Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 98 (2007) 1412 – 1416
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Fisher information for generalised linear mixed models

M.P. Wand

Department of Statistics, School of Mathematics and Statistics, University of New South Wales, Sydney 2052, Australia

Received 25 October 2005
Available online 11 January 2007

Abstract

The Fisher information for the canonical link exponential family generalised linear mixed model is derived. The contribution from the fixed effects parameters is shown to have a particularly simple form.
© 2007 Elsevier Inc. All rights reserved.

AMS 2000 subject classification: 62J12; 62H12; 62F12

Keywords: Longitudinal data analysis; Semiparametric regression; Vector differential calculus; Variance components

1. Introduction

Linear mixed models are an extension of linear models for which covariance structure is based on random effects and their covariance parameters. A further extension is generalised linear mixed models, which specifically cater for non-normal response variables. The use of such models has increased dramatically in the past decade. For example, they are now the main vehicle for analysis of longitudinal data (e.g. [1,2]). Their use in semiparametric regression is advocated in some recent literature (e.g. [6]).

The popularity of linear mixed models has been accompanied by vigorous research on analytic results and computational methods. McCulloch and Searle [5] provides a comprehensive summary of the developments that took place in the 20th Century. In the case of the normal response linear mixed model with variance component structure they give an elegant and succinct expression for the Fisher information matrix of the model parameters. The result is reproduced in Section 2. It is revealing in that asymptotic independence of fixed effects and variance components is apparent and useful in that asymptotic sampling variances may be obtained.

In this note, we derive an explicit expression for the Fisher information for generalised linear mixed models for exponential family response variables. Potentially, this analytic result can be

E-mail address: wand@maths.unsw.edu.au.

0047-259X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2007.01.001

used in the same way that Fisher information matrices are used for standard error approximations in generalised linear models and normal linear mixed models.

In Section 2 we briefly review the Fisher information results for normal linear mixed models. New results for generalised mixed models are presented in Section 3 and their ramifications are discussed.

2. Normal linear mixed models

The normal linear mixed model is given by

$$\mathbf{y}|\mathbf{u} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{R}).$$

Here $\boldsymbol{\beta}$ is the vector of fixed effects, with corresponding design matrix $\mathbf{X}$, and $\mathbf{u}$ is the vector of random effects, with corresponding design matrix $\mathbf{Z}$. The random effects vector $\mathbf{u}$ has density $f(\mathbf{u})$ and is such that $\mathsf{E}(\mathbf{u}) = \mathbf{0}$. Most commonly, the random effects distribution is Gaussian: $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$ for some covariance matrix $\mathbf{G}$. More details are given in Chapter 6 of McCulloch and Searle [5].

Regardless of the random effects distribution, the Fisher information matrix of $\boldsymbol{\beta}$ is $\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}$ where $\mathbf{V} = \text{cov}(\mathbf{y}) = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$ is the covariance matrix of $\mathbf{y}$. Now consider the special case of variance component covariance structure

$$\mathbf{Z}\mathbf{u} = \sum_{i=1}^{r} \mathbf{Z}_i\mathbf{u}_i, \quad \mathbf{G}_{\boldsymbol{\sigma}^2} = \underset{1\leqslant i\leqslant r}{\text{blockdiag}}(\sigma_i^2\mathbf{I}_{q_i}) \quad \text{and} \quad \mathbf{R} = \sigma_0^2\mathbf{I}_n, \tag{1}$$

where $\mathbf{I}_d$ denotes the $d \times d$ identity matrix, q_i is the length of $\mathbf{u}_i$ and n is the length of $\mathbf{y}$. The full parameter vector is $(\boldsymbol{\beta}, \boldsymbol{\sigma}^2)$ where $\boldsymbol{\sigma}^2 \equiv (\sigma_0^2, \ldots, \sigma_r^2)$. Let its Fisher information matrix be written as

$$I(\boldsymbol{\beta}, \boldsymbol{\sigma}^2) \equiv -\mathsf{E}\{\mathsf{H}_{\boldsymbol{\beta},\boldsymbol{\sigma}^2} \log f(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\sigma}^2)\} = \begin{bmatrix} I_{\boldsymbol{\beta}\boldsymbol{\beta}} & I_{\boldsymbol{\beta}\boldsymbol{\sigma}^2} \\ I_{\boldsymbol{\beta}\boldsymbol{\sigma}^2}^T & I_{\boldsymbol{\sigma}^2\boldsymbol{\sigma}^2} \end{bmatrix}, \tag{2}$$

where $f(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\sigma}^2)$ denotes the density of $\mathbf{y}$ and $\mathsf{H}_{\boldsymbol{\beta},\boldsymbol{\sigma}^2}$ denotes the Hessian matrix with respect to $(\boldsymbol{\beta}, \boldsymbol{\sigma}^2)$. From above $I_{\boldsymbol{\beta}\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}$. Also, regardless of $f(\mathbf{u}; \boldsymbol{\sigma}^2)$, $I_{\boldsymbol{\beta}\boldsymbol{\sigma}^2} = \mathbf{0}$ which indicates an asymptotic independence between the maximum likelihood estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}^2$. In the special case of $\mathbf{u}$ being normally distributed the $I_{\boldsymbol{\sigma}^2\boldsymbol{\sigma}^2}$ block has an explicit expression that results in

$$I(\boldsymbol{\beta}, \boldsymbol{\sigma}^2) = \begin{bmatrix} \mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}[\text{tr}\{\mathbf{Z}_i^T\mathbf{V}^{-1}\mathbf{Z}_j(\mathbf{Z}_i^T\mathbf{V}^{-1}\mathbf{Z}_j)^T\}]_{0\leqslant i,j,\leqslant r} \end{bmatrix}, \tag{3}$$

where $\mathbf{Z}_0 \equiv \mathbf{I}$. This is Eq. (6.62) of McCulloch and Searle [5]. In the ensuing discussion, these authors describe asymptotic sampling variances based on $I(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\sigma}}^2)^{-1}$ where $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\sigma}}^2$ denote the maximum likelihood estimators.

3. Generalised linear mixed models

As mentioned previously, generalised linear mixed models extend linear mixed models to non-normal response situations. Alternatively, generalised linear mixed models extend generalised linear models by allowing for the incorporation of random effects. A useful class of generalised linear models is that corresponding to the one-parameter exponential family with canonical link as treated in McCullagh and Nelder [4]. If $\mathbf{y}$ is a random vector and $\mathbf{X}$ is a general design matrix with corresponding parameter vector $\boldsymbol{\beta}$ then the model may be expressed in terms of the density function of $\mathbf{y}$ as

$$f(\mathbf{y}) = \exp\{\mathbf{y}^T(\mathbf{X}\boldsymbol{\beta}) - \mathbf{1}^T b(\mathbf{X}\boldsymbol{\beta}) + \mathbf{1}^T c(\mathbf{y})\}, \tag{4}$$

where, for example, $b(x) = \exp(x)$ for the Poisson model and $b(x) = \log(1+e^x)$ for the Bernoulli model. For a general vector $\mathbf{v} = [v_1, \ldots, v_d]^T$, $b(\mathbf{v})$ denotes the vector $[b(v_1), \ldots, b(v_d)]^T$. The mixed model extension of (4) is

$$f(\mathbf{y}|\mathbf{u}) = \exp\{\mathbf{y}^T(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \mathbf{1}^T b(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) + \mathbf{1}^T c(\mathbf{y})\},$$

where $\mathbf{u}$ is the vector of random effects and $\mathbf{Z}$ is a corresponding design matrix. As in Section 2 let $f(\mathbf{u})$ denote the density of $\mathbf{u}$ and retain the convention that $\mathsf{E}(\mathbf{u}) = \mathbf{0}$.

Result 1. *Regardless of the random effects density* $f(\mathbf{u})$

$$I_{\boldsymbol{\beta}\boldsymbol{\beta}} = \mathbf{X}^T \operatorname{cov}[\mathbf{y} - \mathsf{E}\{\mathsf{E}(\mathbf{y}|\mathbf{u})|\mathbf{y}\}]\mathbf{X}. \tag{5}$$

Remark 1. While (5) is exact, calculation of the (conditional) moments in this expression can be quite difficult. Generally, they involve intractable multivariate integrals although, for some specific cases (e.g. random intercept models), some integrals simplify to low-dimensional forms. Algorithms and approximations for dealing with such integrals have been the subject of a great deal of research in the past several years [5, Chapter 10]. Such research is ongoing.

Now consider the special case of variance component structure where $\mathbf{G} = \mathbf{G}_{\boldsymbol{\sigma}^2}$ is given by (1). Then the full parameter vector is $(\boldsymbol{\beta}, \boldsymbol{\sigma}^2)$ where $\boldsymbol{\sigma}^2 \equiv (\sigma_1^2, \ldots, \sigma_r^2)$. The full Fisher information matrix takes the form of (2) where $I_{\boldsymbol{\beta}\boldsymbol{\beta}}$ is given by Result 1. The contribution from the variance component vector has the following simplified expression:

$$I_{\boldsymbol{\sigma}^2\boldsymbol{\sigma}^2} = \mathsf{E}\left[\mathsf{E}\left\{\frac{\mathsf{D}_{\boldsymbol{\sigma}^2} f(\mathbf{u};\boldsymbol{\sigma}^2)}{f(\mathbf{u};\boldsymbol{\sigma}^2)}\Big|\mathbf{y}\right\}^T \mathsf{E}\left\{\frac{\mathsf{D}_{\boldsymbol{\sigma}^2} f(\mathbf{u};\boldsymbol{\sigma}^2)}{f(\mathbf{u};\boldsymbol{\sigma}^2)}\Big|\mathbf{y}\right\}\right], \tag{6}$$

where $\mathsf{D}_{\boldsymbol{\sigma}^2} f(\mathbf{u};\boldsymbol{\sigma}^2)$ denotes the derivative vector of $f(\mathbf{u};\boldsymbol{\sigma}^2)$ with respect to $\boldsymbol{\sigma}^2$ (see Appendix).

For general $f(\mathbf{u};\boldsymbol{\sigma}^2)$ no further simplification of (6) is possible. However, in the case of $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G}_{\boldsymbol{\sigma}^2})$ an explicit expression can be obtained. It depends on previously defined expressions and the $(\sum_{i=1}^r q_i)^2 \times r$ binary matrix $\mathcal{I} = [\mathcal{I}_1|\ldots|\mathcal{I}_r]$ where

$$\mathcal{I}_i = \operatorname{vec}\{\operatorname{blockdiag}(\mathbf{O}_{q_1}|\ldots|\mathbf{O}_{q_{i-1}}|\mathbf{I}_{q_i}|\mathbf{O}_{q_{i+1}}|\ldots|\mathbf{O}_{q_r})\}$$

and $\mathbf{O}_d$ denotes the $d \times d$ matrix of zeroes.

Result 2. *If* $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G}_{\boldsymbol{\sigma}^2})$, *where* $\mathbf{G}_{\boldsymbol{\sigma}^2}$ *is given by* (1), *then*

$$I_{\boldsymbol{\sigma}^2\boldsymbol{\sigma}^2} = \tfrac{1}{4}\mathcal{I}^T(\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1} \otimes \mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}) \operatorname{cov}[\operatorname{vec}\{\mathsf{E}(\mathbf{u}\mathbf{u}^T|\mathbf{y})\}](\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1} \otimes \mathbf{G}_{\boldsymbol{\sigma}^2}^{-1})\mathcal{I}.$$

The off-diagonal block is

$$I_{\boldsymbol{\beta}\boldsymbol{\sigma}^2} = \mathbf{X}^T \mathsf{E}\left([\mathbf{y} - \mathsf{E}\{\mathsf{E}(\mathbf{y}|\mathbf{u})|\mathbf{y}\}]\mathsf{E}\left\{\frac{\mathsf{D}_{\boldsymbol{\sigma}^2} f(\mathbf{u};\boldsymbol{\sigma}^2)}{f(\mathbf{u};\boldsymbol{\sigma}^2)}\Big|\mathbf{y}\right\}\right). \tag{7}$$

In general, this matrix is non-zero. This indicates that the maximum likelihood estimators of fixed effects and variance components are not asymptotically independent in the general exponential family setting. In the special case of normal $\mathbf{u}$ we have

Result 3. *If* $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G}_{\boldsymbol{\sigma}^2})$, *where* $\mathbf{G}_{\boldsymbol{\sigma}^2}$ *is given by* (1), *then*

$$I_{\boldsymbol{\beta}\boldsymbol{\sigma}^2} = \tfrac{1}{2}\mathbf{X}^T \mathsf{E}\left([\mathbf{y} - \mathsf{E}\{\mathsf{E}(\mathbf{y}|\mathbf{u})|\mathbf{y}\}]\,\mathrm{vec}[\{\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\mathsf{E}(\mathbf{u}\mathbf{u}^T|\mathbf{y}) - \mathbf{I}\}\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}]^T\right)\mathcal{I}.$$

Remark 2. Approximate standard errors for the maximum likelihood estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\sigma}}^2$ can be obtained from the diagonal entries of $I(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\sigma}}^2)^{-1}$. However, as pointed out in Remark 1, implementation is often hindered by intractable multivariate integrals. Additionally, dependence among the entries of $\mathbf{y}$ induced by $\mathbf{u}$ means that central limit theorems of the type: $I(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\sigma}}^2)^{-1}\{(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\sigma}}^2) - (\boldsymbol{\beta}, \boldsymbol{\sigma}^2)\}$ converges in distribution to a $N(\mathbf{0}, \mathbf{I})$ random vector, have not been established in general and, hence, interpretation of standard errors is cloudy. Nevertheless, there are many special cases, such as m-dependence when the data are from a longitudinal study, for which central limit theorems can be established.

Acknowledgments

This article has benefited considerably from the comments of an associate editor.

Appendix. Derivations

Differential calculus preliminaries: Let f be a scalar-valued function with argument $\mathbf{x} \in \mathbb{R}^d$. The *derivative vector* of f, $\mathsf{D}f(\mathbf{x})$, is the $1 \times d$ vector whose ith entry is $\partial f(\mathbf{x})/\partial x_i$. The *Hessian matrix* of f is the $d \times d$ matrix $\mathsf{H}f(\mathbf{x}) = \mathsf{D}\{\mathsf{D}f(\mathbf{x})^T\}$. Magnus and Neudecker [3] and Wand [7] describe techniques for finding derivative vectors and Hessian matrices.

For a general smooth log-likelihood $\ell(\boldsymbol{\theta})$ with parameter vector $\boldsymbol{\theta}$ the Fisher information is

$$I(\boldsymbol{\theta}) \equiv \mathsf{E}\{-\mathsf{H}\ell(\boldsymbol{\theta})\} = \mathsf{E}\{\mathsf{D}\ell(\boldsymbol{\theta})^T\mathsf{D}\ell(\boldsymbol{\theta})\}.$$

Either expression for $I(\boldsymbol{\theta})$ could be used. For generalised linear mixed models the second one leads to more direct derivation of the Fisher information.

Derivation of Result 1: The log-likelihood is

$$\ell(\boldsymbol{\beta}, \boldsymbol{\sigma}^2) = \log \int f(\mathbf{y}, \mathbf{u}; \boldsymbol{\beta}, \boldsymbol{\sigma}^2)\, d\mathbf{u},$$

where

$$f(\mathbf{y}, \mathbf{u}; \boldsymbol{\beta}, \boldsymbol{\sigma}^2) = \exp\{\mathbf{y}^T(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \mathbf{1}^T b(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) + \mathbf{1}^T c(\mathbf{y})\} f(\mathbf{u}; \boldsymbol{\sigma}^2).$$

The first differential is

$$\begin{aligned} d_{\boldsymbol{\beta}}\ell(\boldsymbol{\beta}, \boldsymbol{\sigma}^2) &= \frac{\int f(\mathbf{y}, \mathbf{u}; \boldsymbol{\beta}, \boldsymbol{\sigma}^2)[\{\mathbf{y} - b'(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})\}^T \mathbf{X}\, d\boldsymbol{\beta}]\, d\mathbf{u}}{f(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\sigma}^2)} \\ &= [\mathbf{y} - \mathsf{E}\{\mathsf{E}(\mathbf{y}|\mathbf{u})|\mathbf{y}\}]^T \mathbf{X}\, d\boldsymbol{\beta} \end{aligned}$$

and so

$$\mathsf{D}_{\boldsymbol{\beta}}\ell(\boldsymbol{\beta}, \boldsymbol{\sigma}^2) = [\mathbf{y} - \mathsf{E}\{\mathsf{E}(\mathbf{y}|\mathbf{u})|\mathbf{y}\}]^T \mathbf{X}.$$

Therefore,

$$I_{\boldsymbol{\beta\beta}} = \mathsf{E}\{\mathsf{D}_{\boldsymbol{\beta}}\ell(\boldsymbol{\beta},\boldsymbol{\sigma}^2)^T\mathsf{D}_{\boldsymbol{\beta}}\ell(\boldsymbol{\beta},\boldsymbol{\sigma}^2)\} = \mathbf{X}^T\mathsf{E}\left([\mathbf{y}-\mathsf{E}\{\mathsf{E}(\mathbf{y}|\mathbf{u})|\mathbf{y}\}]^{\otimes 2}\right)\mathbf{X},$$

where, for a general vector $\mathbf{v}$, $\mathbf{v}^{\otimes 2} \equiv \mathbf{v}\mathbf{v}^T$. Result 1 then follows from the fact that $\mathsf{E}[\mathsf{E}\{\mathsf{E}(\mathbf{y}|\mathbf{u})|\mathbf{y}\}] = \mathsf{E}(\mathbf{y})$.

Derivation of Result 2: We have $f(\mathbf{u};\boldsymbol{\sigma}^2) = (2\pi)^{-q/2}|\mathbf{G}_{\boldsymbol{\sigma}^2}|^{-1/2}\exp(-\frac{1}{2}\mathbf{u}^T\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\mathbf{u})$ and

$$\begin{aligned} d_{\boldsymbol{\sigma}^2} f(\mathbf{u};\boldsymbol{\sigma}^2) &= (2\pi)^{-q/2}(-\tfrac{1}{2})|\mathbf{G}_{\boldsymbol{\sigma}^2}|^{-3/2}|\mathbf{G}_{\boldsymbol{\sigma}^2}|\,\mathrm{tr}(\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}d_{\boldsymbol{\sigma}^2}\mathbf{G}_{\boldsymbol{\sigma}^2})\exp(-\tfrac{1}{2}\mathbf{u}^T\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\mathbf{u}) \\ &\quad + (2\pi)^{-q/2}|\mathbf{G}_{\boldsymbol{\sigma}^2}|^{-1/2}\exp(-\tfrac{1}{2}\mathbf{u}^T\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\mathbf{u})(-\tfrac{1}{2})\mathbf{u}^T\{-\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}(d_{\boldsymbol{\sigma}^2}\mathbf{G}_{\boldsymbol{\sigma}^2})\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\}\mathbf{u} \\ &= \tfrac{1}{2}f(\mathbf{u};\boldsymbol{\sigma}^2)\,\mathrm{tr}\{(\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\mathbf{u}\mathbf{u}^T-\mathbf{I})\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}d_{\boldsymbol{\sigma}^2}\mathbf{G}_{\boldsymbol{\sigma}^2}\} \\ &= \tfrac{1}{2}f(\mathbf{u};\boldsymbol{\sigma}^2)\,\mathrm{vec}\{(\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\mathbf{u}\mathbf{u}^T-\mathbf{I})\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\}^T d_{\boldsymbol{\sigma}^2}\,\mathrm{vec}(\mathbf{G}_{\boldsymbol{\sigma}^2}) \\ &= \tfrac{1}{2}f(\mathbf{u};\boldsymbol{\sigma}^2)\,\mathrm{vec}\{(\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\mathbf{u}\mathbf{u}^T-\mathbf{I})\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\}^T\mathcal{I}\,d\boldsymbol{\sigma}^2 \end{aligned}$$

courtesy of the identity $\mathrm{tr}(\mathbf{A}^T\mathbf{B}) = \mathrm{vec}(\mathbf{A})^T\,\mathrm{vec}(\mathbf{B})$ and the result $\mathrm{vec}(\mathbf{G}_{\boldsymbol{\sigma}^2}) = \mathcal{I}\boldsymbol{\sigma}^2$. It follows immediately that, for $\mathbf{u}\sim N(\mathbf{0},\mathbf{G}_{\boldsymbol{\sigma}^2})$,

$$\frac{\mathsf{D}_{\boldsymbol{\sigma}^2}f(\mathbf{u};\boldsymbol{\sigma}^2)}{f(\mathbf{u};\boldsymbol{\sigma}^2)} = \frac{1}{2}\,\mathrm{vec}\{(\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\mathbf{u}\mathbf{u}^T-\mathbf{I})\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\}^T\mathcal{I} \tag{8}$$

and, because of (6),

$$I_{\boldsymbol{\sigma}^2\boldsymbol{\sigma}^2} = \tfrac{1}{4}\mathcal{I}^T\mathsf{E}\left(\mathrm{vec}[\{(\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\mathsf{E}(\mathbf{u}\mathbf{u}^T|\mathbf{y})-\mathbf{I})\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\}]^{\otimes 2}\right)\mathcal{I}. \tag{9}$$

Noting that $\mathsf{E}\{\mathsf{E}(\mathbf{u}\mathbf{u}^T|\mathbf{y})\}=\mathsf{E}(\mathbf{u}\mathbf{u}^T)=\mathbf{G}_{\boldsymbol{\sigma}^2}$ and the matrix identity $\mathrm{vec}(\mathbf{ABC})=(\mathbf{C}^T\otimes\mathbf{B})\,\mathrm{vec}(\mathbf{B})$, the expectation in (9) becomes

$$\begin{aligned} \mathrm{cov}[\mathrm{vec}\{\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\mathsf{E}(\mathbf{u}\mathbf{u}^T|\mathbf{y})\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\}] &= \mathrm{cov}[(\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\otimes\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1})\,\mathrm{vec}\{\mathsf{E}(\mathbf{u}\mathbf{u}^T|\mathbf{y})\}] \\ &= (\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\otimes\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1})\,\mathrm{cov}[\mathrm{vec}\{\mathsf{E}(\mathbf{u}\mathbf{u}^T|\mathbf{y})\}](\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}\otimes\mathbf{G}_{\boldsymbol{\sigma}^2}^{-1}). \end{aligned}$$

Result 2 then follows immediately from this expression and (9).

Derivation of Result 3: This result follows immediately from (7) and (8).

References

[1] P. Diggle, P. Heagerty, K.-L. Liang, S. Zeger, Analysis of Longitudinal Data, second ed., Oxford University Press, Oxford, 2002.

[2] G.M. Fitzmaurice, N.M. Laird, J.H. Ware, Applied Longitudinal Analysis, Wiley, New York, 2004.

[3] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, second ed., Wiley, Chichester, 1999.

[4] P. McCullagh, J.A. Nelder, Generalized Linear Models, second ed., Chapman & Hall, London, 1990.

[5] C.E. McCulloch, S.R. Searle, Generalized, Linear, and Mixed Models, Wiley, New York, 2000.

[6] D. Ruppert, M.P. Wand, R.J. Carroll, Semiparametric Regression, Cambridge University Press, New York, 2003.

[7] M.P. Wand, Vector differential calculus in statistics, Amer. Statist. 56 (2002) 55–62.