Ann. Inst. Statist. Math.
Vol. 52, No. 4, 737–752 (2000)

BAYESIAN INFLUENCE ASSESSMENT IN THE GROWTH CURVE MODEL WITH UNSTRUCTURED COVARIANCE

JIAN-XIN PAN[1]* AND WING-KAM FUNG[2]**

[1] *Statistics Department, IACR-Rothamsted, Harpenden, Herts AL5 2JQ, U.K.*
[2] *Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China*

(Received September 17, 1998; revised May 28, 1999)

Abstract. From a Bayesian point of view, in this paper we discuss the influence of a subset of observations on the posterior distributions of parameters in a growth curve model with unstructured covariance. The measure used to assess the influence is based on a Bayesian entropy, namely Kullback-Leibler divergence (KLD). Several new properties of the Bayesian entropy are studied, and analytically closed forms of the KLD measurement both for the matrix-variate normal distribution and the Wishart distribution are established. In the growth curve model, the KLD measurements for all combinations of the parameters are also studied. For illustration, a practical data set is analyzed using the proposed approach, which shows that the diagnostics measurements are useful in practice.

Key words and phrases: Bayesian analysis, case-deletion method, growth curve model, Kullback-Leibler divergence, statistical diagnostics.

1. Introduction

In general, statistical diagnostics for a certain model are studied under two different frameworks, i.e., likelihood and Bayesian methodologies. Within each framework, diagnostic approaches are further classified into two categories, i.e., global influence or case-deletion and local influence. The former approach is concerned with the change of certain statistical quantities, such as the maximum likelihood estimators (MLEs) of parameters and fitted values of the model, when a subset of observations is deleted from the model. In such a manner, the influence of a subset of observations on the model fittings can be assessed in terms of a certain measure, for example, the Cook's distance (Cook (1979)). On the other hand, the local influence approach proposed by Cook (1986) aims to quantify the effect of a local departure on model assumptions. This method may also reveal the so-called masking and swamping phenomena as well (see, e.g., Cook (1986)). Within the likelihood framework, Cook (1979, 1986) comprehensively studied the identification of multiple outliers and influential observations in an ordinary regression model using the global and local influence techniques. The relationship between those two techniques was also studied by Cook (1986).

From a Bayesian point of view, statistical diagnostics has received much attention recently because it may provide much more information about influential observations,

* Now at Centre for Medical Statistics, Department of Mathematics, Keele University, Keele, Staffordshire, ST5 5BG, U.K. Pan's research was supported by a grant from Agriculture and Fisheries Department, Scottish Office (Project Number: 141627). Part of this work was conducted when he was visiting the Department of Statistics, The University of Hong Kong.

** Fung's work was partly funded by a RGC grant (HKU 7134/98H).

see, for example, Pan *et al.* (1996). In the literature, Box and Tiao (1968) considered the posterior probability of a random event that a subset of observations may be an outlier or influential set. These diagnostics approaches were further improved by other authors including Pettit and Smith (1985), Geisser (1987), and Chaloner and Brant (1988). Johnson and Geisser (1983, 1985) suggested using the Kullback-Leibler divergence (KLD) between certain predictive or posterior distributions to measure the influence of a subset of observations. In ordinary regression models, relevant studies were conducted by Guttman and Peña (1988, 1993) and Ali (1990). In addition, Carlin and Polson (1991) justified taking the KLD measurement as a utility function, and showed the way to compute diagnostics using the Gibbs sampling method for more complicated models.

The model considered in this paper is the *growth curve model* (GCM):

$$Y_{p\times n} = X_{p\times m}B_{m\times r}Z_{r\times n} + \epsilon_{p\times n}, \tag{1.1}$$

where X and Z are known design matrices of rank $m(< p)$ and $r(< n)$ respectively, and B is the regression coefficient matrix. The columns of the error matrix ϵ are independent p-variate normal with a mean vector $\mathbf{0}$ and a common covariance matrix Σ, i.e., the conditional distribution $\epsilon \mid (B, \Sigma) \sim N_{p,n}(\mathbf{0}, \Sigma, I_n)$, where Σ is an arbitrary positive definite matrix, known as the *unstructured covariance* (UC). The model (1.1) is useful especially in longitudinal studies and growth problems over short periods, and has been applied extensively in economics and medical research. The GCM was first proposed by Potthoff and Roy (1964) and then considered subsequently by many other authors, including Rao (1965, 1966), Khatri (1966), Geisser (1970), Lee (1988, 1991) and von Rosen (1989, 1991) among others. In the GCM with UC, the MLEs of the regression coefficient B and the dispersion component Σ are of the forms,

$$\hat{B} = (X^{\tau}S^{-1}X)^{-1}X^{\tau}S^{-1}YZ^{\tau}(ZZ^{\tau})^{-1} \quad \text{and} \quad \hat{\Sigma} = \frac{1}{n}(S + Q_S Y P_{Z^{\tau}} Y^{\tau} Q_S^{\tau}), \tag{1.2}$$

respectively, where $Q_S = SQ(Q^{\tau}SQ)^{-1}Q^{\tau}$, $S = Y(I_n - P_{Z^{\tau}})Y^{\tau}$ and $Q \in \mathcal{Q}$ in which $\mathcal{Q}$ is a set of matrices defined by $\mathcal{Q} = \{Q \mid Q : p \times (p-m),\ \text{rank}(Q) = p - m \text{ and } X^{\tau}Q = \mathbf{0}\}$ (see, e.g., Khatri (1966); von Rosen (1989)). The notation $P_A = A(A^{\tau}A)^{-1}A^{\tau}$ denotes the projection matrix of A provided that $A^{\tau}A$ is nonsingular throughout this paper. The $p \times p$ symmetric matrix S is positive definite with probability one as far as $n > p + r$ (Okamato (1973)).

For the GCM (1.1) with UC, within the likelihood framework, Pan and Fang (1995, 1996) established criteria for multiple outlier detection and influential observation identification in terms of a mean shift regression model and case-deletion techniques. Based on Taylor expansion of a perturbed model, von Rosen (1995) discussed the identification of influential observations. Working with a *spherical covariance structure*, i.e., $\Sigma = \sigma^2 I_p$, Liski (1991) considered the detection of influential observations from a likelihood point of view. Within the Bayesian framework with a non-informative prior

$$p(B, \Sigma^{-1}) \propto \{\det(\Sigma^{-1})\}^{-(p+1)/2} \tag{1.3}$$

Geisser (1970) obtained the posterior density of the regression coefficient B. In fact it is a matrix-variate student t distribution (see e.g. Dickey (1967)). Furthermore, Pan *et al.* (1998) showed that the posterior distribution of the dispersion component Σ^{-1} is a mixture of two Wishart distributions.

From the Bayesian point of view, in this paper we consider the influence of a subset of observations on the posterior distributions of parameters in the GCM with UC. The measure used to assess the influence is based on the Bayesian entropy, i.e., the KLD measurement. In Section 2, several new properties of this Bayesian entropy are outlined, and then analytically closed forms of the KLD measurement for the matrix-variate normal distribution and the Wishart distribution are established. With the non-informative prior (1.3), in Section 3 the posterior distributions of parameters in the model are given for later uses. Specifically, in this section we present the KLD measurement for a single parameter, i.e., $\boldsymbol{B}$ or $\boldsymbol{\Sigma}$, and for their combination, i.e., $(\boldsymbol{B}, \boldsymbol{\Sigma})$ as well. For illustration, a practical data set is analyzed in Section 4. Finally, a discussion on the approach is given in Section 5. All the technical details on proofs of the theorems are postponed until the Appendix.

2. Bayesian influence measurements

In this section, we first introduce the definition of Kullback-Leibler divergence (KLD) and then give several new properties of this Bayesian entropy. Furthermore, analytically closed forms of the KLD measurements for the matrix-variate normal distribution and the Wishart distribution are provided, which play important roles in the detection of influential observations in the GCM with UC. To save space, we do not intend to give the technical details of these results here. Readers who are interested can get the details by contacting the authors.

2.1 *Kullback-Leibler divergence*

Suppose $\boldsymbol{\theta}$ is a parameter vector in a statistical model $M(\boldsymbol{\theta})$. Without loss of generality, the sample matrix $\boldsymbol{Y}$ can be partitioned into $\boldsymbol{Y} = (\boldsymbol{Y}_{(I)} : \boldsymbol{Y}_I)$, where I is a subset of the index set $\{1, 2, \ldots, n\}$ and n is the sample size. With a prior density of $\boldsymbol{\theta}$, say $p(\boldsymbol{\theta})$, the posterior densities of the parameter $\boldsymbol{\theta}$ based on the full observations $\boldsymbol{Y}$ and the partial observations $\boldsymbol{Y}_{(I)}$ are available and denoted by $p(\boldsymbol{\theta} \mid \boldsymbol{Y})$ and $p(\boldsymbol{\theta} \mid \boldsymbol{Y}_{(I)})$, respectively. How should we measure the difference between those two posterior densities $p(\boldsymbol{\theta} \mid \boldsymbol{Y})$ and $p(\boldsymbol{\theta} \mid \boldsymbol{Y}_{(I)})$? One of the commonly used measures is the *Kullback-Leibler divergence* (KLD):

$$K_I(\boldsymbol{\theta}) \equiv K[p(\boldsymbol{\theta} \mid \boldsymbol{Y}_{(I)}), p(\boldsymbol{\theta} \mid \boldsymbol{Y})] = E_{p(\boldsymbol{\theta} \mid \boldsymbol{Y}_{(I)})} \left\{ \log \frac{p(\boldsymbol{\theta} \mid \boldsymbol{Y}_{(I)})}{p(\boldsymbol{\theta} \mid \boldsymbol{Y})} \right\}, \tag{2.1}$$

which measures the effect of $\boldsymbol{Y}_I$ on the posterior distribution of the parameter $\boldsymbol{\theta}$. Since the KLD measurement (2.1) is in general asymmetric, in other words, the measurement $K[p(\boldsymbol{\theta} \mid \boldsymbol{Y}_{(I)}), p(\boldsymbol{\theta} \mid \boldsymbol{Y})]$ does not necessarily equal to $K[p(\boldsymbol{\theta} \mid \boldsymbol{Y}), p(\boldsymbol{\theta} \mid \boldsymbol{Y}_{(I)})]$, some authors such as Johnson and Geisser (1983) and Guttman and Peña (1993) suggested using a symmetric version, for instance, the average of the measurements $K[p(\boldsymbol{\theta} \mid \boldsymbol{Y}_{(I)}), p(\boldsymbol{\theta} \mid \boldsymbol{Y})]$ and $K[p(\boldsymbol{\theta} \mid \boldsymbol{Y}), p(\boldsymbol{\theta} \mid \boldsymbol{Y}_{(I)})]$, known as *Kullback-Leibler symmetric divergence*, to replace the KLD measurement $K_I(\boldsymbol{\theta})$. Since the measurement $K[p(\boldsymbol{\theta} \mid \boldsymbol{Y}), p(\boldsymbol{\theta} \mid \boldsymbol{Y}_{(I)})]$ can be calculated in a similar manner to $K_I(\boldsymbol{\theta})$, our attention here is directly paid to the KLD measurement in (2.1). For the measurement $K_I(\boldsymbol{\theta})$, we have to explore some properties which are related to Bayesian diagnostics.

LEMMA 2.1. *The KLD measurement* (2.1) *is invariant under a one-to-one measurable transformation of the parameter vector* $\boldsymbol{\theta}$.

According to the definition of the KLD measurement, the argument in Lemma 2.1 holds obviously. It indicates that appropriate transformations of parameters can be made and the resulting KLD quantities are invariant in order to calculate analytically the KLD measurement for some complicated models. This conclusion will be frequently used in this paper. The following lemma shows the relationship between the KLD measurement and its marginal form.

LEMMA 2.2. *Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1 : \boldsymbol{\theta}_2)$. The KLD measurement $K_I(\boldsymbol{\theta})$ can then be decomposed into*

$$K_I(\boldsymbol{\theta}) = K_I(\boldsymbol{\theta}_1) + E_{p(\boldsymbol{\theta}_1|\boldsymbol{Y}_{(I)})}[K_I(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1)] = K_I(\boldsymbol{\theta}_2) + E_{p(\boldsymbol{\theta}_2|\boldsymbol{Y}_{(I)})}[K_I(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2)], \tag{2.2}$$

where $K_I(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1) = K[p(\boldsymbol{\theta}_2 \mid (\boldsymbol{\theta}_1, \boldsymbol{Y}_{(I)})), p(\boldsymbol{\theta}_2 \mid (\boldsymbol{\theta}_1, \boldsymbol{Y}))]$ denotes the KLD measurement between the posterior densities $p(\boldsymbol{\theta}_2 \mid (\boldsymbol{\theta}_1, \boldsymbol{Y}_{(I)}))$ and $p(\boldsymbol{\theta}_2 \mid (\boldsymbol{\theta}_1, \boldsymbol{Y}))$.

Lemma 2.2 provides a decomposition of the KLD measurement, which is useful when only a subset of parameters, e.g., the regression coefficient $\boldsymbol{B}$, in the model is of interest. Particularly, if the random variables $\boldsymbol{\theta}_1 \mid \boldsymbol{Y}$ and $\boldsymbol{\theta}_2 \mid \boldsymbol{Y}$ are mutually independent, then we have $K_I(\boldsymbol{\theta}) = K_I(\boldsymbol{\theta}_1) + K_I(\boldsymbol{\theta}_2)$, which indicates that the KLD measurement on the total parameter $\boldsymbol{\theta} = (\boldsymbol{\theta}_1 : \boldsymbol{\theta}_2)$ is the sum of those on the individual components $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$.

2.2 *Kullback-Leibler divergence for matrix-variate distributions*

In the GCM with UC, we will see that the posterior distributions of the regression coefficient $\boldsymbol{B}$ and dispersion component $\boldsymbol{\Sigma}$ involve matrix-variate normal distribution and Wishart distribution, respectively. The KLD measurements for these two distributions thus need to be studied first, and the related results are presented as follows.

LEMMA 2.3. *Suppose $p_1(\bullet)$ and $p_2(\bullet)$ are the density functions of the matrix-variate normal distributions $N_{p,n}(\boldsymbol{M}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\Omega}_1)$ and $N_{p,n}(\boldsymbol{M}_2, \boldsymbol{\Sigma}_2, \boldsymbol{\Omega}_2)$, respectively, then the KLD measurement between $p_1(\bullet)$ and $p_2(\bullet)$ is given by*

$$\begin{aligned} 2K(p_1(\bullet), p_2(\bullet)) = &-pn - p\log\det(\boldsymbol{\Omega}_1\boldsymbol{\Omega}_2^{-1}) - n\log\det(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}) \\ &+ \operatorname{tr}(\boldsymbol{\Omega}_1\boldsymbol{\Omega}_2^{-1})\operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}) + \operatorname{tr}\{(\boldsymbol{M}_2 - \boldsymbol{M}_1)\boldsymbol{\Omega}_2^{-1}(\boldsymbol{M}_2 - \boldsymbol{M}_1)^\tau\boldsymbol{\Sigma}_2^{-1}\}, \end{aligned} \tag{2.3}$$

where $\boldsymbol{M}_i$, $\boldsymbol{\Sigma}_i > \mathbf{0}$ and $\boldsymbol{\Omega}_i > \mathbf{0}$ ($i = 1, 2$) are the $p \times n$, $p \times p$ and $n \times n$ matrices, respectively.

In particular, when $n = 1$, i.e., $p_i(\bullet)$ is the density of the p-variate normal distribution, say $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ ($i = 1, 2$), then (2.3) is reduced to

$$2K(p_1(\bullet), p_2(\bullet)) = -p - \log\det(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}) + \operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}) + (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^\tau\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1).$$

This special case was considered in the literature, for example, by Guttman and Peña (1993). For the Wishart distribution, it seems to us that the KLD measurement was not given in the literature before. We find that it is relatively simple and is presented in the following Lemma.

LEMMA 2.4. *Suppose $p_1(\bullet)$ and $p_2(\bullet)$ are the density functions of the Wishart distributions $W_p(n_1, \boldsymbol{\Sigma}_1)$ and $W_p(n_2, \boldsymbol{\Sigma}_2)$, respectively, then the KLD measurement between $p_1(\bullet)$ and $p_2(\bullet)$ is given by*

$$2K(p_1(\bullet), p_2(\bullet)) = c + n_1\operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}) - n_2\log\det(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}), \tag{2.4}$$

where $n_i \geq p$, $\boldsymbol{\Sigma}_i > \mathbf{0}$ $(i = 1, 2)$ *and* c *is a constant which is independent of the parameters* $\boldsymbol{\Sigma}_1$ *and* $\boldsymbol{\Sigma}_2$.

Obviously, Lemma 2.4 is a generalization of Guttman and Peña (1993), where the KLD measurement for χ^2-distribution was considered.

3. Bayesian influence measurements in the GCM with UC

In this section, the KLD entropy and its properties shown in Section 2 are employed to measure the effects of a subset of observations on the growth fittings. With the non-informative prior (1.3), the Bayesian influence measures based on the KLD measurement are established in the GCM with UC.

3.1 *Posterior distributions*

On the posterior distribution of the regression coefficient $\boldsymbol{B}$ with respect to the non-informative prior distribution (1.3), it is a matrix-variate student t-distribution. More specifically, we have

LEMMA 3.1. *In the GCM with UC, with the non-informative prior density* (1.3), *the posterior distribution of the regression coefficient* $\boldsymbol{B}$ *is a matrix-variate student t-distribution*

$$\boldsymbol{B} \mid \boldsymbol{Y} \sim t_{m,r}(\hat{\boldsymbol{B}}, \boldsymbol{N}^{-1}, \boldsymbol{R}, \nu), \tag{3.1}$$

where $\hat{\boldsymbol{B}}$ *is the MLE of* $\boldsymbol{B}$, *i.e.*, $\hat{\boldsymbol{B}} = (\boldsymbol{X}^\tau \boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\tau \boldsymbol{S}^{-1}\boldsymbol{Y}\boldsymbol{Z}^\tau(\boldsymbol{Z}\boldsymbol{Z}^\tau)^{-1}$, $\boldsymbol{N} = \boldsymbol{X}^\tau \boldsymbol{S}^{-1}\boldsymbol{X}$, $\boldsymbol{S} = \boldsymbol{Y}(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Z}^\tau})\boldsymbol{Y}^\tau$, $\boldsymbol{R} = (\boldsymbol{Z}\boldsymbol{Z}^\tau)^{-1} + (\boldsymbol{Z}\boldsymbol{Z}^\tau)^{-1}\boldsymbol{Z}\boldsymbol{Y}^\tau\boldsymbol{Q}(\boldsymbol{Q}^\tau\boldsymbol{S}\boldsymbol{Q})^{-1}\boldsymbol{Q}^\tau\boldsymbol{Y}\boldsymbol{Z}^\tau(\boldsymbol{Z}\boldsymbol{Z}^\tau)^{-1}$ *and* $\nu = n - m - r + 1$.

On the matrix-variate student t-distribution (3.1), the first component $\hat{\boldsymbol{B}}$ denotes its location, the second and third represent the associated dispersion component, and the fourth component is the degree of freedom in the distribution. For more details on the matrix-variate student t-distribution, see Dickey (1967) and Muirhead (1982). In Lemma 3.1, the matrix $\boldsymbol{R}$ involved in (3.1) seems to be somewhat complicated, but its inverse has a relatively simple form $\boldsymbol{R}^{-1} = \boldsymbol{Z}(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Y}^\tau\boldsymbol{Q}})\boldsymbol{Z}^\tau$, which is independent of the specific choice of $\boldsymbol{Q} \in \mathcal{Q}$ because $\boldsymbol{P}_{\boldsymbol{Y}^\tau\boldsymbol{Q}}$ is a projection matrix. This fact will be useful in our subsequent consideration. On the other hand, our major concern about the regression coefficient $\boldsymbol{B}$ is to calculate its KLD measurement. The exact form of the KLD entropy for the matrix-variate student t-distribution, however, is too complicated so that an appropriate approximation to the KLD measurement could be considered alternatively. For the matrix-variate student t-distribution (3.1), in particular when the sample size n is sufficiently large, it can be approximated by a matrix-variate normal distribution

$$\boldsymbol{B} \mid \boldsymbol{Y} \dot{\sim} N_{m,r}(\hat{\boldsymbol{B}}, [\nu\boldsymbol{N}]^{-1}, \boldsymbol{R}) \tag{3.2}$$

(see, e.g., Pan (1995)), where the notation "$\dot{\sim}$" means "approximately distributed". In fact the approximation is in density convergence so that the KLD measurement for matrix-variate normal distribution can be used to approximate that for matrix-variate student t-distribution. More details can be referred to Box and Tiao (1968) and Dickey (1967).

The posterior distribution of $\mathbf{\Sigma}^{-1}$ is somewhat complicated and actually it can be viewed as a mixture of two Wishart distributions, which makes the calculation of the KLD measurement of $\mathbf{\Sigma}^{-1}$, say $K_I(\mathbf{\Sigma}^{-1})$, rather difficult. Nevertheless, Lemma 2.1 guarantees that $K_I(\mathbf{\Sigma}^{-1})$ is invariant under all one-to-one transformations of $\mathbf{\Sigma}^{-1}$. We thus pay our attention to looking for such appropriate transformations so that $K_I(\mathbf{\Sigma}^{-1})$ can be calculated analytically. In what follows two such transformations are proposed and the posterior distributions of transformed parameters are established with respect to the non-informative prior (1.3). First, since rank$(\boldsymbol{X}) = m$, there exist two orthogonal matrices $\mathbf{\Gamma}$ and $\mathbf{\Gamma}^*$ with orders $p \times p$ and $m \times m$, respectively, such that

$$\boldsymbol{X} = \mathbf{\Gamma}\begin{pmatrix}\mathbf{\Lambda}\\ \mathbf{0}\end{pmatrix}\mathbf{\Gamma}^* \equiv (\mathbf{\Gamma}_1 : \mathbf{\Gamma}_2)\begin{pmatrix}\mathbf{\Lambda}\\ \mathbf{0}\end{pmatrix}\mathbf{\Gamma}^* \tag{3.3}$$

where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$, $\lambda_i^2 > 0$ $(1 \le i \le m)$ are the eigenvalues of $\boldsymbol{X}^\tau\boldsymbol{X}$, and $\mathbf{\Gamma}_1$ with order $p \times m$ is the submatrix of $\mathbf{\Gamma}$. Now, make the transformation

$$\mathbf{\Omega} = \mathbf{\Gamma}^\tau\mathbf{\Sigma}^{-1}\mathbf{\Gamma} = \begin{pmatrix}\mathbf{\Omega}_{11} & \mathbf{\Omega}_{12}\\ \mathbf{\Omega}_{21} & \mathbf{\Omega}_{22}\end{pmatrix}, \tag{3.4}$$

where the partition of $\mathbf{\Omega}$ corresponds to that of $\mathbf{\Gamma}$. Second, set a further transformation

$$\boldsymbol{V} = \begin{pmatrix}\boldsymbol{V}_{11} & \boldsymbol{V}_{12}\\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22}\end{pmatrix} \equiv \begin{pmatrix}\mathbf{\Omega}_{11} & \mathbf{\Omega}_{11}^{-1}\mathbf{\Omega}_{12}\\ \mathbf{\Omega}_{21}\mathbf{\Omega}_{11}^{-1} & \mathbf{\Omega}_{22.1}\end{pmatrix}, \tag{3.5}$$

where $\mathbf{\Omega}_{22.1} = \mathbf{\Omega}_{22} - \mathbf{\Omega}_{21}\mathbf{\Omega}_{11}^{-1}\mathbf{\Omega}_{12}$. For those transformed parameters, we have the following conclusions.

LEMMA 3.2. *In the GCM with UC, with the non-informative prior* (1.3), *the posterior distribution of the reparameterized dispersion matrix* $\boldsymbol{V}$ *consists of*

$$\begin{aligned} &\boldsymbol{V}_{11} \mid \boldsymbol{Y} \sim W_m(n-r, \mathbf{\Psi}_1),\\ &\boldsymbol{V}_{12} \mid (\boldsymbol{V}_{11}, \boldsymbol{Y}) \sim N_{m,(p-m)}(\boldsymbol{V}^*, \boldsymbol{V}_{11}^{-1}, \mathbf{\Psi}_2^{-1}),\\ &\boldsymbol{V}_{22} \mid \boldsymbol{Y} \sim W_{p-m}(n-m, \mathbf{\Psi}_3^{-1}), \end{aligned} \tag{3.6}$$

and $\boldsymbol{V}_{22} \mid \boldsymbol{Y}$ *is independent of* $(\boldsymbol{V}_{11}, \boldsymbol{V}_{12}) \mid \boldsymbol{Y}$, *where* $\boldsymbol{V}^* = -(\mathbf{\Gamma}_1^\tau\boldsymbol{S}\mathbf{\Gamma}_2)(\mathbf{\Gamma}_2^\tau\boldsymbol{S}\mathbf{\Gamma}_2)^{-1}$, $\mathbf{\Psi}_1 = \mathbf{\Gamma}_1^\tau\boldsymbol{S}^{-1}\mathbf{\Gamma}_1$, $\mathbf{\Psi}_2 = \mathbf{\Gamma}_2^\tau\boldsymbol{S}\mathbf{\Gamma}_2$, $\mathbf{\Psi}_3 = \mathbf{\Gamma}_2^\tau\boldsymbol{Y}\boldsymbol{Y}^\tau\mathbf{\Gamma}_2$, $\boldsymbol{S}$ *is given in Lemma* 3.1, *and* $\mathbf{\Gamma}_i$ $(i = 1, 2)$ *is defined by* (3.3).

By the use of Lemma 3.1 and Lemma 3.2, we can also obtain the conditional posterior distribution of $\boldsymbol{B}$ given $\mathbf{\Sigma}^{-1}$, which will be useful for calculating the KLD measurement of the parameter pair $(\boldsymbol{B}, \mathbf{\Sigma})$.

LEMMA 3.3. *In the GCM with UC, with the non-informative prior* (1.3), *the conditional posterior distribution of* $\boldsymbol{B}$ *given* $\mathbf{\Sigma}^{-1}$ *can be written as*

$$\boldsymbol{B} \mid (\mathbf{\Sigma}^{-1}, \boldsymbol{Y}) \sim N_{m,r}(\hat{\boldsymbol{B}}_{\mathbf{\Sigma}}, (\boldsymbol{X}^\tau\mathbf{\Sigma}\boldsymbol{X})^{-1}, (\boldsymbol{Z}\boldsymbol{Z}^\tau)^{-1}), \tag{3.7}$$

where $\hat{\boldsymbol{B}}_{\mathbf{\Sigma}} = (\boldsymbol{X}^\tau\mathbf{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\tau\mathbf{\Sigma}^{-1}\boldsymbol{Y}\boldsymbol{Z}^\tau(\boldsymbol{Z}\boldsymbol{Z}^\tau)^{-1}$, *i.e., the matrix* $\boldsymbol{S}$ *in the MLE of* $\hat{\boldsymbol{B}}$ *is replaced by the matrix* $\mathbf{\Sigma}$.

For the technical details on proofs of Lemma 3.2 and Lemma 3.3, one can refer to Pan *et al.* (1998).

3.2 *Bayesian influence measurements*

In this subsection the KLD measurements for the regression coefficient $\boldsymbol{B}$, the dispersion component $\boldsymbol{\Sigma}$, and their combination $(\boldsymbol{B}, \boldsymbol{\Sigma})$ are considered, respectively. Since the technical details of the derivations on the KLD measurements are troublesome, we only summarize the results in the following theorems and postpone the proofs until the Appendix of this paper.

Let $I = \{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots, n\}$ $(n > p + k)$ be a subset containing the indices of the k individuals to be deleted, where the number k is given in advance. Without loss of generality, the index set can be assumed to be $I = \{n-k+1, n-k+2, \ldots, n\}$ so that the response matrix $\boldsymbol{Y}$ can be partitioned into $\boldsymbol{Y} = (\boldsymbol{Y}_{(I)} : \boldsymbol{Y}_I)$, where $\boldsymbol{Y}_I = (\boldsymbol{y}_{n-k+1}, \boldsymbol{y}_{n-k+2}, \ldots, \boldsymbol{y}_n)$. Correspondingly, the matrices $\boldsymbol{Z}$ and $\boldsymbol{\epsilon}$ are partitioned into $\boldsymbol{Z} = (\boldsymbol{Z}_{(I)} : \boldsymbol{Z}_I)$ and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_{(I)} : \boldsymbol{\epsilon}_I)$, respectively. Then, the GCM after deleting $\boldsymbol{Y}_I$ becomes

$$\begin{cases} \boldsymbol{Y}_{(I)} = \boldsymbol{X}\boldsymbol{B}\boldsymbol{Z}_{(I)} + \boldsymbol{\epsilon}_{(I)} \\ \boldsymbol{\epsilon}_{(I)} \sim N_{p,(n-k)}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_{n-k}), \end{cases} \tag{3.8}$$

which is known as the *multiple individual deletion model* (MIDM), see, e.g., Pan and Fang (1995). For the MIDM (3.8) with UC, Lemma 3.1 also implies that

$$\boldsymbol{B} \mid \boldsymbol{Y}_{(I)} \sim t_{m,r}(\hat{\boldsymbol{B}}_{(I)},\ \boldsymbol{N}_{(I)}^{-1},\ \boldsymbol{R}_{(I)}, \nu^*), \tag{3.9}$$

where $\hat{\boldsymbol{B}}_{(I)} = (\boldsymbol{X}^{\tau}\boldsymbol{S}_{(I)}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}_{(I)}^{-1}\boldsymbol{Y}_{(I)}\boldsymbol{Z}_{(I)}^{\tau}(\boldsymbol{Z}_{(I)}\boldsymbol{Z}_{(I)}^{\tau})^{-1}$, $\boldsymbol{S}_{(I)} = \boldsymbol{Y}_{(I)}(\boldsymbol{I}_{n-k} - \boldsymbol{P}_{\boldsymbol{Z}_{(I)}^{\tau}})\boldsymbol{Y}_{(I)}^{\tau}$, $\boldsymbol{R}_{(I)} = (\boldsymbol{Z}_{(I)}\boldsymbol{Z}_{(I)}^{\tau})^{-1} + (\boldsymbol{Z}_{(I)}\boldsymbol{Z}_{(I)}^{\tau})^{-1}\boldsymbol{Z}_{(I)}\boldsymbol{Y}_{(I)}^{\tau}\boldsymbol{Q}(\boldsymbol{Q}^{\tau}\boldsymbol{S}_{(I)}\boldsymbol{Q})^{-1}\boldsymbol{Q}^{\tau}\boldsymbol{Y}_{(I)}\boldsymbol{Z}_{(I)}^{\tau}(\boldsymbol{Z}_{(I)}\boldsymbol{Z}_{(I)}^{\tau})^{-1}$, $\boldsymbol{N}_{(I)} = \boldsymbol{X}^{\tau}\boldsymbol{S}_{(I)}^{-1}\boldsymbol{X}$ and $\nu^* = n - k - m - r + 1$. Due to the reasons mentioned above, an approximation to the matrix-variate student t-distribution (3.9) should be considered alternatively. In particular, when the sample size n is sufficiently large, (3.9) can be approximated by a matrix-variate normal distribution

$$\boldsymbol{B} \mid \boldsymbol{Y}_{(I)} \dot{\sim} N_{m,r}(\hat{\boldsymbol{B}}_{(I)}, [\nu^*\boldsymbol{N}_{(I)}]^{-1}, \boldsymbol{R}_{(I)}). \tag{3.10}$$

Based on (3.2) and (3.10), the KLD measurement for the regression coefficient $\boldsymbol{B}$ can be established approximately, which is summarized in the following theorem.

THEOREM 3.1. *In the GCM with UC, with the non-informative prior* (1.3), *the KLD measurement of the regression coefficient* $\boldsymbol{B}$ *can be written approximately as*

$$\begin{aligned} 2K_I(\boldsymbol{B}) \approx{}& (n-m-r+1) \\ &\cdot \operatorname{tr}\{\boldsymbol{K}_I^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1}\boldsymbol{Z}(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Y}^{\tau}\boldsymbol{Q}})\boldsymbol{Z}^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1}\boldsymbol{K}_I\boldsymbol{V}_I^{-1}\boldsymbol{A}_I\boldsymbol{V}_I^{-1}\} \\ &+ \frac{n-m-r+1}{n-k-m-r+1}[m - \operatorname{tr}\{\boldsymbol{A}_I\boldsymbol{V}_I^{-1}\}] \\ &\cdot [\operatorname{tr}\{(\boldsymbol{P}_{\boldsymbol{Z}^{\tau}} + \boldsymbol{Z}^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1}\boldsymbol{Z}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}\boldsymbol{Z}_I^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1}\boldsymbol{Z})(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Y}^{\tau}\boldsymbol{Q}})\} \\ &\quad + \operatorname{tr}\{(\boldsymbol{Y}\boldsymbol{Z}^{\tau} - \boldsymbol{E}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}\boldsymbol{Z}_I^{\tau})(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1}\boldsymbol{Z}(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Y}^{\tau}\boldsymbol{Q}})\boldsymbol{Z}^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1} \\ &\quad \cdot (\boldsymbol{Y}\boldsymbol{Z}^{\tau} - \boldsymbol{E}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}\boldsymbol{Z}_I^{\tau})^{\tau}\boldsymbol{Q}(\boldsymbol{Q}^{\tau}\boldsymbol{S}\boldsymbol{Q})^{-1}\boldsymbol{Q}^{\tau} \\ &\quad \cdot (\boldsymbol{I}_p + \boldsymbol{E}_I\boldsymbol{V}_I^{-1}\boldsymbol{E}_I^{\tau}\boldsymbol{Q}(\boldsymbol{Q}^{\tau}\boldsymbol{S}\boldsymbol{Q})^{-1}\boldsymbol{Q}^{\tau})\}] \\ &- m\log\det\{\boldsymbol{I}_p + (\boldsymbol{Y}\boldsymbol{P}_{\boldsymbol{Z}^{\tau}}\boldsymbol{Y}^{\tau} - \boldsymbol{Y}_I\boldsymbol{Y}_I^{\tau} + \boldsymbol{E}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}\boldsymbol{E}_I^{\tau}) \\ &\qquad \cdot \boldsymbol{Q}(\boldsymbol{Q}^{\tau}\boldsymbol{S}\boldsymbol{Q})^{-1}\boldsymbol{Q}^{\tau}(\boldsymbol{I}_p + \boldsymbol{E}_I\boldsymbol{V}_I^{-1}\boldsymbol{E}_I^{\tau}\boldsymbol{Q}(\boldsymbol{Q}^{\tau}\boldsymbol{S}\boldsymbol{Q})^{-1}\boldsymbol{Q}^{\tau})\} \\ &- m\log\det\{(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}\} - r\log\Lambda_I + c_1, \end{aligned} \tag{3.11}$$

where $K_I = Z_I - ZY^{\tau}Q_S^{\tau}S^{-1}Q_S E_I$, $V_I = I_k - H_I - E_I^{\tau}Q_S^{\tau}S^{-1}Q_S E_I$, $Q_S = SQ(Q^{\tau}SQ)^{-1}Q^{\tau}$, $Q \in \mathcal{Q}$, $A_I = E_I^{\tau}S^{-1}X(X^{\tau}S^{-1}X)^{-1}X^{\tau}S^{-1}E_I$, $P_{Y^{\tau}Q} = Y^{\tau}Q(Q^{\tau}YY^{\tau}Q)^{-1}Q^{\tau}Y$, $H_I = Z_I^{\tau}(ZZ^{\tau})^{-1}Z_I$, $E = Y(I_n - P_{Z^{\tau}}) = (E_{(I)} : E_I)$, $\Lambda_I = \det\{I_k + A_I(I_k - H_I - E_I^{\tau}S^{-1}E_I)^{-1}\}^{-1}$, *and the constant* c_1 *is independent of the index subset* I.

Theorem 3.1 implies that the KLD measurement of the regression coefficient B not only depends on the generalized Cook's distance $D_I(R^{-1}, (X^{\tau}S^{-1}X)^{-1})$ given by (A.3) in the Appendix, but also contains the information of discordant outliers provided by the statistic Λ_I, where Λ_I is actually the likelihood ratio statistic for detecting multiple discordant outliers, see Pan and Fang (1995) for more details. On the other hand, for the dispersion component Σ, the following theorem presents the analytical form of its KLD measurement.

THEOREM 3.2. *In the GCM with UC, with the non-informative prior* (1.3), *the KLD measurement of the dispersion component* Σ *can be expressed as*

$$
\begin{aligned}
(3.12)\quad 2K_I(\Sigma) = {} & (n-k-r)\,\mathrm{tr}\{A_I(I_k - H_I - E_I^{\tau}S^{-1}E_I)^{-1}\} + (n-r)\log\Lambda_I \\
& + (n-k-m)\,\mathrm{tr}\{(I_k - \Delta_I^*)^{-1}\} + (n-m)\log\det\{(I_k - \Delta_I^*)\} \\
& + m\,\mathrm{tr}\{(I_k - H_I)V_I^{-1}\} - m\log\det\{(I_k - H_I)V_I^{-1}\} \\
& + (n-k-r)\,\mathrm{tr}\{A_I(I_k - H_I - E_I^{\tau}S^{-1}E_I)^{-1}(I_k - H_I - V_I)V_I^{-1}\} + c_2,
\end{aligned}
$$

where the definitions of A_I, H_I, E_I, V_I *and* Λ_I *are the same as those given in Theorem* 3.1, *and* $\Delta_I^* = Y_I^{\tau}Q(Q^{\tau}YY^{\tau}Q)^{-1}Q^{\tau}Y_I$. *The constant* c_2 *does not depend on the index subset* I.

When both of the regression coefficient B and the dispersion component Σ are of interest, the associated KLD measurement can be calculated in terms of those on B and Σ, according to Lemma 2.2. The main results are summarized in what follows.

THEOREM 3.3. *In the GCM with UC, with the non-informative prior* (1.3), *the KLD measurement of the parameter pair* (B, Σ) *can be simplified to*

$$
\begin{aligned}
(3.13)\quad 2K_I(B,\Sigma) = {} & (n-k-r)\,\mathrm{tr}\{A_I(I_k - H_I - E_I^{\tau}S^{-1}E_I)^{-1}\} + (n-r)\log\Lambda_I \\
& + (n-k-m)\,\mathrm{tr}\{(I_k - \Delta_I^*)^{-1}\} + (n-m)\log\det\{(I_k - \Delta_I^*)\} \\
& + m\,\mathrm{tr}\{(I_k - H_I)V_I^{-1}\} - m\log\det\{(I_k - H_I)V_I^{-1}\} \\
& + (n-k-r)\,\mathrm{tr}\{A_I(I_k - H_I - E_I^{\tau}S^{-1}E_I)^{-1}(I_k - H_I - V_I)V_I^{-1}\} \\
& + m\log\det\{(I_k - H_I)\} + m\,\mathrm{tr}\{H_I(I_k - H_I - E_I^{\tau}S^{-1}E_I)^{-1}\} \\
& + (n-k-m-r)\,\mathrm{tr}\{H_I(I_k - H_I - E_I^{\tau}S^{-1}E_I)^{-1} \\
& \qquad \cdot (A_I - A_I V_I^{-1}A_I)(I_k - H_I - E_I^{\tau}S^{-1}E_I)^{-1}\} + c_3,
\end{aligned}
$$

where the definitions of A_I, Δ_I^*, V_I *and* Λ_I *are the same as those defined in Theorems* 3.1–3.2, *and the constant* c_3 *is independent of the index subset* I.

Although the results in Theorems 3.1–3.3 are somewhat troublesome, the KLD measurements are of analytically closed forms. So, the computations on those measurements are straightforward and intensive computations such as Gibbs sampling methods used by Carlin and Polson (1991) are not necessary in the GCM.

4. An illustrative example

In this section, the diagnostic techniques proposed in this paper are applied to a practical data set analyzed by Rao (1987), and Lee (1988, 1991). The primary objective is to illustrate applications of the Bayesian influence measures. Thus, an unstructured covariance matrix is assumed for the practical data set in our analysis. For more details about the model selection criteria with respect to the covariance structure, one can refer to Lee (1991). In addition, for simplicity, we only consider single case-deletion. In other words, each single observation is deleted in turn with $k = 1$ and $I = \{i\}$ $(1 \leq i \leq n)$. In the data analysis, when either one of the parameters $\boldsymbol{B}$ and $\boldsymbol{\Sigma} > \mathbf{0}$ or both of them are of interest, the KLD measurements $K_i^*(\boldsymbol{B})$, $K_i^*(\boldsymbol{\Sigma})$ and $K_i^*(\boldsymbol{B}, \boldsymbol{\Sigma})$ are calculated using Theorems 3.1–3.3, respectively, where $K_i^*(\bullet)$ is the KLD measurement $K_i(\bullet)$ without the associated constant c_i. In other words, the measurement is defined by $K_i^*(\bullet) = (2K_i(\bullet) - c_i)/2$, where c_i is the constant involved in the theorem, see (3.11)–(3.13).

Dental Data This data set was first considered by Potthoff and Roy (1964) and later analyzed by Lee and Geisser (1975), Rao (1987) and Lee (1991) for different study purposes, focusing mainly on statistical inferences such as estimation, testing hypothesis and prediction. Dental measurements were made on 11 girls and 16 boys at age 8, 10, 12 and 14 years. Each measurement is the distance, in millimeters, from the center of the pituitary to the pterygomaxillary fissure.

Since the measurements are obtained at equal time intervals, the design matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$, respectively, take the following forms:

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 8 & 10 & 12 & 14 \end{pmatrix}^{T} \quad \text{and} \quad \boldsymbol{Z} = \begin{pmatrix} \mathbf{1}_{11}^{T} & 0 \\ 0 & \mathbf{1}_{16}^{T} \end{pmatrix},$$

where $\mathbf{1}_m$ is a $m \times 1$ vector with all components 1's. The structure of the between-design matrix $\boldsymbol{Z}$ emphasizes that two different groups of the observations are involved in the study. For this data set, based on the likelihood case-deletion approach, Pan and Fang (1995, 1996) indicated that the 24th observation, which belongs to the boy group, is a discordant outlier. Also the 20th observation is the most influential observation. Analogous conclusions were obtained by von Rosen (1995) in terms of a neighborhood method based on the Taylor expansions.

Now, we use the Bayesian case-deletion approach to analyze this data set. In this example the regression coefficient $\boldsymbol{B}$ and the dispersion component $\boldsymbol{\Sigma} > \mathbf{0}$ are 2×2 and 4×4 matrices, respectively. When we are concerned with influence analysis on either parameter, i.e., $\boldsymbol{B}$ or $\boldsymbol{\Sigma}$, or both parameters, i.e., $(\boldsymbol{B}, \boldsymbol{\Sigma})$, the numerical results are calculated in terms of the influence measurements given in Section 3. Figure 1 gives the index plot of the KLD measurement $K_i^*(\boldsymbol{B})$. We observe that the effects of the 24th, 20th and 15th observations on the posterior distribution of $\boldsymbol{B}$ stand out. In particular, the 24th observation has the largest influence on the regression coefficient $\boldsymbol{B}$. In fact, it is also a discordant outlier in terms of a mean shift regression model (Pan and Fang (1995)).

For the KLD measurement of the dispersion parameter $\boldsymbol{\Sigma}$ displayed in Fig. 2, however, the largest influence is achieved at the 20th observation. The 24th observation has the second largest influence on the posterior distributions of $\boldsymbol{\Sigma}$. This is in agreement with the results of Pan and Fang (1995) and von Rosen (1995). In addition, Fig. 2 also

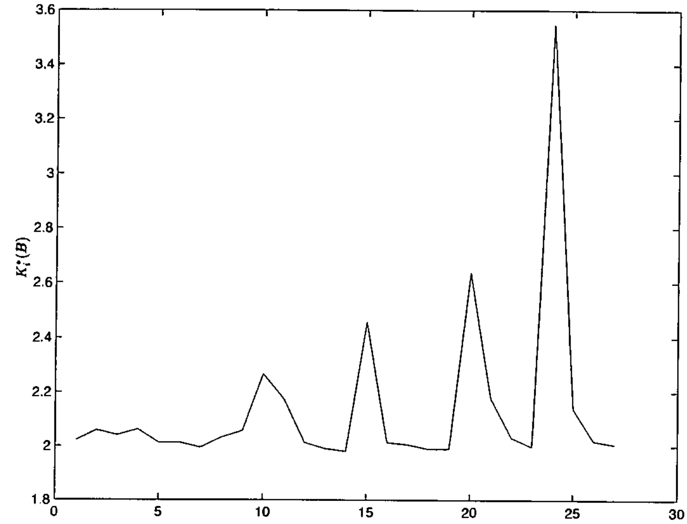

Fig. 1. Index plot of $K_i^*(\boldsymbol{B})$ for the Dental data.

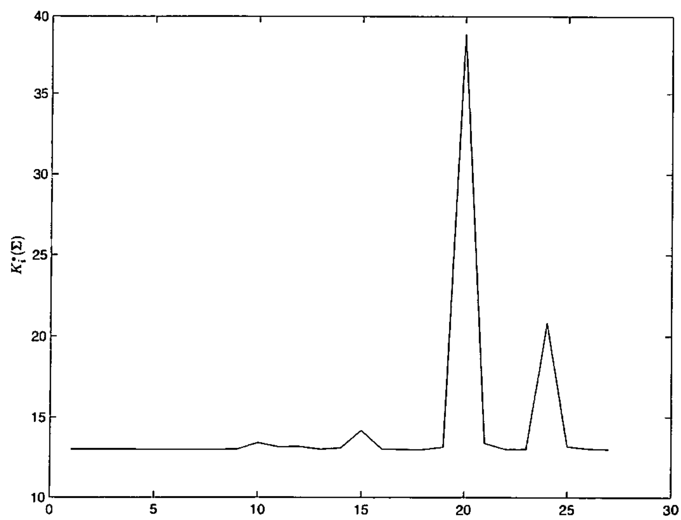

Fig. 2. Index plot of $K_i^*(\boldsymbol{\Sigma})$ for the Dental data.

shows that the effects of the girl group, indexed by the first eleven observations, on the Bayesian inference in the GCM are significantly less than those of the boy group, indexed from the 12th to 27th observations. The influence measures of the joint parameters $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ for this data set is also calculated. It gives us very similar diagnostics information to the KLD measurement of the dispersion component, i.e., $K_i^*(\boldsymbol{\Sigma})$. For brevity, it is omitted here.

We are suggested by a referee that it may be worthwhile looking into the results when the boys and girls are treated separately, because Lee (1988) already indicated that the prediction is better when they are treated as from two different populations. By treating the boys and girls separately, we found that in the boy group the information on influential observations is very similar to that when the boys and girls are treated together. In the girl group, however, we do find more information when it is studied

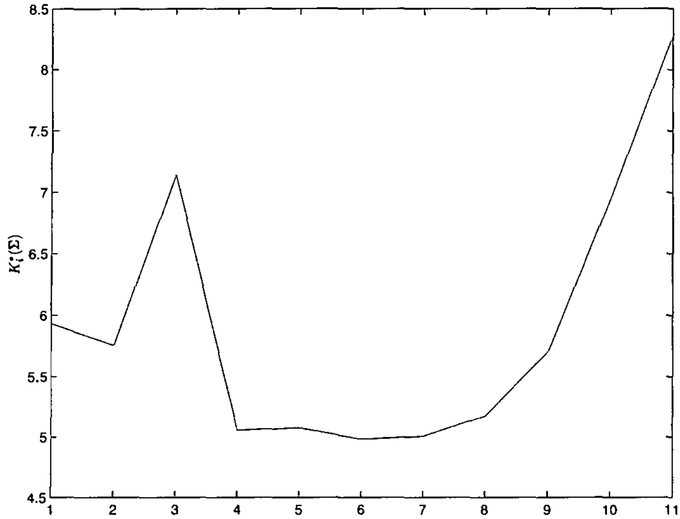

Fig. 3. Index plot of $K_i^*(\Sigma)$ for the girl group in Dental data.

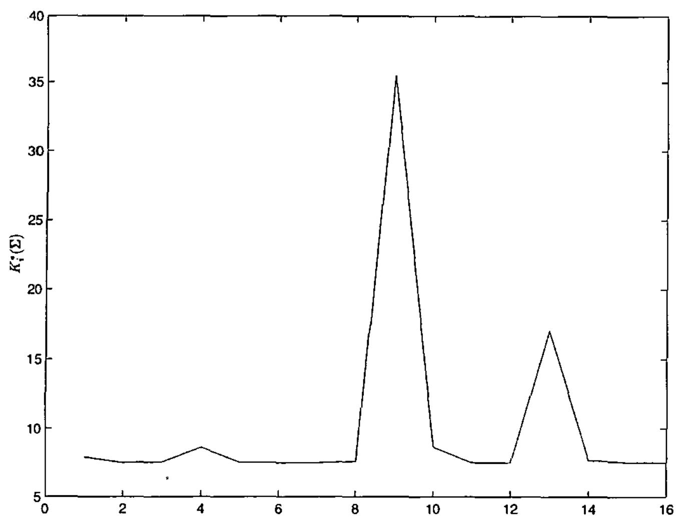

Fig. 4. Index plot of $K_i^*(\Sigma)$ for the boy group in Dental data.

separately. To save space, we only display the index plots of KLD measurements for the dispersion component $\mathbf{\Sigma}$ here. Figure 3 presents the KLD measurement $K_i^*(\mathbf{\Sigma})$ for the girl group while Fig. 4 is for the boy group. Figure 3 indicates that the 11th, 10th and 3rd observations in the girl group stand out and may be the influential observations on the Bayesian inference for the girls. Figure 4 reveal quite similar information to Fig. 2, in other words, the 9th and 13th observations in the boy group have large influence on the Bayesian inference of the boy population, which are the 20th and 24th observations in Fig. 2, respectively.

In summary, the performance of the KLD measurement on the girl group is hidden when the boys and girls are treated together because the influence of the boy group is considerably much larger than that of the girl group. Therefore, it is better to treat them separately, in particular when there is heterogeneity between the groups, which confirms Lee's (1988) conclusion from the view point of diagnostics.

5. Discussion

In comparison with the likelihood case-deletion approach in the GCM with UC (Pan and Fang (1996)), the Bayesian case-deletion technique can reveal more information on influential observations. Both our theoretical and numerical results show this point well. For example, for the regression coefficient $\boldsymbol{B}$, (3.11) implies that the KLD measurement $K_I(\boldsymbol{B})$ not only depends on the generalized Cook's distance $D_I(\boldsymbol{R}^{-1}, (\boldsymbol{X}^T\boldsymbol{S}^{-1}\boldsymbol{X})^{-1})$ given by (A.3), but also contains the information of multiple discordant outlier detection, i.e. Λ_I (Pan and Fang (1995)). Numerical analysis on the dental data shows that the Bayesian case-deletion method detects not only the 20th observation as influential but also the 24th observation as a potential influential point. The 24th observation is actually a discordant outlier (Pan and Fang (1995)).

Acknowledgements

We are grateful to the Managing Editor, Associate Editor and two anonymous referees for their very stimulating and helpful comments, which makes the presentation of this paper be improved significantly.

Appendix. Proofs of theorems

PROOF OF THEOREM 3.1. Firstly, by applying Lemma 2.3 to both (3.2) and (3.10) we know that the KLD measurement of $\boldsymbol{B}$ is approximately equal to

$$
\begin{aligned}
\text{(A.1)}\qquad 2K_I(\boldsymbol{B}) \approx & -mr - m\log\det(\boldsymbol{R}_{(I)}\boldsymbol{R}^{-1}) \\
& - r\log\left(\frac{\nu}{\nu^*}\right)^m \det\{(\boldsymbol{X}^T\boldsymbol{S}_{(I)}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{S}^{-1}\boldsymbol{X})\} \\
& + \frac{\nu}{\nu^*}\operatorname{tr}\{\boldsymbol{R}_{(I)}\boldsymbol{R}^{-1}\}\operatorname{tr}\{(\boldsymbol{X}^T\boldsymbol{S}_{(I)}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{S}^{-1}\boldsymbol{X})\} \\
& + \nu\operatorname{tr}\{(\hat{\boldsymbol{B}} - \hat{\boldsymbol{B}}_{(I)})\boldsymbol{R}^{-1}(\hat{\boldsymbol{B}} - \hat{\boldsymbol{B}}_{(I)})^T(\boldsymbol{X}^T\boldsymbol{S}^{-1}\boldsymbol{X})\},
\end{aligned}
$$

where $\boldsymbol{R}_{(I)}$ is defined in (3.9). Secondly, by using the relation between $\boldsymbol{S}$ and $\boldsymbol{S}_{(I)}$ given by Pan and Fang (1995) and setting $\boldsymbol{A}_I$ as the form in Theorem 3.1, we have

$$
\begin{aligned}
\text{(A.2)}\qquad & \operatorname{tr}\{(\boldsymbol{X}^T\boldsymbol{S}_{(I)}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{S}^{-1}\boldsymbol{X})\} = m - \operatorname{tr}\{\boldsymbol{A}_I\boldsymbol{V}_I^{-1}\}, \\
& \det\{(\boldsymbol{X}^T\boldsymbol{S}_{(I)}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{S}^{-1}\boldsymbol{X})\} = \det\{\boldsymbol{I}_k - \boldsymbol{A}_I\boldsymbol{V}_I^{-1}\} \equiv \Lambda_I,
\end{aligned}
$$

where $\boldsymbol{V}_I = \boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{E}_I^T\boldsymbol{S}^{-1}\boldsymbol{E}_I + \boldsymbol{A}_I$. Thirdly, the relation between $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{B}}_{(I)}$ presented by Pan and Fang (1995) shows

$$
\begin{aligned}
\text{(A.3)}\qquad & \operatorname{tr}\{(\hat{\boldsymbol{B}} - \hat{\boldsymbol{B}}_{(I)})\boldsymbol{R}^{-1}(\hat{\boldsymbol{B}} - \hat{\boldsymbol{B}}_{(I)})^T(\boldsymbol{X}^T\boldsymbol{S}^{-1}\boldsymbol{X})\} \\
& = \operatorname{tr}\{\boldsymbol{K}_I^T(\boldsymbol{Z}\boldsymbol{Z}^T)^{-1}\boldsymbol{Z}(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Y}^T\boldsymbol{Q}})\boldsymbol{Z}^T(\boldsymbol{Z}\boldsymbol{Z}^T)^{-1}\boldsymbol{K}_I\boldsymbol{V}_I^{-1}\boldsymbol{A}_I\boldsymbol{V}_I^{-1}\},
\end{aligned}
$$

where $\boldsymbol{K}_I = \boldsymbol{Z}_I - \boldsymbol{Z}\boldsymbol{Y}^T\boldsymbol{Q}(\boldsymbol{Q}^T\boldsymbol{S}\boldsymbol{Q})^{-1}\boldsymbol{Q}^T\boldsymbol{E}_I$ and $\boldsymbol{P}_{\boldsymbol{Y}^T\boldsymbol{Q}} = \boldsymbol{Y}^T\boldsymbol{Q}(\boldsymbol{Q}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{Q})^{-1}\boldsymbol{Q}^T\boldsymbol{Y}$, which is the generalized Cook's distance and denoted by $D_I(\boldsymbol{R}^{-1}, (\boldsymbol{X}^T\boldsymbol{S}^{-1}\boldsymbol{X})^{-1})$ (Pan and Fang

(1996)). Fourthly, we can show that

$$\text{(A.4)}\quad \operatorname{tr}\{R_{(I)}R^{-1}\} = \operatorname{tr}[\{P_{Z^\tau} + Z^\tau(ZZ^\tau)^{-1}Z_I(I_n - H_I)^{-1}Z_I^\tau(ZZ^\tau)^{-1}Z\}(I_n - P_{Y^\tau Q})]$$
$$\operatorname{tr}[\{YZ^\tau - E_I(I_n - H_I)^{-1}Z_I^\tau\}(ZZ^\tau)^{-1}Z(I_n - P_{Y^\tau Q})Z^\tau \cdot (ZZ^\tau)^{-1}\{ZY^\tau - Z_I(I_n - H_I)^{-1}E_I^\tau\} \cdot Q(Q^\tau SQ)^{-1}Q^\tau\{I_p + E_I V_I^{-1}E_I^\tau Q(Q^\tau SQ)^{-1}Q^\tau\}]$$

and

$$\text{(A.5)}\quad \det\{R_{(I)}R^{-1}\} = \det\{I_p + (YP_{Z^\tau}Y^\tau - Y_I Y_I^\tau + E_I(I_n - H_I)^{-1}E_I^\tau) \cdot Q(Q^\tau SQ)^{-1}Q^\tau(I_p + E_I V_I^{-1}E_I^\tau Q(Q^\tau SQ)^{-1}Q^\tau)\} \cdot \det\{(I_n - H_I)^{-1}\}\det\{P_{Z^\tau}(I_n - P_{Y^\tau Q})\}.$$

where the matrices H_I and E_I are defined in Theorem 3.1. Finally, by substituting (A.2)–(A.5) into (A.1), we know Theorem 3.1 holds and the proof is complete. □

PROOF OF THEOREM 3.2. Since the transformations from Σ^{-1} to V, given by (3.4) and (3.5), are obviously one-to-one, we have $K_I(\Sigma) = K_I(\Sigma^{-1}) = K_I(V)$ according to Lemma 2.1. On the other hand, using the partition of V given by (3.5), Lemma 2.2 shows that $K_I(V) = K_I(V_{11}, V_{12}) + K_I(V_{22})$ because $(V_{11}, V_{12}) \mid Y$ is independent of $V_{22} \mid Y$. Furthermore, $K_I(V)$ can be written as

$$\text{(A.6)}\qquad K_I(V) = K_I(V_{11}) + E_{p(V_{11}|Y_{(I)})}[K_I(V_{12} \mid V_{11})] + K_I(V_{22}).$$

According to Lemma 3.2, for the MIDM (3.8) with UC, we have $V_{11} \mid Y_{(I)} \sim W_m(n - k - r, \Gamma_1^\tau S_{(I)}^{-1}\Gamma_1)$, $V_{12} \mid (V_{11}, Y_{(I)}) \sim N_{m,(p-m)}(V_{(I)}^*, V_{11}^{-1}, (\Gamma_2^\tau S_{(I)}\Gamma_2)^{-1})$, $V_{22} \mid Y_{(I)} \sim W_{p-m}(n - k - m, (\Gamma_2^\tau Y_{(I)}Y_{(I)}^\tau\Gamma_2)^{-1})$, and $V_{22} \mid Y_{(I)}$ is independent of $(V_{11}, V_{12}) \mid Y$, where the definition of Γ_i $(i = 1, 2)$ is given by (3.3) and $V_{(I)}^* = -(\Gamma_1^\tau S_{(I)}\Gamma_2)(\Gamma_2^\tau S_{(I)}\Gamma_2)^{-1}$. By applying Lemma 2.3 and Lemma 2.4 to those facts, the KLD measurements of the parameters V_{11}, $V_{12} \mid V_{11}$ and V_{22} can be expressed as

$$\begin{aligned}
2K_I(V_{11}) &= c_1^* + (n - k - r)\operatorname{tr}\{(\Gamma_1^\tau S_{(I)}^{-1}\Gamma_1)(\Gamma_1^\tau S^{-1}\Gamma_1)^{-1}\} \\
&\quad - (n - r)\log\det\{(\Gamma_1^\tau S_{(I)}^{-1}\Gamma_1)(\Gamma_1^\tau S^{-1}\Gamma_1)^{-1}\}, \\
2K_I(V_{12} \mid V_{11}) &= -m(p - m) - m\log\det\{(\Gamma_2^\tau S_{(I)}\Gamma_2)^{-1}(\Gamma_2^\tau S\Gamma_2)\} \\
&\quad + m\operatorname{tr}\{(\Gamma_2^\tau S_{(I)}\Gamma_2)^{-1}(\Gamma_2^\tau S\Gamma_2)\} \qquad \text{(A.7)} \\
&\quad + \operatorname{tr}\{(V^* - V_{(I)}^*)(\Gamma_2^\tau S\Gamma_2)(V^* - V_{(I)}^*)^\tau V_{11}\}, \\
2K_I(V_{22}) &= c_2^* + (n - k - m)\operatorname{tr}\{(\Gamma_2^\tau Y_{(I)}Y_{(I)}^\tau\Gamma_2)^{-1}(\Gamma_2^\tau YY^\tau\Gamma_2)\}, \\
&\quad - (n - m)\log\det\{(\Gamma_2^\tau Y_{(I)}Y_{(I)}^\tau\Gamma_2)^{-1}(\Gamma_2^\tau YY^\tau\Gamma_2)\},
\end{aligned}$$

respectively, where the constant c_i^* $(i = 1, 2)$ does not depend upon the index subset I. In addition, by noting that $\Gamma_1(\Gamma_1^\tau S\Gamma_1)^{-1}\Gamma_1^\tau = X(X^\tau SX)^{-1}X^\tau$ and $\Gamma_1^\tau S_{(I)}^{-1}\Gamma_1 = \Gamma_1^\tau S^{-1}\Gamma_1 + \Gamma_1^\tau S^{-1}E_I(I_k - H_I - E_I^\tau S^{-1}E_I)^{-1}E_I^\tau S^{-1}\Gamma_1$, we establish that

$$\text{(A.8)}\qquad 2K_I(V_{11}) = c_1^* + m(n - k - r) + (n - r)\log\Lambda_I + (n - k - r)\operatorname{tr}\{A_I(I_k - H_I - E_I^\tau S^{-1}E_I)^{-1}\},$$

where A_I is defined in Theorem 3.1. Similarly, for $K_I(V_{22})$, since $\Gamma_2(\Gamma_2^\tau YY^\tau\Gamma_2)^{-1}\Gamma_2^\tau = Q(Q^\tau YY^\tau Q)^{-1}Q^\tau$ $Q \in \mathcal{Q}$, and $(\Gamma_2^\tau Y_{(I)}Y_{(I)}^\tau\Gamma_2)^{-1} = (\Gamma_2^\tau YY^\tau\Gamma_2)^{-1} + (\Gamma_2^\tau YY^\tau\Gamma_2)^{-1}\Gamma_2^\tau Y_I (I_k - \Delta_I^*)^{-1}Y_I^\tau\Gamma_2(\Gamma_2^\tau YY^\tau\Gamma_2)^{-1}$, where $\Delta_I^* = Y_I^\tau Q(Q^\tau YY^\tau Q)^{-1}Q^\tau Y_I$, we have

$$\text{(A.9)}\quad 2K_I(V_{22}) = c_2^* + (n-k-m)(p-m-k) + (n-k-m)\,\mathrm{tr}\{(I_k-\Delta_I^*)^{-1}\} + (n-m)\log\det\{(I_k-\Delta_I^*)\}.$$

Finally, for $K_I(V_{12} \mid V_{11})$, by using the following facts, $\Gamma_2(\Gamma_2^\tau S\Gamma_2)^{-1}\Gamma_2^\tau = Q(Q^\tau SQ)^{-1}Q^\tau = S^{-1} - S^{-1}X(X^\tau S^{-1}X)^{-1}X^\tau S^{-1}$, $(\Gamma_2^\tau S_{(I)}\Gamma_2)^{-1} = (\Gamma_2^\tau S\Gamma_2)^{-1} + (\Gamma_2^\tau S\Gamma_2)^{-1}\Gamma_2^\tau E_I V_I^{-1}E_I^\tau\Gamma_2(\Gamma_2^\tau S\Gamma_2)^{-1}$, and $E_I^\tau Q(Q^\tau SQ)^{-1}Q^\tau E_I = I_k - H_I - V_I$, it can be concluded that $V_{(I)}^* = V^* + \Gamma_1^\tau X(X^\tau S^{-1}X)^{-1}X^\tau S^{-1}E_I V_I^{-1}E_I^\tau\Gamma_2(\Gamma_2^\tau S\Gamma_2)^{-1}$. Therefore, the KLD measurement of $V_{12} \mid V_{11}$ can be written as

$$\begin{aligned}2K_I(V_{12} \mid V_{11}) = &-mk + m\,\mathrm{tr}\{(I_k - H_I)V_I^{-1}\} - m\log\det\{(I_k-H_I)V_I^{-1}\}\\ &+ \mathrm{tr}\{E_I^\tau S^{-1}X(X^\tau S^{-1}X)^{-1}X^\tau\Gamma_1 V_{11}\Gamma_1^\tau X\\ &\quad\cdot (X^\tau S^{-1}X)^{-1}X^\tau S^{-1}E_I V_I^{-1}(I_k - H_I - V_I)V_I^{-1}\}.\end{aligned}$$

Furthermore, since $E_{p(V_{11}|Y_{(I)})}[V_{11}] = (n-k-r)(\Gamma_1^\tau S_{(I)}^{-1}\Gamma_1)$ and $X^\tau\Gamma_1(\Gamma_1^\tau S_{(I)}^{-1}\Gamma_1)\Gamma_1^\tau X = X^\tau S_{(I)}^{-1}X = X^\tau S^{-1}X + X^\tau S^{-1}E_I(I_k - H_I - E_I^\tau S^{-1}E_I)^{-1}E_I^\tau S^{-1}X$, it is calculated that

$$\begin{aligned}\text{(A.10)}\quad &2E_{p(V_{11}|Y_{(I)})}[K_I(V_{12} \mid V_{11})]\\ &= -mk + m\,\mathrm{tr}\{(I_k-H_I)V_I^{-1}\} - m\log\det\{(I_k-H_I)V_I^{-1}\}\\ &\quad + (n-k-r)\,\mathrm{tr}\{A_I(I_k - H_I - E_I^\tau S^{-1}E_I)^{-1}(I_k - H_I - V_I)V_I^{-1}\}.\end{aligned}$$

Substituting (A.8)–(A.10) into (A.6), then (3.12) is obtained and the proof is complete. □

Proof of Theorem 3.3. On the one hand, Lemma 2.2 suggests that the KLD measurement of (B,Σ) can be decomposed into

$$\text{(A.11)}\qquad K_I(B,\Sigma) = K_I(B,\Sigma^{-1}) = K_I(\Sigma^{-1}) + E_{p(\Sigma^{-1}|Y_{(I)})}[K_I(B \mid \Sigma^{-1})],$$

where the KLD measurement $K_I(\Sigma^{-1})$ is given by (3.12). On the other hand, for the MIDM (3.8) with UC, Lemma 3.3 implies $B \mid (\Sigma^{-1}, Y_{(I)}) \sim N_{m,r}(\hat{B}_{\Sigma(I)}, (X^\tau\Sigma^{-1}X)^{-1}, (Z_{(I)}Z_{(I)}^\tau)^{-1})$, where $\hat{B}_{\Sigma(I)} = (X^\tau\Sigma^{-1}X)^{-1}X^\tau\Sigma^{-1}Y_{(I)}Z_{(I)}^\tau(Z_{(I)}Z_{(I)}^\tau)^{-1}$. Using Lemma 2.3 we have $2K_I(B \mid \Sigma^{-1}) = -mr - m\log\det\{(Z_{(I)}Z_{(I)}^\tau)^{-1}(ZZ^\tau)\} + \mathrm{tr}\{(\hat{B}_\Sigma - \hat{B}_{\Sigma(I)})(ZZ^\tau)\ (\hat{B}_\Sigma - \hat{B}_{\Sigma(I)})^\tau(X^\tau\Sigma^{-1}X)\} + m\,\mathrm{tr}\{(Z_{(I)}Z_{(I)}^\tau)^{-1}(ZZ^\tau)\}$. It can be shown that $\hat{B}_{\Sigma(I)} = \hat{B}_\Sigma - (X^\tau\Sigma^{-1}X)^{-1}X^\tau\Sigma^{-1}E_I(I_k - H_I)^{-1}Z_I^\tau(ZZ^\tau)^{-1}$. So, the KLD measurement $K_I(B \mid \Sigma^{-1})$ can be further simplified as

$$\begin{aligned}\text{(A.12)}\quad &2K_I(B \mid \Sigma^{-1})\\ &= -mr + m(r-k) + m\log\det\{(I_k - H_I)\} + m\,\mathrm{tr}\{(I_k-H_I)^{-1}\}\\ &\quad + \mathrm{tr}\{\Sigma^{-1}X(X^\tau\Sigma^{-1}X)^{-1}X^\tau\Sigma^{-1}E_I(I_k-H_I)^{-1}H_I(I_k-H_I)^{-1}E_I^\tau\}.\end{aligned}$$

The remaining task is to calculate the expectation of (A.12) with respect to the posterior $p(\Sigma^{-1} \mid Y_{(I)})$. In the similar manner to the proof in Theorem 3.2, the quantity

$\Sigma^{-1}X(X^{\tau}\Sigma^{-1}X)^{-1}X^{\tau}\Sigma^{-1}$ involved in the last term in (A.12) can be expressed as a function of the transformed dispersion components V_{11}, V_{12} and V_{22}. In this way, the posterior distributions of V_{11}, V_{12} and V_{22} can be used to calculate $E_{p(\Sigma^{-1}|Y_{(I)})}[K_I(B \mid \Sigma^{-1})]$, which can be simplified into

$$[m \log \det\{(I_k - H_I)\} + m\,\mathrm{tr}\{H_I(I_k - H_I - E_I^{\tau}S^{-1}E_I)^{-1}\} + (n-k-m-r)$$
$$\mathrm{tr}\{H_I(I_k - H_I - E_I^{\tau}S^{-1}E_I)^{-1}(A_I - A_I V_I^{-1} A_I)(I_k - H_I - E_I^{\tau}S^{-1}E_I)^{-1}\}]/2.$$

Substituting this fact and (3.12) into (A.11), then (3.13) is obvious and the proof is complete. □

REFERENCES

Ali, M. A. (1990). A Bayesian approach to detect informative observations in an experiment, *Comm. Statist. Theory Methods*, **19**, 2567–2575.

Box, G. E. P. and Tiao, G. C. (1968). A Bayesian approach to some outlier problems, *Biometrika*, **55**, 119–129.

Carlin, B. P. and Polson, N. G. (1991). An expected utility approach to influence diagnostics, *J. Amer. Statist. Assoc.*, **86**, 1013–1021.

Chaloner, K. and Brant, R. (1988). A Bayesian approach to outlier detection and residual analysis, *Biometrika*, **75**, 651–659.

Cook, R. D. (1979). Influential observations in linear regression, *J. Amer. Statist. Assoc.*, **74**, 169–174.

Cook, R. D. (1986). Assessment of local influence (with discussion), *J. Roy. Statist. Soc. Ser. B*, **48**, 133–169.

Dickey, J. M. (1967). Multivariate generalization of multivariate t-distribution, *Ann. Math. Statist.*, **38**, 511–518.

Geisser, S. (1970). Bayesian analysis of growth curves, *Sankhyā Ser. A*, **32**, 53–64.

Geisser, S. (1987). Influential observations, diagnostics and discovery tests, *J. Appl. Statist.*, **14**, 133–142.

Guttman, I. and Peña, D. (1988). Outliers and influence: Evaluation of posteriors of parameters in linear model, *Bayesian Statistics 3* (eds. by J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith), 631-640. Oxford University Press.

Guttman, I. and Peña, D. (1993). A Bayesian look at diagnostics in the univariate linear model, *Statist. Sinica*, **3**, 367–390.

Johnson, W. and Geisser, S. (1983). A predictive view of the detection and characterization of influential observations in regression analysis, *J. Amer. Statist. Assoc.*, **78**, 137–144.

Johnson, W. and Geisser, S. (1985). Estimative influence measures for the multivariate general linear model, *J. Statist. Plann. Inference*, **11**, 33–56.

Khatri, C. G. (1966). A note on a MANOVA model applied to problems in growth curve, *Ann. Inst. Statist. Math.*, **18**, 75–86.

Lee, J. C. (1988). Prediction and estimation of growth curve with special covariance structure, *J. Amer. Statist. Assoc.*, **83**, 432–440.

Lee, J. C. (1991). Tests and model selection for the general growth curve model, *Biometrics*, **47**, 147–159.

Lee, J. C. and Geisser, S. (1975). Growth curve prediction, *Sankhyā Ser. A*, **34**, 393–412.

Liski, E. P. (1991). Detecting influential measurements in a growth curve model, *Biometrics*, **47**, 659–668.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

Okamato, M. (1973). Distinctness of the eigenvalues of a quadratic form in a multivariate sample, *Ann. Statist.*, **1**, 763–765.

Pan, J. X. (1995). multivariate statistical diagnostics with application to the growth curve model, Ph. D. dissertation, Department of Mathematics, Hong Kong Baptist University (unpublished).

Pan, J. X. and Fang, K. T. (1995). Multiple outlier detection in growth curve model with unstructured covariance matrix, *Ann. Inst. Statist. Math.*, **47**, 137–153.

Pan, J. X. and Fang, K. T. (1996). Detecting influential observations in growth curve model with unstructured covariance, *Comput. Statist. Data Anal.*, **22**, 71–87.

Pan, J. X., Fang, K. T. and Liski, E. P. (1996). Bayesian local influence in the growth curve model with Rao's simple covariance structure, *J. Multivariate Anal.*, **58**, 55-81.

Pan, J. X., Fang, K. T. and von Rosen, D. (1998). On the posterior distribution of the covariance matrix of the growth curve model, *Statist. Probab. Lett.*, **38**, 33–39.

Pettit, L. I. and Smith, A. F. M. (1985). Outliers and influential observations in linear model, *Bayesian Statistics 2* (eds. J. M. Bernardo, M. H. DeGoot, D. V. Lindley and A. F. M. Smith), 473–494, Elsevier Science Publish, North-Holland.

Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika*, **51**, 313–326.

Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves, *Biometrika*, **52**, 447–458.

Rao, C. R. (1966). Covariance adjustment and related problems in multivariate analysis, *Multivariate Analysis* (ed. P. Krishnaiah), 87–103, Academic Press, New York.

Rao, C. R. (1987). Prediction of future observations in growth curve models, *Statist. Sci.*, **2**, 434–471.

von Rosen, D. (1989). Maximum likelihood estimates in multivariate linear normal model, *J. Multivariate Anal.*, **31**, 187–200.

von Rosen, D. (1991). The growth curve model: a review, *Comm. Statist. Theory Methods*, **20**, 2791–2822.

von Rosen, D. (1995). Influential observations in multivariate linear models, *Scand. J. Statist.*, **22**, 207–222.