), \psi)}{h_{10}(h^{-1}(\mu_i), \psi)} \text{tr} \left(\Omega_i^{-1} \frac{\partial \Omega_i}{\partial \rho} \right)$$

and

$$\frac{N}{2\theta^2} \frac{\partial \theta}{\partial \psi} + i_{\theta\rho} \frac{\partial \rho}{\partial \psi} = \frac{1}{\theta} \sum_i n_i \frac{h_{11}(h^{-1}(\mu_i), \psi)}{h_{10}(h^{-1}(\mu_i), \psi)}.$$

These are solved by ρ equal to a constant provided the following condition is satisfied:

$$\left\{ \sum_i \text{tr} \left(\Omega_i^{-1} \frac{\partial \Omega_i}{\partial \rho} \right) \right\} \left\{ \sum_i n_i \frac{h_{11}(h^{-1}(\mu_i), \psi)}{h_{10}(h^{-1}(\mu_i), \psi)} \right\} = N \sum_i \frac{h_{11}(h^{-1}(\mu_i), \psi)}{h_{10}(h^{-1}(\mu_i), \psi)} \text{tr} \left(\Omega_i^{-1} \frac{\partial \Omega_i}{\partial \rho} \right).$$

The condition is satisfied in the case of balanced data and is approximately satisfied when the effect of the covariates is small.

For the differential equations involving the β s we have for each k

$$\sum_{j=1}^Q \sum_i X_{ij} 1_i' \Omega_i^{-1} 1_i X_{ik} \frac{\partial \beta_j}{\partial \psi} = \sum_i h_{01}(h^{-1}(\mu_i), \psi) 1_i' \Omega_i^{-1} 1_i X_{ik}. \quad (4)$$

Following a Taylor series expansion the right hand side of (5) can be written as

$$\sum_i \left(h_{01}(h^{-1}(\beta_1), \psi) + \sum_{j=2}^Q \beta_j X_{ij} \frac{h_{11}(h^{-1}(\beta_1), \psi)}{h_{10}(h^{-1}(\beta_1), \psi)} \right) 1_i' \Omega_i^{-1} 1_i X_{ik}. \quad (5)$$

Equating terms on the left hand side of (4) with (5) we see that the equations are solved if

$$\frac{\partial \beta_1}{\partial \psi} = h_{01}(h^{-1}(\beta_1), \psi) \quad (6)$$

and

$$\frac{\partial \beta_j}{\partial \psi} = \beta_j \frac{h_{11}(h^{-1}(\beta_1), \psi)}{h_{10}(h^{-1}(\beta_1), \psi)}. \quad (7)$$

A solution to (6) and (7) in terms of the new parameters $(\lambda_1, \dots, \lambda_Q)$ is $\beta_1 = h(\lambda_1, \psi)$ and $\beta_j = \lambda_1 \lambda_j h_{10}(\lambda_1, \psi)$, $j = 2, \dots, Q$. This gives a reformulation of $\sum X_{ik} \beta_k$ as

$$h(\lambda_1, \psi) + \lambda_1 h_{10}(\lambda_1, \psi) \sum_{j=2}^Q X_{ij} \lambda_j,$$

where $\lambda_1 = h^{-1}(\beta_1)$ and $\lambda_j = \beta_j / \{h^{-1}(\beta_1) h_{01}(h^{-1}(\beta_1), \psi)\}$.

We note for j and k larger than 1, that $\beta_j / \beta_k = \lambda_j / \lambda_k$. This reproduces the result that the ratio of regression coefficients are interpretable quantities which can be robustly estimated without being strongly dependent on the specific assumptions in the model. This result has been obtained in a number of different ways for a variety of models, although we believe not models containing variance components, by previous authors (Brillinger, 1982; Solomon, 1984; Li and Duan, 1989; Skinner, 1989; Taylor, 1989, among others).

4 Parameters orthogonal to intraclass correlation

In this section we consider the simple one-way model

$$Y_{ij} = \mu + a_i + e_{ij}$$

without any transformation. The covariance structure of the observations are determined by two parameters. We have established in the earlier sections of this paper that the parameter $\rho = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2)$ has important orthogonality properties. The question we ask in this section is can we find a function of σ_a^2 and σ_e^2 which is orthogonal to ρ . We again use the Cox and Reid (1987) formulation and look for a reparametrisation from μ, θ, ρ to $\lambda_0, \lambda_1, \rho$, where $\theta = (\sigma_a^2 + \sigma_e^2)$, such that λ_0 and λ_1 are orthogonal to ρ . Clearly μ is already orthogonal to ρ so $\lambda_0 = \mu$. The required differential equation for θ is

$$i_{\theta\theta} \frac{\partial \theta}{\partial \rho} = -i_{\theta\rho}.$$

Using the forms of $i_{\theta\theta}$ and $i_{\theta\rho}$ given earlier in equations (1) and (2), we obtain

$$\frac{1}{\theta} \frac{\partial \theta}{\partial \rho} = -\frac{1}{N} \sum_i \frac{\rho(1 - n_i)n_i}{(1 - \rho)(1 - \rho + \rho n_i)}.$$

The expression on the right can be integrated exactly to give the solution

$$\theta = \lambda_1 \prod_{i=1}^I [(1 - \rho)^{-(n_i-1)} (n_i - 1)^{-(n_i/2-1)} \{1 + (n_i - 1)\rho\}^{-1}]^{1/N}$$

where λ_1 is an arbitrary constant. Rearranging this and dropping terms which only depend on n_i we obtain

$$\lambda_1 = \sigma_e^2 \left\{ \prod_{i=1}^I \left(\frac{\sigma_e^2 + n_i \sigma_a^2}{\sigma_e^2} \right)^{1/N} \right\}$$

as the parameter orthogonal to ρ .

It is interesting that this expression contains terms like $\sigma_e^2 + n_i \sigma_a^2$ which frequently occur in expressions for the maximum likelihood estimates of σ_e^2 and σ_a^2 and their variances. In the balanced case λ_1 reduces to $(\sigma_e^2)^{1-(1/n)} (\sigma_e^2 + n \sigma_a^2)^{1/n}$, which can be thought of as a form of the geometric mean of the two variances σ_e^2 and $\sigma_e^2 + n \sigma_a^2$. For the balanced one-way model Searle *et al* (1992) demonstrate that the variance-covariance matrix of the estimates of σ_e^2 and $\sigma_e^2 + n \sigma_a^2$ can be written as a diagonal matrix. The orthogonal parametrisation in the current paper provides an alternative to the one given by Searle *et al*.

5 Discussion

Our focus in this paper has been on variance component models and models which have specific covariance structures. We have considered broad families of models which cover many practical situations and applications to longitudinal data and repeated measures data generally. In many of the models we investigated we considered the correlation structure and the variance structure as separate components determining the overall covariance structure of the observations, with different sets of parameters determining the two components.

The correlation parameters may or may not be of primary interest in any given application. It is useful to know however, which features of the data and aspects of the model are robust to transformation, since the choice of transformation is often made for convenience or acceptability in a particular field of application. In view of the robustness results we have obtained concerning the correlation parameters, we believe this is a useful way to formulate models for repeated measures data.

6 Acknowledgements

This work was supported by grants from the Australian Research Council and the United States National Institutes of Health.

References

- Brillinger D.R. (1982). A generalised linear model with ‘Gaussian’ regressor variables. In *A Festschrift for Erich Lehmann* (eds P.J.Bickel, K.A.Doksum and J.L.Hodges). San Francisco: Wadsworth.
- Cox D.R. and Reid N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J.R. Statist. Soc. B* **49**, 1-39.
- Diggle P.J. (1988). An approach to the analysis of repeated measurements. *Biometrics* **44**, 959-971.
- Li, K.-C. and Duan, N. (1989). Regression analysis under link violation. *Ann. Stat.* **17**, 1009-1052.
- Searle, S.R., Casella, G. and McCulloch, C.E. (1992). *Variance Components*. Wiley: New York.

- Skinner C.J. (1989). GLMs and coefficients ratios. In *Statistical Modelling* (eds A. Decarli, B.J.Francis, R.Gilchrist and S.U.H.Seeber). New York, Springer.
- Solomon P.J. (1984). Effect of misspecification of regression models in the analysis of survival data. *Biometrika* **71**, 291-298; amendment (1986) **73**, 245.
- Solomon P.J. (1985). Transformations for components of variance and covariance. *Biometrika* **72**, 233-239.
- Taylor J.M.G. (1989). A note on the cost of estimating the ratio of regression parameters after fitting a power transformation. *J. Stat. Plan. Infer.* **21**, 223-230.
- Taylor J.M.G. , Cumberland W.G. and Sy J.P. (1994). A stochastic model for analysis of longitudinal AIDS data. *J. Amer. Statist. Assoc.* **89**, 727-736.
- Taylor J.M.G., Cumberland W.G. and Meng X. (1996). Components of variance models with transformations. *Australian J. Stat.* **38**, 183-191.