A Diagnostic of Influential Cases Based on the Information Complexity Criteria in Mixed Models

Junfeng Shang
Department of Mathematics and Statistics
Bowling Green State University
Bowling Green, OH 43403, USA

Abstract: In the mixed modeling framework, we develop a diagnostic for detecting influential cases based on the information complexity criteria. The diagnostic compares the information complexity criteria between the full data set and a case-deleted data set. The computational formula of the information complexity criterion is derived using the Fisher information matrix.

Key words: Model selection, ICOMP, Kullback-Leibler discrepancy, case-deletion

1 Introduction

In collecting data, recording errors or anomaly may result in certain amount of influential cases which have a considerable influence on specific inferences. To effectively keep away from such influence on statistical inferences, modeling diagnostics provide measures which discriminate influential cases by gauging their impact on a particular inferential objective.

In the literature, modeling diagnostics have attracted sufficient attention. In linear regression, measures such as COOK's distance, DFBETAS, DFFITS, studentized residuals, and COVRATIO are popularly applied to reveal influential cases which substantially impact the fitted model and associated results (see Belsley, Kuh, and Welsh, 1980; Cook and Weisberg, 1982). The modeling diagnostics have been furthermore developed in the work of Johnson and Geisser (1983), Johnson (1985), Cavanaugh and Johnson (1999), Cavanaugh and Oleson (2001).

Very little work has been published on mixed modeling diagnostics. Christensen, Pearson, and Johnson (1992) proposed case-deletion diagnostics for detecting influential observations in mixed linear models. Cavanaugh and Shang (2005) developed a predictive influence function (PIF) for discovering the influential cases for the prediction of the random effects in a mixed model.

The identification of influential cases involves the problem of model selection since the detected influential cases may be caused by the simplicity of the model. Bozdogan and Bearse (2003) developed a modeling diagnostic using the information complexity (ICOMP) criteria in dynamic multivariate linear models. In their work, influential case detection and model selection have been addressed jointly.

In this paper, we develop a diagnostic for detecting influential cases based on the information complexity criteria in the mixed modeling framework. The diagnostic compares the information complexity criteria between the full data set and a case-deleted data set. The information complexity criterion is computed from the Fisher information matrix.

2 The Information Complexity (ICOMP) Criterion in Mixed Models

2.1 Model

For $i = 1, \ldots, m$, let y_i denote an $n_i \times 1$ vector of responses observed on the ith subject or case, and let b_i denote a $q \times 1$ vector of associated random effects. Assume the vectors b_i are independently distributed as $N(0, D)$. Let $N = \sum_{i=1}^{m} n_i$ denote the total number of response measurements.

The general linear mixed model can be represented as

$$Y = X\beta + Zb + \varepsilon, \tag{2.1}$$

where Y denotes the $N \times 1$ response vector $({y_1}', \ldots, {y_m}')'$; X is an $N \times p$ design matrix of full column rank; Z is an $N \times mq$ block diagonal design matrix comprised of m blocks, where each block is an $n_i \times q$ matrix; β is the $p \times 1$ fixed effects parameter vector; b is the $mq \times 1$ random effects vector $({b_1}', \ldots, {b_m}')'$; and ε is the $N \times 1$ error vector. We assume $b \sim N(0, G)$ and $\varepsilon \sim N(0, \sigma_0^2 R)$, with b and ε distributed independently. Here, R and G are positive definite block diagonal matrices and G is $mq \times mq$

and comprised of m identical blocks, each of which is D.

Let θ denote the unknown parameter vector, consisting of the elements of the vector β, the matrix D, and the scalar σ_0^2. Let $V = ZGZ' + \sigma_0^2 R$. Note that V represents the covariance matrix of Y and that V is positive definite.

Let $\hat{\theta}$ denote the MLE of θ, consisting of the elements of the vector $\hat{\beta}$, the matrix $\hat{D}$, and the scalar $\hat{\sigma}_0^2$. Note that $\hat{G}$ is a positive definite block diagonal matrix and is comprised of m identical blocks, each of which is $\hat{D}$. For a given set of estimates $\hat{\theta}$, the covariance matrix of Y is given by $\hat{V} = Z\hat{G}Z' + \hat{\sigma}_0^2 R$.

2.2 The Information Complexity (ICOMP) Criterion in Mixed Models

Prior to the introduction of the information complexity (ICOMP) criterion, the most recognized model selection criterion, the Akaike Information Criterion (AIC, Akaike, 1973, 1974), is presented and discussed for the purpose of comparison with the ICOMP criterion.

Akaike's (1973) original AIC is given by

$$\text{AIC} = -2\,\log L(\hat{\theta} \mid Y) + 2k,$$

where $L(\hat{\theta} \mid Y)$ is the maximized likelihood function, and k represents the dimension of estimated parameter $\hat{\theta}$ under the given model. Here, the "goodness of fit" term, $-2\,\log L(\hat{\theta} \mid Y)$, gauges how well the model fits the data, and the penalty term, $2k$, measures the complexity that compensates for the bias in the lack of fit when the maximum likelihood estimators are used. The success of AIC depends on its approximation to the bias adjustment by $2k$ for large samples.

Suppose the generating model or the true model, which presumably gave rise to the data. Also, suppose that a candidate or approximating model is a model that could potentially be used to describe the data and a fitted model is a candidate model that has been fit to the data. AIC is justified as an asymptotic unbiased estimator of Kullback-Leibler discrepancy between the generating model and a fitted model. The formation of AIC reflects an underlying principle for model selection criteria, that is, a model selection criterion involves both a goodness of fit term gauging how well the model fits the data and a penalty term measuring the model complexity. AIC penalizes the complexity of model by two times of the number of estimated parameters.

Similar to AIC, the ICOMP criterion combines a goodness-of-fit term with a term for measuring the complexity of model. In what follows, we will see that instead of penalizing the number of estimated parameters, the ICOMP criterion penalizes the covariance complexity of model.

The ICOMP criterion is based on the covariance complexity index of van Emdan (1971) in parametric estimation. The ICOMP criterion is defined as

$$ICOMP = -2\log L(\hat{\theta} \mid Y) + 2\,C(\hat{\Sigma}), \qquad (2.2)$$

where $L(\hat{\theta} \mid Y)$ represents the maximized likelihood function, $\hat{\theta}$ represents the maximum likelihood estimator of the unknown parameter θ, C represents a complexity measure, Σ represents the covariance matrix of the estimated parameters for the model, and correspondingly $\hat{\Sigma}$ represents the estimated covariance matrix of Σ. Note that in original definition of the ICOMP criterion, the $\hat{\theta}$ could be any estimator of θ. In this paper, we utilize the maximum likelihood estimator (MLE) of θ.

Note that the ICOMP criterion and the AIC share the similarity in containing two terms, one is the goodness of fit term, $-2\log L(\hat{\theta} \mid Y)$; the other one is the penalty term. However, the penalty term of AIC is $2k$, two times of the number of estimated parameters, whereas the penalty term of the ICOMP criterion is the measure of the covariance complexity for model.

To evaluate the complexity measure of the ICOMP criterion, Bozdogan (1988, 1990, 1993, 1994) proposed a maximal information complexity measure which is expressed as

$$C_m(\Sigma) = \frac{m_k}{2}\log\frac{\text{tr}(\Sigma)}{m_k} - \frac{1}{2}\log|\Sigma|, \qquad (2.3)$$

where m_k is the dimension of Σ. This measure is optimal in that it is invariant with respect to scaler multiplication and orthonormal transformation and in that it is a monotonically increasing function of the dimension m_k of Σ. (See Bozdogan, 1988, 1990 for details.)

In the mixed model (2.1), recall that $V = ZGZ' + \sigma^2 R$ is the covariance matrix of Y. To generalize model (2.1), Z matrix can be partitioned in r submatrices. Write $Z = [Z_1, \cdots, Z_r]$ and $b' = [b_1', \cdots, b_r']$ with $Cov(b_i) = \sigma_i^2 I_{q(i)}$ and $Cov(b_i, b_j) = 0$. Let $q(i)$ denote the number of columns in Z_i and then $I_{q(i)}$ is a $q(i) \times q(i)$ identity matrix. Therefore, the covariance matrix of b is a block diagonal matrix with blocks $\sigma_i^2 I_{q(i)}$. As usual, assume $R = I_N$, and $R = Z_0 Z_0'$, we then can write

$$V = \sum_{i=0}^{r} \sigma_i^2 Z_i Z_i'.$$

Thus, in model (2.1), the covariance of b can be rewritten as

$$G = \sum_{i=1}^{r} \sigma_i^2 Z_i Z_i'.$$

Now, the unknown parameter vector θ consists of the elements of the vector β and the scalars $\sigma_0^2, \cdots, \sigma_r^2$. Instead of estimating the matrix D, we need to estimate scalars $\sigma_0^2, \cdots, \sigma_r^2$. We have $p + r + 1$ parameters to estimate. Let $\hat{\theta}$ denote the MLE of θ, and $\hat{\theta} = (\hat{\beta}', \hat{\sigma}_0^2, \cdots, \hat{\sigma}_r^2)$. In this paper, we utilize the EM algorithm to estimate the MLE's.

In model (2.1), the covariance matrix of the estimated parameters Σ is unknown in closed form, we therefore employ the estimated inverse-Fisher information matrix to assess the complexity. Let F represents the Fisher information matrix for the model, then let F^{-1} denote the inverse of F. The estimated inverse-Fisher information matrix $\hat{F}^{-1}$ is obtained with $\hat{\theta}$ in place of θ in the matrix F^{-1}.

By the expressions (2.2) and (2.3), we therefore rewrite the ICOMP criterion as

$$\begin{aligned} ICOMP &= -2 \log L(\hat{\theta} \mid Y) \\ &\quad + m_k \log \frac{\operatorname{tr}(\hat{F}^{-1})}{m_k} - \log |\hat{F}^{-1}| \\ &= N \log(2\pi) + \log |\hat{V}| \\ &\quad + (Y - X\hat{\beta})' \hat{V}^{-1} (Y - X\hat{\beta}) \\ &\quad + m_k \log \frac{\operatorname{tr}(\hat{F}^{-1})}{m_k} \\ &\quad - \log |\hat{F}^{-1}|. \qquad (2.4) \end{aligned}$$

For the evaluation of (2.4), apart from the estimation of the MLE's of the parameters, we need to derive the Fisher information matrix F in model (2.1). The Appendix presents the formal derivation of the Fisher information matrix for model (2.1). Corresponding to a class of models which are considered to fit the data, the value of the ICOMP criterion for each model is calculated by the formula (2.4), and the model is chosen under which the ICOMP criterion is minimized.

3 A Diagnostic of Influential Cases Based on the ICOMP Criterion

For the identification of influential cases, the idea of leave-one-out method is typically utilized to develop measures for identifying influential cases. This idea compares inferential quantities such as regression parameter estimates, fitted values, and estimated variances based on a fitted model to the full data set with those based on fitting a model to the data set with a case deleted. For instance, Cook (1977, 1979) effectively applied leave-one-out method and developed numerous measures for detecting influential observations in linear regression modeling framework.

We propose a diagnostic which makes use of the deletion of a case at a time based on the ICOMP criteria. The diagnostic is defined by the discrepancy of the two ICOMP criteria, one is computed based on the full data; the other one is computed based on a case-deleted data set. We define the diagnostic as

$$\delta_{ICOMP}(i) = ICOMP_{\text{Full-Data}} - ICOMP^{(i)}, \quad (3.1)$$

where $ICOMP_{\text{Full-Data}}$ is the ICOMP criterion value for a fitted mixed model when the full data set is utilized; $ICOMP^{(i)}$ is the ICOMP criterion value for the same fitted mixed model when the ith case is deleted.

Straightforwardly, the magnitude of $\delta_{ICOMP}(i)$ reflected on definition (3.1) evaluates the influence of y_i on the ICOMP criterion. We recall that the ICOMP criterion consists of two terms and essentially takes into account of both goodness of fit and model complexity. The magnitude of $\delta_{ICOMP}(i)$ therefore combines the influences of y_i on both goodness of fit and model complexity.

For the evaluated $\delta_{ICOMP}(i)$ values, we need a standard to determine which cases are influential. If a case is not abnormal or not influential, it should be indispensable to the fitting of the model to the data. Once this case is removed, the leave-one-out data will make the model less fit. In this sense, the leave-one-out ICOMP criterion, i.e., $ICOMP^{(i)}$, will grow to be larger than the ICOMP criterion under the full data set. Thus, the $\delta_{ICOMP}(i)$ value is negative. On the contrary, if a case is abnormal or influential, the removal of this case will make the model better fit. Correspondingly, the leave-one-out ICOMP criterion $ICOMP^{(i)}$ shrinks, and the diagnostic value is positive. However, positive diagnostics only indicate that the corresponding cases are potentially influential. When the diagnostic values are positive for some cases, we hope to reveal the most influential cases. As a result, among all the evaluated $\delta_{ICOMP}(i)$ for the cases in a data set, the outstanding positive ones specify the influential cases. Furthermore, under a given model, it is possible that all cases are potentially influential.

Although AIC is similar to the ICOMP criterion for model selection, the analogous criterion based on AIC cannot provide a diagnostic of influential cases as effective as $\delta_{ICOMP}(i)$ because the dimension of estimated parameters is identical for both the full data set and a case-deleted data set, and the difference of the model complexity cannot be measured.

In fact, detecting influential cases in mixed models is quite crucial. For instance, in biostatistical studies, mixed models are appropriate for a various types of repeated measurements where a response variable and a collection of covariates are measured on each subject. However, there exist some subjects which provide extreme response measurements because anomaly occurs when the data are collected. In this situation, the diagnostic proposed here can be applied to this type of data to reveal the influential subjects. Then the data set excluding the measurements from the influential subjects will greatly improve the accuracy of statistical inferences for the data set.

4 An Application

For the illustration of the effectiveness of the diagnostic, we apply the diagnostic to an exam scores data set. This data set can be described by model (2.1) and was utilized in (Cavanaugh and Shang, 2005). In what follows, the applied results demonstrate that the proposed diagnostic can effectively flag the influential cases in the mixed model.

This exam data set consists of 3 exam scores for each of 72 students. For such a data set, it is reasonable to assume that the scores for a student are correlated, and yet sets of scores for different students are uncorrelated. The student effect can be regarded as random and the exam effect as fixed. Thus, the data is amenable to the mixed model (2.1). The purpose of our analysis will be to explore which cases have a substantial impact on the ICOMP criterion for a given mixed model.

In model (2.1) for exam data, specifically we note that $m = 72, n_i = n = 3$, $p = 3$, $q = 1$; X is an $N \times n$ design matrix of full column rank; Z matrix is an $N \times m$ block diagonal design matrix, where each block is an $n \times 1$ vector; and b is the $m \times 1$ random effects vector. We assume that β is the $p \times 1$ fixed effects parameter vector consisting of three population means for exams, μ_1, μ_2, and μ_3. We also assume that $b \sim N(0, \sigma_1^2 I)$ and $\epsilon \sim \sigma_0^2 I)$, and they are mutually independent. Therefore, the unknown parameter vector θ consists of the population means μ_1, μ_2, and μ_3, and two variance components σ_0^2 and σ_1^2. They need to be estimated for the calculation of the $\delta_{ICOMP}(i)$. Based on the parameter estimates in both the full data set and case-deleted data sets, we can evaluate the diagnostic $\delta_{ICOMP}(i)$ using the expressions (2.4) and (3.1).

Figure 1 features the values of $\delta_{ICOMP}(i)$ for exam data. As addressed previously, a positive outstanding $\delta_{ICOMP}(i)$ value indicates the corresponding case is influential. Since the most outstanding values are for cases 55, 71, and 53, these three cases are identified influential and marked with "#" in Figure 1. The observations for each of influential cases are shown in Table 1.

Table 2 features the parameter estimates for the full data set and case-deleted data sets. When one case is omitted from the full data set, the parameter estimates significantly modify. It is noted from Table 2 that quite large differences of the parameter estimates exist between the full data set and case-deleted data sets corresponding to cases 55, 71, and 53. As a result, as demonstrated in both Figure 1 and Table 1, $\delta_{ICOMP}(55)$, $\delta_{ICOMP}(71)$, and $\delta_{ICOMP}(53)$ are quite substantial.

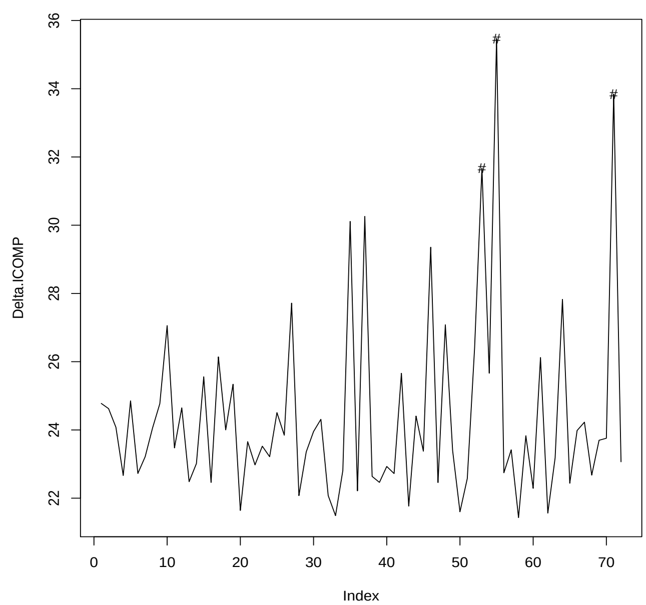

Figure 1: $\delta_{ICOMP(i)}$ vs. Case Index i

Table 1: Influential Cases in Exam Data

Data	Exam1	Exam2	Exam3	$\delta_{ICOMP(i)}$
Case 55	32	86	92	35.47
Case 71	20	69	37	33.84
Case 53	100	99	57	31.67

It is noteworthy that the same exam data set is utilized to illustrate the effectiveness of both δ_{ICOMP} and PIF in Cavanaugh and Shang (2005). The diagnostic δ_{ICOMP} detects that cases 55, 71, and 53

Table 2: Parameter Estimates for Exam Data

Data set	μ_1	μ_2	μ_3	σ_1^2	σ_0^2
Full data	69.25	84.58	73.61	125.22	120.71
(a)	69.77	84.56	73.35	130.37	110.80
(b)	69.94	84.80	74.13	112.03	118.28
(c)	68.81	84.37	73.84	128.44	114.14

Note: (a) =Case 55 deleted, (b)= Case 71 deleted, (c)=Case 53 deleted.

are influential; whereas the diagnostic PIF detects that cases 48, 55, and 71 are influential. To explain these different results, we notice that the diagnostic δ_{ICOMP} proposed here aims to measure the influence of a case on both goodness of fit and model complexity for a given model; the diagnostic PIF proposed in Cavanaugh and Shang (2005) aims to measure the influence of a case on the prediction of the random effects for a given model. Indeed, one diagnostic cannot identify influential cases which have substantial impact on all inferential objectives, it is therefore quite possible that different influential cases are revealed under different diagnostics.

5 Concluding Remarks

We develop a diagnostic for detecting influential cases based on the ICOMP criteria in mixed models. Given the mixed model, we define the diagnostic for revealing influential case as the discrepancy of the ICOMP criteria based on the full data set and a case-deleted data set. Since the covariance matrix of estimated parameters in the mixed modeling framework is unknown, the Fisher information matrix is employed to compute the ICOMP criterion. The proposed diagnostic takes into account of both goodness-of-fit and model complexity.

Appendix: The Fisher Information Matrix in Mixed Models

Here, we presents the derivation of the Fisher information matrix in model (2.1).

For model (2.1), without considering the constant 2π, we have the log-likelihood function

$$\begin{aligned} & \log L(\beta, \sigma_0^2, \cdots, \sigma_r^2 \mid Y) \\ = & -\frac{1}{2}\log|V| \\ & -(Y - X\beta)'V^{-1}(Y - X\beta)/2. \end{aligned}$$

Now, we need to find the matrix F, i.e., the matrix $-E_\theta\{\frac{\partial^2 \log L(\beta,\sigma_0^2,\cdots,\sigma_r^2|Y)}{\partial\theta\partial\theta'}\}$.

To find the first derivatives of the log-likelihood function with respect to the parameters, we have

$$\begin{aligned} \frac{\partial \log L(\cdot)}{\partial\beta} &= X'V^{-1}Y - X'V^{-1}X\beta \\ &= X'V^{-1}(Y - X\beta) \quad \text{and} \end{aligned}$$

$$\frac{\partial \log L(\cdot)}{\partial\sigma_i^2} = -\frac{1}{2}\text{tr}(V^{-1}Z_iZ_i')$$
$$+\frac{1}{2}(Y - X\beta)'V^{-1}Z_iZ_i'V^{-1}(Y - X\beta),$$
$$i = 0, 1, \cdots, r.$$

Based on the first derivatives, we can obtain

$$\begin{aligned} \frac{\partial^2 \log L(\cdot)}{\partial\beta\partial\beta'} &= -X'V^{-1}X, \\ \frac{\partial^2 \log L(\cdot)}{\partial\beta\partial\sigma_i^2} &= (Y - X\beta)'V^{-1}Z_iZ_i'V^{-1}X, \quad \text{and} \\ \frac{\partial \log L(\cdot)}{\partial\sigma_i^2\partial\beta} &= (Y - X\beta)'V^{-1}Z_iZ_i'V^{-1}X, \\ & i = 0, 1, \cdots, r. \end{aligned}$$

To compute $\frac{\partial^2 \log L(\cdot)}{\partial\sigma_i^2\partial\sigma_j^2}$ for $i = 0, 1, \cdots, r$ and $j = 0, 1, \cdots, r$, we first have

$$\begin{aligned} & \frac{\text{tr}(V^{-1}Z_iZ_i')}{\partial\sigma_j^2} \\ = & -\text{tr}(V^{-1}Z_iZ_i'V^{-1}Z_jZ_j') \quad \text{and} \\ & \frac{\partial[(Y - X\beta)'V^{-1}Z_iZ_i'V^{-1}(Y - X\beta)]}{\partial\sigma_j^2} \\ = & -2(Y - X\beta)'V^{-1}Z_iZ_i'V^{-1}Z_jZ_j'V^{-1}(Y - X\beta). \end{aligned}$$

Then we can arrive at

$$\frac{\partial^2 \log L(\cdot)}{\partial\sigma_i^2\partial\sigma_j^2} = \frac{1}{2}\text{tr}(V^{-1}Z_iZ_i'V^{-1}Z_jZ_j')$$
$$-(Y - X\beta)'V^{-1}Z_iZ_i'V^{-1}Z_jZ_j'V^{-1}(Y - X\beta).$$

To obtain the expectations of the second derivatives, we have

$$\begin{aligned} E_\theta\left[\frac{\partial^2 \log L(\cdot)}{\partial\beta\partial\beta'}\right] &= -X'V^{-1}X, \\ E_\theta\left[\frac{\partial^2 \log L(\cdot)}{\partial\beta\partial\sigma_i^2}\right] &= E_\theta\left[(Y - X\beta)'V^{-1}Z_iZ_i'V^{-1}X\right] \\ &= 0, \\ E_\theta\left[\frac{\partial \log L(\cdot)}{\partial\sigma_i^2\partial\beta}\right] &= E_\theta\left[(Y - X\beta)'V^{-1}Z_iZ_i'V^{-1}X\right] \\ &= 0, \quad \text{and} \end{aligned}$$

$$
\begin{aligned}
& E_\theta \left[\frac{\partial^2 \log L(\cdot)}{\partial \sigma_i^2 \partial \sigma_j^2} \right] \\
= & \frac{1}{2}\mathrm{tr}(V^{-1}Z_iZ_i'V^{-1}Z_jZ_j') \\
& -E_\theta \left[(Y - X\beta)'V^{-1}Z_iZ_i'V^{-1}Z_jZ_j'V^{-1}(Y - X\beta) \right] \\
= & \frac{1}{2}\mathrm{tr}(V^{-1}Z_iZ_i'V^{-1}Z_jZ_j') \\
& -\mathrm{tr}(V^{-1}Z_iZ_i'V^{-1}Z_jZ_j') \\
= & -\frac{1}{2}\mathrm{tr}(V^{-1}Z_iZ_i'V^{-1}Z_jZ_j').
\end{aligned}
$$

Thus, we can have

$$
\begin{aligned}
F &= -E_\theta\{ \frac{\partial^2 \log L(\beta, \sigma_0^2, \cdots, \sigma_r^2 | Y)}{\partial\theta\partial\theta'} \} \\
&= \begin{bmatrix} X'V^{-1}X_{(p\times p)} & \mathbf{0}_{(p\times(r+1))} \\ \mathbf{0}_{((r+1)\times p)} & F_{1((r+1)\times(r+1))} \end{bmatrix},
\end{aligned}
$$

where the element of F_1 is $\frac{1}{2}\mathrm{tr}(V^{-1}Z_iZ_i'V^{-1}Z_jZ_j')$ with diagonal elements $i = j$ and non-diagonal elements $i \neq j$.

Hence, the Fisher information matrix F for model (2.1) is established.

References

[1] Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F.Csaki ed. *2nd International Symposium on Information Theory* 267–281. Akademia Kiado, Budapest.

[2] Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **AC-19**, 716–723.

[3] Belsley, D.A., Kuh, E., Welsch, R.E., 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* New York: John Wiley and Sons.

[4] Bozdogan, H., 1988. ICOMP: A new model selection criterion. *In: H.H. Bock(Ed.), Classification and Related Methods of Data Analysis, North-Holland, Amsterdam,* 599–608.

[5] Bozdogan, H., 1990. On the information-based measure of covariance complexity and its application on the evaluation of multivariate linear models. *Communications in Statistics, Theory Methods* **19(1)**, 221–278.

[6] Bozdogan, H., 1993. Choosing the number of component clusters in the mixture model using a new informational complexity criterion of the inverse Fisher information matrix. *In: O.Opitz, B. Lausen, and R.Klar (Eds.), Studies in Classification, Data Analysis, and Knowledge Organization,* Springer, Heidelberg, pp. 40–54.

[7] Bozdogan, H., 1994. Mixture-model cluster analysis using a new informational complexity and model selection criteria. *In: H. Bozdogan (Ed.), Multivariate Statistical Modeling,* Vol 2, Proc. 1st US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach, Kluwer Academic Publishers, the Netherlands, Dordrecht, pp. 69–113.

[8] Bozdogan, H., Bearse, P., 2003. Information complexity criteria for detecting influential observations in dynamic multivariate linear models using the genetic algorithm. *Journal of Statistical Planning and Inference* **114**, 31–44.

[9] Cavanaugh, J.E., Johnson, W.O. 1999. Assessing the predictive influence of cases in a state-space process. *Biometrika* **86**, 183–190.

[10] Cavanaugh, J.E., Oleson, J.J., 2001. A diagnostic for assessing the influence of cases on the prediction of missing data. *Journal of the Royal Statistical Society, Series D* **50**, 427–440.

[11] Cavanaugh, J.E., Shang, J., 2005. A diagnostic for assessing the influence of cases on the prediction of random effects in a mixed model. *Journal of Data Science* **3**, 137–151.

[12] Christensen, R., Pearson, L.M., Johnson, W., 1992. Case-deletion diagnostics for mixed models. *Technometrics* **34**, 38–45.

[13] Cook, R.D., 1977. Detection of influential observations in linear regression. *Technometrics* **19**, 15–18.

[14] Cook, R.D., 1979. Influential observations in linear regression. *Journal of the American Statistical Association* **74**, 169–174.

[15] Cook, R.D., Weisberg, S., 1982. *Residuals and Influence in Regression.* London: Chapman and Hall.

[16] Johnson, W., 1985. Influence measures for logistic regression: Another point of view. *Biometrika* **72**, 59–65.

[17] Johnson, W., Geisser, S., 1983. A predictive view of the detection and characterization of influential observations in regression analysis. *Journal of the American Statistical Association* **78**, 137–144.

[18] van Emden, M., 1971. An analysis of complexity. Mathematisch Centrum Tracts 35, Mathematisch Centrum, Amsterdam.