Ann. Inst. Statist. Math.
Vol. 47, No. 1, 137–153 (1995)

MULTIPLE OUTLIER DETECTION IN GROWTH CURVE MODEL WITH UNSTRUCTURED COVARIANCE MATRIX

JIAN-XIN PAN[1] * AND KAI-TAI FANG[2] **

[1] *Department of Mathematics, Hong Kong Baptist University, 224 Waterloo Road, Kowloon, Hong Kong and Department of Statistics, Yunnan University, Kunming 650091, China*
[2] *Department of Mathematics, Hong Kong Baptist University, 224 Waterloo Road, Kowloon, Hong Kong and Institute of Applied Mathematics, Academia Sinica, Beijing 100080, China*

(Received January 31, 1994; revised June 21, 1994)

Abstract. Under a normal assumption, Liski (1991, *Biometrics*, **47**, 659–668) gave some measurements for assessing influential observations in a Growth Curve Model (GCM) with a known covariance. For the GCM with an arbitrary (p.d.) covariance structure, known as unstructured covariance matrix (UCM), the problems of detecting multiple outliers are discussed in this paper. When a multivariate normal error is assumed, the MLEs of the parameters in the Multiple-Individual-Deletion model (MIDM) and the Mean-Shift-Regression model (MSRM) are derived, respectively. In order to detect multiple outliers in the GCM with UCM, the likelihood ratio testing statistic in MSRM is established and its null distribution is derived. For illustration, two numerical examples are discussed, which shows that the criteria presented in this paper are useful in practice.

Key words and phrases: Elliptically contoured distribution, growth curve model, influential observation, multiple outlier detection criterion, statistical diagnostic.

1. Introduction

The *growth curve model* (GCM) is a generalized multivariate model of variance analysis, which is useful especially for growths of animals and plants so that it is applied extensively to biostatistics, medical research and epidemiology. It was first proposed by Potthoff and Roy (1964) and then subsequently considered by many authors, including Rao (1965, 1966, 1967), Khatri (1966), Geisser (1970) and von Rosen (1989, 1990, 1991).

* Supported partially by the WAI TAK Investment and Loan Company Ltd. Research Scholarship of Hong Kong for 1992–93.

** Supported partially by the Hong Kong UPGC Grant.

Consider a GCM:

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

where X and Z are known design matrices of rank $m < p$ and $r < n$, respectively, and the regression coefficient B is unknown. Furthermore, the columns of the error matrix E are independent p-variate normal with a mean vector $\mathbf{0}$ and a common unknown covariance matrix $\Sigma > \mathbf{0}$, i.e., $Y \sim N_{p,n}(XBZ, \Sigma \otimes I_n)$, where the notation $\otimes$ denotes the Kronecker product of matrices.

Under the normal assumption, Rao (1965, 1966), Khatri (1966) and von Rosen (1989), using different methods, obtained the maximum likelihood estimates (MLEs) of the parameters B and Σ as follows

$$\hat{B} = (X^\tau S^{-1} X)^{-1} X^\tau S^{-1} Y Z^\tau (Z Z^\tau)^{-1} \tag{1.2}$$

and

$$\hat{\Sigma} = \frac{1}{n}(Y - X\hat{B}Z)(Y - X\hat{B}Z)^\tau = \frac{1}{n}(S + Q_S Y P_{Z^\tau} Y^\tau Q_S^\tau), \tag{1.3}$$

where $Q_S = SQ(Q^\tau SQ)^{-1}Q^\tau$, $S = Y(I_n - P_{Z^\tau})Y^\tau$ and $Q \in \mathcal{Q}$, a set of matrices defined as

$$\mathcal{Q} = \{Q \mid Q : p \times (p-m), \mathrm{rank}(Q) = p - m \text{ and } X^\tau Q = \mathbf{0}\}. \tag{1.4}$$

Throughout this paper $P_A = A(A^\tau A)^{-1}A^\tau$ denotes the projection matrix of A on condition that $A^\tau A$ is nonsingular. The matrix S is positive definite with probability one as long as $n > p + r$ (Okamato (1973)). Furthermore, von Rosen (1990) discussed some asymptotic properties of the MLEs (1.2) and (1.3). He also derived some useful formulae for higher moments of $\hat{B}$ and $\hat{\Sigma}$ (see von Rosen (1990, 1991)).

On the statistical diagnostic, a number of papers and books have been published dealing with the problems of detecting outliers and influential observations, especially in an ordinary regression analysis (ORA). Based on the empirical influence function of the regression coefficient, Cook (1977) introduced a statistic to investigate the influence of an observation on the regression fit. Since then many measurements in a certain sense have been proposed to identify whether or not a subset of the observations is an outlier or influential set. According to Cook's definition, an observation can be judged to be influential if some important features of the analysis are substantially altered when it is deleted from the data set. The so-called outliers are the observations that do not follow the pattern of the majority of the data. The problems of detecting multiple outliers and influential observations in the GCM, however, are more complicated than those in the ORA, and few works on this subject have been developed. Recently, Liski (1991) has presented some methods of detecting outliers and influential observations in the GCM with a covariance matrix $\Sigma = \sigma^2 G$, where G is a known positive definite (p.d.) matrix ($G > \mathbf{0}$) and $\sigma^2 > 0$ is an unknown scalar. When $\Sigma > \mathbf{0}$ is an unknown arbitrary covariance matrix, known as *unstructured covariance matrix* (UCM), the problems of effectively detecting outliers and influential observations in the GCM become more difficult because the MLE $\hat{\boldsymbol{B}}$ given by (1.2) is a nonlinear function of the response matrix $\boldsymbol{Y}$, as pointed out by Liski (1991). However, the problem with an unknown UCM is the most usual case in practice and should be investigated completely. In this paper, a solution to this important case is presented.

In the next section, a convenient formula for the empirical influence function of $\hat{\boldsymbol{B}}$ is established. Also the relationship between the MLEs of the parameters in the *Multiple-Individual-Deletion model* (MIDM) and the *Mean-Shift-Regression model* (MSRM) is investigated. This relation implies that MIDM is equivalent to MSRM in the sense of the MLE of the regression coefficient. This conclusion, however, does not hold for the MLE of the covariance matrix $\boldsymbol{\Sigma}$. In order to detect multiple outliers in the GCM with UCM, the likelihood ratio testing statistic for MSRM is established and its exact null distribution is derived in Section 3. For illustration, the Dental Data (Potthoff and Roy (1964)) and the Mouse Data (Rao (1984)) are analyzed in the last section, which shows that the presented criteria are useful in practice.

2. Multiple-Individual-Deletion and Mean-Shift-Regression

In this section the MLEs of the parameters in MIDM and MSRM are derived. Also, the relationships of the MLEs for the GCM, MIDM and MSRM are investigated.

2.1 *MLEs of $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ for MIDM*

Let $I = \{i_1, i_2, \ldots, i_k\}$ $(n > p + k)$ be a set containing the indices of the k individuals to be deleted, where the number k is given. Without loss of generality, the index set can be assumed to be $I = \{n-k+1, n-k+2, \ldots, n\}$ so that $\boldsymbol{Y}$ can be partitioned as $\boldsymbol{Y} = (\boldsymbol{Y}_{(I)}, \boldsymbol{Y}_I)$. The matrices $\boldsymbol{Z}$ and $\boldsymbol{E}$ are partitioned as $\boldsymbol{Z} = (\boldsymbol{Z}_{(I)}, \boldsymbol{Z}_I)$ and $\boldsymbol{E} = (\boldsymbol{E}_{(I)}, \boldsymbol{E}_I)$, respectively. Thus, the GCM (1.1) after deleting $\boldsymbol{Y}_I$ becomes

$$\boldsymbol{Y}_{(I)} = \boldsymbol{X}\boldsymbol{B}\boldsymbol{Z}_{(I)} + \boldsymbol{E}_{(I)}, \tag{2.1}$$

which is known as the Multiple-Individual-Deletion model (MIDM), where $\boldsymbol{E}_{(I)} \sim N_{p,n-k}(\boldsymbol{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_{n-k})$. Similar to (1.2) and (1.3), the MLEs of $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ for MIDM (2.1) are

$$\hat{\boldsymbol{B}}_{(I)} = (\boldsymbol{X}^\tau \boldsymbol{S}_{(I)}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\tau \boldsymbol{S}_{(I)}^{-1} \boldsymbol{Y}_{(I)} \boldsymbol{Z}_{(I)}^\tau (\boldsymbol{Z}_{(I)} \boldsymbol{Z}_{(I)}^\tau)^{-1} \tag{2.2}$$

and

$$\hat{\boldsymbol{\Sigma}}_{(I)} = \frac{1}{n-k}(\boldsymbol{S}_{(I)} + \boldsymbol{Q}_{\boldsymbol{S}_{(I)}} \boldsymbol{Y}_{(I)} \boldsymbol{P}_{\boldsymbol{Z}_{(I)}^\tau} \boldsymbol{Y}_{(I)}^\tau \boldsymbol{Q}_{\boldsymbol{S}_{(I)}}^\tau), \tag{2.3}$$

respectively, where $\boldsymbol{S}_{(I)} = \boldsymbol{Y}_{(I)}(\boldsymbol{I}_{n-k} - \boldsymbol{P}_{\boldsymbol{Z}_{(I)}^\tau})\boldsymbol{Y}_{(I)}^\tau$. Throughout this paper we assume $n > r + p + k$ so that $\boldsymbol{S}_{(I)}$ is positive definite with probability one.

In order to derive the empirical influence function of $\hat{\boldsymbol{B}}$, i.e., the difference between $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{B}}_{(I)}$, we need first to find out the empirical influence function of

$\boldsymbol{S}$. This was previously derived in the literature (see, e.g. Chatterjee and Hadi (1988), pp. 42–46).

LEMMA 2.1. *The relationship between* $\boldsymbol{S}$ *and* $\boldsymbol{S}_{(I)}$ *is given by*

$$\boldsymbol{S}_{(I)} = \boldsymbol{S} - \boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}\boldsymbol{e}_I^{\tau} \tag{2.4}$$

where $\boldsymbol{H}_I = \boldsymbol{Z}_I^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1}\boldsymbol{Z}_I$ *and* $\boldsymbol{e} = \boldsymbol{Y}(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Z}^{\tau}}) = (\boldsymbol{e}_{(I)}, \boldsymbol{e}_I)$ *is the residual of* $\boldsymbol{Y}$ *regressed on* $\boldsymbol{Z}$. *Therefore, we have*

$$\boldsymbol{S}_{(I)}^{-1} = \boldsymbol{S}^{-1} + \boldsymbol{S}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1}\boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}. \tag{2.5}$$

With help of Lemma 2.1 the empirical influence function of the MLE of $\boldsymbol{B}$ can be derived as follows.

THEOREM 2.1. *The relationship between the MLEs* $\hat{\boldsymbol{B}}$ *and* $\hat{\boldsymbol{B}}_{(I)}$ *is given by*

$$\hat{\boldsymbol{B}}_{(I)} = \hat{\boldsymbol{B}} - (\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I\boldsymbol{V}_I^{-1}\boldsymbol{K}_I^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1}, \tag{2.6}$$

where $\boldsymbol{V}_I = \boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I + \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X}(\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I$ *and* $\boldsymbol{K}_I = \boldsymbol{Z}_I - \boldsymbol{Z}\boldsymbol{Y}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I + \boldsymbol{Z}\boldsymbol{Y}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X}(\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I$.

PROOF. On the one hand, it is easy to show

$$\boldsymbol{Y}_{(I)}\boldsymbol{Z}_{(I)}^{\tau}(\boldsymbol{Z}_{(I)}\boldsymbol{Z}_{(I)}^{\tau})^{-1} = \boldsymbol{Y}\boldsymbol{Z}^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1} - \boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}\boldsymbol{Z}_I^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1} \tag{2.7}$$

(see, e.g. Cook and Weisberg (1982), pp. 135–137). By using (2.5) and (2.7) we have

$$\begin{aligned}
&\boldsymbol{X}^{\tau}\boldsymbol{S}_{(I)}^{-1}\boldsymbol{Y}_{(I)}\boldsymbol{Z}_{(I)}^{\tau}(\boldsymbol{Z}_{(I)}\boldsymbol{Z}_{(I)}^{\tau})^{-1} \\
&= \{\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1} + \boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1}\boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\} \\
&\quad \cdot \{\boldsymbol{Y}\boldsymbol{Z}^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1} - \boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}\boldsymbol{Z}_I^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1}\} \\
&= \boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{Y}\boldsymbol{Z}^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1} \\
&\quad + \boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1}\boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{Y}\boldsymbol{Z}^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1} \\
&\quad - \boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}\boldsymbol{Z}_I^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1} \\
&\quad - \boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1} \\
&\quad \cdot \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}\boldsymbol{Z}_I^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1} \\
&= \boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{Y}\boldsymbol{Z}^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1} \\
&\quad + \boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1}\boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{Y}\boldsymbol{Z}^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1} \\
&\quad - \boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1}\boldsymbol{Z}_I^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1} \\
&= \boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{Y}\boldsymbol{Z}^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1} \\
&\quad - \boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1} \\
&\quad \cdot (\boldsymbol{Z}_I - \boldsymbol{Z}\boldsymbol{Y}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1}.
\end{aligned} \tag{2.8}$$

On the other hand, from (2.5) we have

$$(2.9)\qquad \begin{aligned}(\boldsymbol{X}^{\tau}\boldsymbol{S}_{(I)}^{-1}\boldsymbol{X})^{-1} = (\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1} - (\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I \\ \cdot (\boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I \\ + \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X}(\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1} \\ \cdot \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X}(\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}.\end{aligned}$$

By using (2.8) and (2.9), we obtain that

$$(2.10)\quad \begin{aligned}\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{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} \\ &\quad - (\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I \\ &\quad \cdot (\boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I + \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X}(\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1} \\ &\quad \cdot \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X}(\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} \\ &\quad - (\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1} \\ &\quad \cdot (\boldsymbol{Z}_I - \boldsymbol{Z}\boldsymbol{Y}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1} + (\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I \\ &\quad \cdot (\boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I + \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X}(\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1} \\ &\quad \cdot \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X}(\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1} \\ &\quad \cdot (\boldsymbol{Z}_I - \boldsymbol{Z}\boldsymbol{Y}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1}.\end{aligned}$$

Noting that the last term of (2.10) can be expressed as

$$\begin{aligned}&(\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1} \\ &\quad \cdot (\boldsymbol{Z}_I - \boldsymbol{Z}\boldsymbol{Y}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1} - (\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I \\ &\quad \cdot (\boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I + \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X}(\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1} \\ &\quad \cdot (\boldsymbol{Z}_I - \boldsymbol{Z}\boldsymbol{Y}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1},\end{aligned}$$

the MLE $\hat{\boldsymbol{B}}_{(I)}$ can be simplified as

$$\begin{aligned}\hat{\boldsymbol{B}}_{(I)} &= \hat{\boldsymbol{B}} - (\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I \\ &\quad \cdot (\boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I + \boldsymbol{e}_I^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X}(\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1} \\ &\quad \cdot (\boldsymbol{Z}_I - \boldsymbol{Z}\boldsymbol{Y}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I + \boldsymbol{Z}\boldsymbol{Y}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X}(\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1} \\ &\equiv \hat{\boldsymbol{B}} - (\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{e}_I\boldsymbol{V}_I^{-1}\boldsymbol{K}_I^{\tau}(\boldsymbol{Z}\boldsymbol{Z}^{\tau})^{-1},\end{aligned}$$

and the proof is complete. □

Remark 2.1. By using the formula

$$(2.11)\qquad \begin{aligned}&\boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{X}(\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}^{-1} \\ &\quad = \boldsymbol{Q}(\boldsymbol{Q}^{\tau}\boldsymbol{S}\boldsymbol{Q})^{-1}\boldsymbol{Q}^{\tau} = \boldsymbol{S}^{-1}\boldsymbol{Q}_S = \boldsymbol{Q}_S^{\tau}\boldsymbol{S}^{-1} = \boldsymbol{Q}_S^{\tau}\boldsymbol{S}^{-1}\boldsymbol{Q}_S\end{aligned}$$

where $\boldsymbol{Q} \in \mathcal{Q}$ (see, e.g., von Rosen (1990)), we obtain

$$(2.12)\quad \boldsymbol{V}_I \equiv \boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^{\tau}\boldsymbol{Q}_S^{\tau}\boldsymbol{S}^{-1}\boldsymbol{Q}_S\boldsymbol{e}_I \quad \text{and} \quad \boldsymbol{K}_I \equiv \boldsymbol{Z}_I - \boldsymbol{Z}\boldsymbol{Y}^{\tau}\boldsymbol{Q}_S^{\tau}\boldsymbol{S}^{-1}\boldsymbol{Q}_S\boldsymbol{e}_I.$$

These are other simplified forms of the matrices $\boldsymbol{V}_I$ and $\boldsymbol{K}_I$. According to (2.11) and (2.12), however, both $\boldsymbol{V}_I$ and $\boldsymbol{K}_I$ do not depend upon the choice of the matrix $\boldsymbol{Q}$ in the set $\mathcal{Q}$.

2.2 *MLEs of $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ for MSRM*

Consider the following Mean-Shift-Regression model (MSRM):

$$\boldsymbol{Y}_{p\times n} = \boldsymbol{X}_{p\times m}\boldsymbol{B}_{m\times r}\boldsymbol{Z}_{r\times n} + \boldsymbol{X}_{p\times m}\boldsymbol{\Phi}_{m\times k}\boldsymbol{D}_{k\times n} + \boldsymbol{E}_{p\times n} \tag{2.13}$$

where $\boldsymbol{E} \sim N_{p,n}(0, \boldsymbol{\Sigma}\otimes\boldsymbol{I}_n)$, $\boldsymbol{\Phi}$ is a mean shift parameter and $\boldsymbol{D} = (\boldsymbol{d}_{n-k+1}, \boldsymbol{d}_{n-k+2}, \ldots, \boldsymbol{d}_n)^\tau$ is a matrix of indicator variables, that is $\boldsymbol{d}_i$, the i-th column of $\boldsymbol{D}^\tau$, is a n-variate vector whose i-th element is one and others equal to zero $(n-k+1 \le i \le n)$. Obviously, $\boldsymbol{Z}\boldsymbol{D}^\tau = \boldsymbol{Z}_I$, $\boldsymbol{Y}\boldsymbol{D}^\tau = \boldsymbol{Y}_I$ and $\boldsymbol{D}\boldsymbol{D}^\tau = \boldsymbol{I}_k$.

THEOREM 2.2. *The MLEs of $\boldsymbol{B}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ for MSRM are, respectively,*

$$\begin{aligned} \hat{\boldsymbol{B}}_a &= \hat{\boldsymbol{B}}_{(I)}, \\ \hat{\boldsymbol{\Phi}} &= (\boldsymbol{X}^\tau \boldsymbol{S}_{(I)}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\tau \boldsymbol{S}_{(I)}^{-1} \boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}, \\ \hat{\boldsymbol{\Sigma}}_a &= \frac{1}{n}\cdot\{(n-k)\hat{\boldsymbol{\Sigma}}_{(I)} + \boldsymbol{Q}_{\boldsymbol{S}_{(I)}}\boldsymbol{Y}_I\boldsymbol{Y}_I^\tau \boldsymbol{Q}_{\boldsymbol{S}_{(I)}}^\tau\}. \end{aligned} \tag{2.14}$$

PROOF. Let $\tilde{\boldsymbol{B}} = (\boldsymbol{B}, \boldsymbol{\Phi})$ and $\tilde{\boldsymbol{Z}}^\tau = (\boldsymbol{Z}^\tau, \boldsymbol{D}^\tau)$, then $\boldsymbol{Y} \sim N_{p,n}(\boldsymbol{X}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{Z}}, \boldsymbol{\Sigma}\otimes \boldsymbol{I}_n)$ and the MLEs of $\tilde{\boldsymbol{B}}$ and $\boldsymbol{\Sigma}$ are

$$\hat{\tilde{\boldsymbol{B}}} = (\boldsymbol{X}^\tau \boldsymbol{S}_a^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\tau \boldsymbol{S}_a^{-1}\boldsymbol{Y}\tilde{\boldsymbol{Z}}^\tau(\tilde{\boldsymbol{Z}}\tilde{\boldsymbol{Z}}^\tau)^{-1} \tag{2.15}$$

and

$$\hat{\boldsymbol{\Sigma}}_a = \frac{1}{n}\cdot\{\boldsymbol{S}_a + \boldsymbol{Q}_{\boldsymbol{S}_a}\boldsymbol{Y}\boldsymbol{P}_{\tilde{\boldsymbol{Z}}^\tau}\boldsymbol{Y}^\tau\boldsymbol{Q}_{\boldsymbol{S}_a}^\tau\}, \tag{2.16}$$

respectively, where $\boldsymbol{S}_a = \boldsymbol{Y}(\boldsymbol{I}_n - \boldsymbol{P}_{\tilde{\boldsymbol{Z}}^\tau})\boldsymbol{Y}^\tau$. Partition $\hat{\tilde{\boldsymbol{B}}}$ into $\hat{\tilde{\boldsymbol{B}}} = (\hat{\boldsymbol{B}}_a, \hat{\boldsymbol{\Phi}})$, where $\hat{\boldsymbol{B}}_a$ and $\hat{\boldsymbol{\Phi}}$ are the MLEs of $\boldsymbol{B}$ and $\boldsymbol{\Phi}$ for MSRM, respectively. Since $\boldsymbol{P}_{\tilde{\boldsymbol{Z}}^\tau} = \boldsymbol{P}_{\boldsymbol{Z}^\tau} + \boldsymbol{P}_{(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Z}^\tau})\boldsymbol{D}^\tau}$, so that

$$\begin{aligned} \boldsymbol{S}_a &= \boldsymbol{Y}(\boldsymbol{I}_n - \boldsymbol{P}_{\tilde{\boldsymbol{Z}}^\tau})\boldsymbol{Y}^\tau \\ &= \boldsymbol{Y}(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Z}^\tau})\boldsymbol{Y}^\tau - \boldsymbol{Y}\boldsymbol{P}_{(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Z}^\tau})\boldsymbol{D}^\tau}\boldsymbol{Y}^\tau \\ &= \boldsymbol{S} - \boldsymbol{Y}(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Z}^\tau})\boldsymbol{D}^\tau\{\boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Z}^\tau})\boldsymbol{D}^\tau\}^{-1}\boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Z}^\tau})\boldsymbol{Y}^\tau \\ &= \boldsymbol{S} - \boldsymbol{Y}(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Z}^\tau})\boldsymbol{D}^\tau(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}\boldsymbol{D}(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Z}^\tau})\boldsymbol{Y}^\tau \\ &= \boldsymbol{S} - \boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}\boldsymbol{e}_I^\tau \\ &= \boldsymbol{S}_{(I)}. \end{aligned} \tag{2.17}$$

On the other hand, because

$$(\tilde{\boldsymbol{Z}}\tilde{\boldsymbol{Z}}^\tau)^{-1} = \begin{pmatrix} (\boldsymbol{Z}_{(I)}\boldsymbol{Z}_{(I)}^\tau)^{-1} & -(\boldsymbol{Z}\boldsymbol{Z}^\tau)^{-1}\boldsymbol{Z}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1} \\ -\boldsymbol{Z}_I^\tau(\boldsymbol{Z}_{(I)}\boldsymbol{Z}_{(I)}^\tau)^{-1} & (\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1} \end{pmatrix}$$

and

$$\boldsymbol{Y}\tilde{\boldsymbol{Z}}^\tau(\tilde{\boldsymbol{Z}}\tilde{\boldsymbol{Z}}^\tau)^{-1} = (\boldsymbol{Y}_{(I)}\boldsymbol{Z}_{(I)}^\tau(\boldsymbol{Z}_{(I)}\boldsymbol{Z}_{(I)}^\tau)^{-1}, \boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}),$$

the MLE $\hat{\hat{\boldsymbol{B}}}$ in (2.15) can be written as

$$\begin{aligned}\hat{\hat{\boldsymbol{B}}} &= (\boldsymbol{X}^\tau \boldsymbol{S}_{(I)}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\tau \boldsymbol{S}_{(I)}^{-1}(\boldsymbol{Y}_{(I)}\boldsymbol{Z}_{(I)}^\tau(\boldsymbol{Z}_{(I)}\boldsymbol{Z}_{(I)}^\tau)^{-1}, \boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}) \\ &= (\hat{\boldsymbol{B}}_{(I)}, (\boldsymbol{X}^\tau \boldsymbol{S}_{(I)}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\tau \boldsymbol{S}_{(I)}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}),\end{aligned}$$

which implies that $\hat{\boldsymbol{B}}_a = \hat{\boldsymbol{B}}_{(I)}$ and $\hat{\boldsymbol{\Phi}} = (\boldsymbol{X}^\tau \boldsymbol{S}_{(I)}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\tau \boldsymbol{S}_{(I)}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}$. In order to derive the relationship between $\hat{\boldsymbol{\Sigma}}_a$ in (2.16) and $\hat{\boldsymbol{\Sigma}}_{(I)}$ in (2.3), we notice that $\boldsymbol{S}_a = \boldsymbol{S}_{(I)}$ results in $\boldsymbol{Y}\boldsymbol{P}_{\tilde{\boldsymbol{Z}}^\tau}\boldsymbol{Y}^\tau = \boldsymbol{Y}_{(I)}\boldsymbol{P}_{\boldsymbol{Z}_{(I)}^\tau}\boldsymbol{Y}_{(I)}^\tau + \boldsymbol{Y}_I\boldsymbol{Y}_I^\tau$. Therefore,

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_a &= \frac{1}{n}(\boldsymbol{S}_{(I)} + \boldsymbol{Q}_{\boldsymbol{S}_{(I)}}\boldsymbol{Y}\boldsymbol{P}_{\tilde{\boldsymbol{Z}}^\tau}\boldsymbol{Y}^\tau\boldsymbol{Q}_{\boldsymbol{S}_{(I)}}^\tau) \\ &= \frac{1}{n}(\boldsymbol{S}_{(I)} + \boldsymbol{Q}_{\boldsymbol{S}_{(I)}}\boldsymbol{Y}_{(I)}\boldsymbol{P}_{\boldsymbol{Z}_{(I)}^\tau}\boldsymbol{Y}_{(I)}^\tau\boldsymbol{Q}_{\boldsymbol{S}_{(I)}}^\tau + \boldsymbol{Q}_{\boldsymbol{S}_{(I)}}\boldsymbol{Y}_I\boldsymbol{Y}_I^\tau\boldsymbol{Q}_{\boldsymbol{S}_{(I)}}^\tau) \\ &= \frac{n-k}{n}\hat{\boldsymbol{\Sigma}}_{(I)} + \frac{1}{n}\boldsymbol{Q}_{\boldsymbol{S}_{(I)}}\boldsymbol{Y}_I\boldsymbol{Y}_I^\tau\boldsymbol{Q}_{\boldsymbol{S}_{(I)}}^\tau\end{aligned}$$

and the proof is complete. □

Theorem 2.2 implies that for a GCM with UCM, the MLE of the regression coefficient in MIDM is the same as that in MSRM, which is coincident with the corresponding fact in the ORA (see, e.g., Cook and Weisberg (1982)). This conclusion, however, does not hold for the MLE of the covariance parameter.

Theorems 2.1 and 2.2 establish the relationships among the MLEs $\hat{\boldsymbol{B}}$, $\hat{\boldsymbol{B}}_{(I)}$ and $\hat{\boldsymbol{B}}_a$. Although the relation of the MLEs of $\boldsymbol{\Sigma}$ in MIDM and MSRM is derived, the relationship between $\hat{\boldsymbol{\Sigma}}_a$ (or $\hat{\boldsymbol{\Sigma}}_{(I)}$) and $\hat{\boldsymbol{\Sigma}}$ is not clear. The ratio of the determinant of $\hat{\boldsymbol{\Sigma}}_a$ to that of $\hat{\boldsymbol{\Sigma}}$, however, can be established, which is extremely useful in multiple outlier detection.

THEOREM 2.3. *The relationship between the determinants of* $\hat{\boldsymbol{\Sigma}}_a$ *and* $\hat{\boldsymbol{\Sigma}}$ *is given by*

$$\begin{aligned}T_I &\equiv \det(\hat{\boldsymbol{\Sigma}})/\det(\hat{\boldsymbol{\Sigma}}_a) \\ &= \det\{\boldsymbol{I}_k + \boldsymbol{e}_I^\tau\boldsymbol{S}^{-1}\boldsymbol{X}(\boldsymbol{X}^\tau\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\tau\boldsymbol{S}^{-1}\boldsymbol{e}_I \\ &\qquad \cdot (\boldsymbol{I}_k - \boldsymbol{H}_I - \boldsymbol{e}_I^\tau\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1}\}\end{aligned} \tag{2.18}$$

or

$$\begin{aligned}\Lambda_I &\equiv \det(\hat{\boldsymbol{\Sigma}}_a)/\det(\hat{\boldsymbol{\Sigma}}) \\ &= \det\{\boldsymbol{I}_k - \boldsymbol{e}_I^\tau\boldsymbol{S}_{(I)}^{-1}\boldsymbol{X}(\boldsymbol{X}^\tau\boldsymbol{S}_{(I)}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\tau\boldsymbol{S}_{(I)}^{-1}\boldsymbol{e}_I \\ &\qquad \cdot (\boldsymbol{I}_k - \boldsymbol{H}_I + \boldsymbol{e}_I^\tau\boldsymbol{S}_{(I)}^{-1}\boldsymbol{e}_I)^{-1}\}.\end{aligned} \tag{2.19}$$

PROOF. From (1.3), noting (2.11) and the definition of $\boldsymbol{S}$, we know that

$$
\begin{aligned}
(2.20)\quad \det(\hat{\Sigma}) &= \frac{1}{n^p}\cdot\det\{\boldsymbol{S}+\boldsymbol{Q}_S\boldsymbol{Y}\boldsymbol{P}_{Z^\tau}\boldsymbol{Y}^\tau\boldsymbol{Q}_S^\tau\}\\
&= \frac{1}{n^p}\cdot\det(\boldsymbol{S})\cdot\det\{\boldsymbol{I}_p+\boldsymbol{Q}_S^\tau\boldsymbol{S}^{-1}\boldsymbol{Q}_S\cdot\boldsymbol{Y}\boldsymbol{P}_{Z^\tau}\boldsymbol{Y}^\tau\}\\
&= \frac{1}{n^p}\cdot\det(\boldsymbol{S})\cdot\det\{\boldsymbol{I}_p+\boldsymbol{Q}(\boldsymbol{Q}^\tau\boldsymbol{S}\boldsymbol{Q})^{-1}\boldsymbol{Q}^\tau\cdot\boldsymbol{Y}\boldsymbol{P}_{Z^\tau}\boldsymbol{Y}^\tau\}\\
&= \frac{1}{n^p}\cdot\det(\boldsymbol{S})\cdot\det\{\boldsymbol{I}_{p-m}+(\boldsymbol{Q}^\tau\boldsymbol{S}\boldsymbol{Q})^{-1}\cdot\boldsymbol{Q}^\tau\boldsymbol{Y}\boldsymbol{P}_{Z^\tau}\boldsymbol{Y}^\tau\boldsymbol{Q}\}\\
&= \frac{1}{n^p}\cdot\det(\boldsymbol{S})\cdot\det\{(\boldsymbol{Q}^\tau\boldsymbol{S}\boldsymbol{Q})^{-1}\}\\
&\quad\cdot\det\{(\boldsymbol{Q}^\tau\boldsymbol{S}\boldsymbol{Q})+\boldsymbol{Q}^\tau\boldsymbol{Y}\boldsymbol{P}_{Z^\tau}\boldsymbol{Y}^\tau\boldsymbol{Q}\}\\
&= \frac{1}{n^p}\cdot\det(\boldsymbol{S})\cdot\det\{(\boldsymbol{Q}^\tau\boldsymbol{S}\boldsymbol{Q})^{-1}\}\cdot\det\{\boldsymbol{Q}^\tau\boldsymbol{Y}\boldsymbol{Y}^\tau\boldsymbol{Q}\}.
\end{aligned}
$$

In the same manner it follows

$$
(2.21)\qquad \det(\hat{\Sigma}_a) = \frac{1}{n^p}\cdot\det(\boldsymbol{S}_a)\cdot\det\{(\boldsymbol{Q}^\tau\boldsymbol{S}_a\boldsymbol{Q})^{-1}\}\cdot\det\{\boldsymbol{Q}^\tau\boldsymbol{Y}\boldsymbol{Y}^\tau\boldsymbol{Q}\}.
$$

On the one hand, (2.17) and (2.4) imply

$$
\begin{aligned}
\det(\boldsymbol{S}_a) &= \det\{\boldsymbol{S}_{(I)}\}\\
&= \det(\boldsymbol{S})\cdot\det\{\boldsymbol{I}_p-\boldsymbol{S}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k-\boldsymbol{H}_I)^{-1}\boldsymbol{e}_I^\tau\}\\
&= \det(\boldsymbol{S})\cdot\det\{\boldsymbol{I}_k-\boldsymbol{e}_I^\tau\boldsymbol{S}^{-1}\boldsymbol{e}_I\cdot(\boldsymbol{I}_k-\boldsymbol{H}_I)^{-1}\}\\
&= \det\{(\boldsymbol{I}_k-\boldsymbol{H}_I)^{-1}\}\cdot\det(\boldsymbol{S})\cdot\det\{\boldsymbol{I}_k-\boldsymbol{H}_I-\boldsymbol{e}_I^\tau\boldsymbol{S}^{-1}\boldsymbol{e}_I\}
\end{aligned}
$$

and

$$
\begin{aligned}
\det(\boldsymbol{Q}^\tau\boldsymbol{S}_a\boldsymbol{Q}) &= \det\{\boldsymbol{Q}^\tau\boldsymbol{S}_{(I)}\boldsymbol{Q}\}\\
&= \det(\boldsymbol{Q}^\tau\boldsymbol{S}\boldsymbol{Q})\cdot\det\{\boldsymbol{I}_{p-m}-(\boldsymbol{Q}^\tau\boldsymbol{S}\boldsymbol{Q})^{-1}\boldsymbol{Q}^\tau\boldsymbol{e}_I(\boldsymbol{I}_k-\boldsymbol{H}_I)^{-1}\boldsymbol{e}_I^\tau\boldsymbol{Q}\}\\
&= \det(\boldsymbol{Q}^\tau\boldsymbol{S}\boldsymbol{Q})\cdot\det\{\boldsymbol{I}_k-\boldsymbol{e}_I^\tau\boldsymbol{Q}(\boldsymbol{Q}^\tau\boldsymbol{S}\boldsymbol{Q})^{-1}\boldsymbol{Q}^\tau\boldsymbol{e}_I\cdot(\boldsymbol{I}_k-\boldsymbol{H}_I)^{-1}\}\\
&= \det\{(\boldsymbol{I}_k-\boldsymbol{H}_I)^{-1}\}\cdot\det(\boldsymbol{Q}^\tau\boldsymbol{S}\boldsymbol{Q})\\
&\quad\cdot\det\{\boldsymbol{I}_k-\boldsymbol{H}_I-\boldsymbol{e}_I^\tau\boldsymbol{Q}(\boldsymbol{Q}^\tau\boldsymbol{S}\boldsymbol{Q})^{-1}\boldsymbol{Q}^\tau\boldsymbol{e}_I\}.
\end{aligned}
$$

Therefore, according to (2.20) and (2.21) we can conclude that

$$
\begin{aligned}
T_I &\equiv \frac{\det(\hat{\Sigma})}{\det(\hat{\Sigma}_a)}\\
&= \left\{\frac{\det(\boldsymbol{S})}{\det(\boldsymbol{S}_a)}\right\}\cdot\left\{\frac{\det(\boldsymbol{Q}^\tau\boldsymbol{S}_a\boldsymbol{Q})}{\det(\boldsymbol{Q}^\tau\boldsymbol{S}\boldsymbol{Q})}\right\}\\
&= \frac{\det\{\boldsymbol{I}_k-\boldsymbol{H}_I-\boldsymbol{e}_I^\tau\boldsymbol{Q}(\boldsymbol{Q}^\tau\boldsymbol{S}\boldsymbol{Q})^{-1}\boldsymbol{Q}^\tau\boldsymbol{e}_I\}}{\det\{\boldsymbol{I}_k-\boldsymbol{H}_I-\boldsymbol{e}_I^\tau\boldsymbol{S}^{-1}\boldsymbol{e}_I\}}\\
&= \frac{\det\{(\boldsymbol{I}_k-\boldsymbol{H}_I-\boldsymbol{e}_I^\tau\boldsymbol{S}^{-1}\boldsymbol{e}_I)+\boldsymbol{e}_I^\tau\boldsymbol{S}^{-1}\boldsymbol{X}(\boldsymbol{X}^\tau\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\tau\boldsymbol{S}^{-1}\boldsymbol{e}_I\}}{\det\{(\boldsymbol{I}_k-\boldsymbol{H}_I-\boldsymbol{e}_I^\tau\boldsymbol{S}^{-1}\boldsymbol{e}_I)\}}\\
&= \det\{\boldsymbol{I}_k+\boldsymbol{e}_I^\tau\boldsymbol{S}^{-1}\boldsymbol{X}(\boldsymbol{X}^\tau\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\tau\boldsymbol{S}^{-1}\boldsymbol{e}_I\cdot(\boldsymbol{I}_k-\boldsymbol{H}_I-\boldsymbol{e}_I^\tau\boldsymbol{S}^{-1}\boldsymbol{e}_I)^{-1}\}.
\end{aligned}
$$

On the other hand, (2.4) and (2.17) also imply that

$$(2.22)\quad \begin{aligned}\det(\boldsymbol{S}) &= \det\{\boldsymbol{S}_{(I)}\}\cdot\det\{\boldsymbol{I}_p+\boldsymbol{S}_{(I)}^{-1}e_I(\boldsymbol{I}_k-\boldsymbol{H}_I)^{-1}e_I^{\tau}\}\\ &= \det\{\boldsymbol{S}_{(I)}\}\cdot\det\{\boldsymbol{I}_k+e_I^{\tau}\boldsymbol{S}_{(I)}^{-1}e_I\cdot(\boldsymbol{I}_k-\boldsymbol{H}_I)^{-1}\}\\ &= \det\{(\boldsymbol{I}_k-\boldsymbol{H}_I)^{-1}\}\cdot\det\{\boldsymbol{S}_{(I)}\}\cdot\det\{\boldsymbol{I}_k-\boldsymbol{H}_I+e_I^{\tau}\boldsymbol{S}_{(I)}^{-1}e_I\}\\ &= \det\{(\boldsymbol{I}_k-\boldsymbol{H}_I)^{-1}\}\cdot\det(\boldsymbol{S}_a)\cdot\det\{\boldsymbol{I}_k-\boldsymbol{H}_I+e_I^{\tau}\boldsymbol{S}_{(I)}^{-1}e_I\}\end{aligned}$$

and

$$(2.23)\quad \begin{aligned}\det(\boldsymbol{Q}^{\tau}\boldsymbol{S}\boldsymbol{Q}) &= \det\{\boldsymbol{Q}^{\tau}\boldsymbol{S}_{(I)}\boldsymbol{Q}\}\\ &\quad\cdot\det\{\boldsymbol{I}_{p-m}+(\boldsymbol{Q}^{\tau}\boldsymbol{S}_{(I)}\boldsymbol{Q})^{-1}\boldsymbol{Q}^{\tau}e_I(\boldsymbol{I}_k-\boldsymbol{H}_I)^{-1}e_I^{\tau}\boldsymbol{Q}\}\\ &= \det(\boldsymbol{Q}^{\tau}\boldsymbol{S}_a\boldsymbol{Q})\\ &\quad\cdot\det\{\boldsymbol{I}_k+e_I^{\tau}\boldsymbol{Q}(\boldsymbol{Q}^{\tau}\boldsymbol{S}_{(I)}\boldsymbol{Q})^{-1}\boldsymbol{Q}^{\tau}e_I\cdot(\boldsymbol{I}_k-\boldsymbol{H}_I)^{-1}\}\\ &= \det\{(\boldsymbol{I}_k-\boldsymbol{H}_I)^{-1}\}\cdot\det(\boldsymbol{Q}^{\tau}\boldsymbol{S}_a\boldsymbol{Q})\\ &\quad\cdot\det\{(\boldsymbol{I}_k-\boldsymbol{H}_I+e_I^{\tau}\boldsymbol{Q}(\boldsymbol{Q}^{\tau}\boldsymbol{S}_{(I)}\boldsymbol{Q})^{-1}\boldsymbol{Q}^{\tau}e_I)\}.\end{aligned}$$

Furthermore, (2.22) and (2.23) provide that

$$(2.24)\quad \begin{aligned}\Lambda_I &\equiv \frac{\det(\hat{\boldsymbol{\Sigma}}_a)}{\det(\hat{\boldsymbol{\Sigma}})}\\ &= \left\{\frac{\det(\boldsymbol{S}_a)}{\det(\boldsymbol{S})}\right\}\cdot\left\{\frac{\det(\boldsymbol{Q}^{\tau}\boldsymbol{S}\boldsymbol{Q})}{\det(\boldsymbol{Q}^{\tau}\boldsymbol{S}_a\boldsymbol{Q})}\right\}\\ &= \frac{\det\{(\boldsymbol{I}_k-\boldsymbol{H}_I+e_I^{\tau}\boldsymbol{Q}(\boldsymbol{Q}^{\tau}\boldsymbol{S}_{(I)}\boldsymbol{Q})^{-1}\boldsymbol{Q}^{\tau}e_I)\}}{\det\{(\boldsymbol{I}_k-\boldsymbol{H}_I+e_I^{\tau}\boldsymbol{S}_{(I)}^{-1}e_I)\}}\\ &= \frac{\det\{(\boldsymbol{I}_k-\boldsymbol{H}_I+e_I^{\tau}\boldsymbol{S}_{(I)}^{-1}e_I)-e_I^{\tau}\boldsymbol{S}_{(I)}^{-1}\boldsymbol{X}(\boldsymbol{X}^{\tau}\boldsymbol{S}_{(I)}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}_{(I)}^{-1}e_I\}}{\det\{(\boldsymbol{I}_k-\boldsymbol{H}_I+e_I^{\tau}\boldsymbol{S}_{(I)}^{-1}e_I)\}}\\ &= \det\{\boldsymbol{I}_k-e_I^{\tau}\boldsymbol{S}_{(I)}^{-1}\boldsymbol{X}(\boldsymbol{X}^{\tau}\boldsymbol{S}_{(I)}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\tau}\boldsymbol{S}_{(I)}^{-1}e_I\\ &\qquad\cdot(\boldsymbol{I}_k-\boldsymbol{H}_I+e_I^{\tau}\boldsymbol{S}_{(I)}^{-1}e_I)^{-1}\}\end{aligned}$$

and the proof is complete. □

It is noted that the expression (2.19) is particularly important for deriving the distribution of Λ_I because $\boldsymbol{S}_{(I)}$ is independent of e_I, which will be explained in detail in the next section.

3. Multiple outlier detection in GCM

In this section, we deal with the problem of detecting multiple outliers in a GCM with UCM. In regression diagnostic, it is well-known that the mean-shift model is one of the most common outlier-generating models (see, e.g., Barnett and Lewis (1984)). Whether or not there are discordant outliers in the observations is reduced to testing if the mean of the population is shifted. For the GCM with

UCM, this problem is transformed into testing whether or not the mean shift parameter $\boldsymbol{\Phi}$ in MSRM is zero. In other words, it is sufficient to test the following hypothesis

$$H : \boldsymbol{\Phi} = \mathbf{0} \leftrightarrow K : \boldsymbol{\Phi} \neq \mathbf{0}. \tag{3.1}$$

When the null hypothesis is rejected at level α, $\boldsymbol{Y}_I = (\boldsymbol{y}_{n-k+1}, \boldsymbol{y}_{n-k+2}, \ldots, \boldsymbol{y}_n)$ are declared as k discordant outliers at level α (ref., e.g., Cook and Weisberg (1982), pp. 28–30; Chatterjee and Hadi (1988), pp. 187–190).

From the definition of the statistic Λ_I (or T_I), it is obvious that the likelihood ratio testing criterion of (3.1) is equivalent to rejecting the null hypothesis H if Λ_I (or T_I) is significantly small (or large). What we need to do here is to derive the exact null distribution of the testing statistic Λ_I (or T_I). It seems that the hypothesis test problem (3.1) can be reduced to a specific case of general linear hypotheses concerning the regression coefficient $\boldsymbol{B}$, and the exact null distribution of the latter was obtained by Tang and Gupta (1986) and Nagarsenker (1977) by aid of solving Wilk's type-B integral equations and zonal polynomials, respectively. By using those methods, however, the computation for looking for the critical value of Λ_I (or T_I) is complex and burdensome. On the other hand, when the general methodology provided by Khatri (1966) is applied to looking for the null distribution of Λ_I (or T_I), it is involved inevitably in investigating the relationship between Λ_I (or T_I) and Khatri's criterion. Actually, it can be shown that the testing criterion based on Λ_I (or T_I) is equivalent to that of Khatri (1966), but the proof of this conclusion is involved in some heavy loads of matrix derivation. Instead of doing directly in such a way, we present an alternative approach to derive briefly the null distribution of Λ_I, which emphasizes a much highlight and intuitive background on statistical diagnostic. In fact, there is a much simpler distribution form for the statistic Λ_I under the null hypothesis H, that is a Wilk's distribution with degree freedom m, $n-k-r-p+m$ and k, i.e., $\Lambda(m, n-k-r-p+m, k)$. Before deriving this conclusion, we need the following lemma.

LEMMA 3.1. *The matrix $\boldsymbol{S}_{(I)}$ given in (2.3) is independent of $\boldsymbol{e}_I$ in (2.4).*

PROOF. According to (2.7), the residual $\boldsymbol{e}_I$ can be simplified as

$$\begin{aligned} \boldsymbol{e}_I &= \boldsymbol{Y}_I - \boldsymbol{Y}\boldsymbol{Z}^\tau(\boldsymbol{Z}\boldsymbol{Z}^\tau)^{-1}\boldsymbol{Z}_I \\ &= \boldsymbol{Y}_I - \boldsymbol{Y}_{(I)}\boldsymbol{Z}_{(I)}^\tau(\boldsymbol{Z}_{(I)}\boldsymbol{Z}_{(I)}^\tau)^{-1}\boldsymbol{Z}_I - \boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}\boldsymbol{H}_I, \end{aligned}$$

which induces

$$\boldsymbol{e}_I(\boldsymbol{I}_k + (\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}\boldsymbol{H}_I) = \boldsymbol{Y}_I - \boldsymbol{Y}_{(I)}\boldsymbol{Z}_{(I)}^\tau(\boldsymbol{Z}_{(I)}\boldsymbol{Z}_{(I)}^\tau)^{-1}\boldsymbol{Z}_I,$$

equivalently,

$$\boldsymbol{e}_I = (\boldsymbol{Y}_I - \boldsymbol{Y}_{(I)}\boldsymbol{Z}_{(I)}^\tau(\boldsymbol{Z}_{(I)}\boldsymbol{Z}_{(I)}^\tau)^{-1}\boldsymbol{Z}_I)(\boldsymbol{I}_k - \boldsymbol{H}_I) \tag{3.2}$$

where we use the fact

$$(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1} = \boldsymbol{I}_k + (\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}\boldsymbol{H}_I = \boldsymbol{I}_k + \boldsymbol{H}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1}.$$

From the definition of $\boldsymbol{S}_{(I)} = \boldsymbol{Y}_{(I)}(\boldsymbol{I}_{n-k} - \boldsymbol{P}_{\boldsymbol{Z}^{\tau}_{(I)}})\boldsymbol{Y}^{\tau}_{(I)}$, it is clear that $\boldsymbol{S}_{(I)}$ is independent of $\boldsymbol{Y}_I$ and $\boldsymbol{Y}_{(I)}\boldsymbol{Z}^{\tau}_{(I)}(\boldsymbol{Z}_{(I)}\boldsymbol{Z}^{\tau}_{(I)})^{-1}$. From (3.2), we know that $\boldsymbol{S}_{(I)}$ is independent of $\boldsymbol{e}_I$, and the proof is complete. □

THEOREM 3.1. *For MSRM, the likelihood ratio test of level α of $H : \boldsymbol{\Phi} = \boldsymbol{0}$ versus $K : \boldsymbol{\Phi} \neq \boldsymbol{0}$ is equivalent to rejecting H if $T_I \geq C^*_\alpha \equiv 1/C_\alpha$, where the statistic T_I is defined by (2.18) and C_α denotes the lower $100\alpha\%$ critical point of the Wilk's distribution $\Lambda(m, n-k-r-p+m, k)$.*

PROOF. According to the relationship $T_I \equiv \Lambda_I^{-1}$, it is enough to prove that $\Lambda_I \sim \Lambda(m, n-k-r-p+m, k)$ under the null hypothesis H. From (2.24), first, the statistic Λ_I can be simplified as follows

$$\text{(3.3)} \qquad \Lambda_I = \frac{\det\{\boldsymbol{I}_k + (\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1/2}\boldsymbol{e}_I^{\tau}\boldsymbol{Q}(\boldsymbol{Q}^{\tau}\boldsymbol{S}_{(I)}\boldsymbol{Q})^{-1}\boldsymbol{Q}^{\tau}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1/2}\}}{\det\{\boldsymbol{I}_k + (\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1/2}\boldsymbol{e}_I^{\tau}\boldsymbol{S}_{(I)}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1/2}\}}$$

where $\boldsymbol{Q} \in \mathcal{Q}$. Let $\boldsymbol{Q} = \boldsymbol{\Gamma}_1\boldsymbol{\Lambda}_1\boldsymbol{\Gamma}_2$ be the singular value decomposition (SVD) of $\boldsymbol{Q}$, where $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ are $p \times p$ and $(p-m)\times(p-m)$ orthogonal matrices, respectively, and $\boldsymbol{\Lambda}_1 = (\boldsymbol{\Lambda}, \boldsymbol{0})^{\tau}$ is a $p \times (p-m)$ matrix in which $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_{p-m})$ $(\lambda_i \neq 0, 1 \leq i \leq p-m)$. Denote $\boldsymbol{A} = \boldsymbol{\Gamma}_1^{\tau}\boldsymbol{S}_{(I)}\boldsymbol{\Gamma}_1$ and $\boldsymbol{\nu} = \boldsymbol{\Gamma}_1^{\tau}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1/2}$. Corresponding to the orders of $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$, the matrix $\boldsymbol{A}$ and $\boldsymbol{\nu}$ can be partitioned into, respectively,

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\nu} = \begin{pmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \end{pmatrix}$$

where the sizes of $\boldsymbol{A}_{11}$, $\boldsymbol{A}_{22}$, $\boldsymbol{A}_{12}(= \boldsymbol{A}_{21}^{\tau})$, $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ are $(p-m)\times(p-m)$, $m \times m$, $(p-m)\times m$, $(p-m)\times k$ and $m \times k$, respectively. It can be shown that

$$\text{(3.4)} \qquad (\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1/2}\boldsymbol{e}_I^{\tau}\boldsymbol{Q}(\boldsymbol{Q}^{\tau}\boldsymbol{S}_{(I)}\boldsymbol{Q})^{-1}\boldsymbol{Q}^{\tau}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1/2} = \boldsymbol{\nu}_1^{\tau}\boldsymbol{A}_{11}^{-1}\boldsymbol{\nu}_1$$

and

$$\text{(3.5)} \qquad (\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1/2}\boldsymbol{e}_I^{\tau}\boldsymbol{S}_{(I)}^{-1}\boldsymbol{e}_I(\boldsymbol{I}_k - \boldsymbol{H}_I)^{-1/2} = \boldsymbol{\nu}^{\tau}\boldsymbol{A}^{-1}\boldsymbol{\nu}.$$

By using (3.4) and (3.5), the statistic Λ_I in (3.3) becomes

$$\text{(3.6)} \qquad \Lambda_I = \frac{\det(\boldsymbol{I}_k + \boldsymbol{\nu}_1^{\tau}\boldsymbol{A}_{11}^{-1}\boldsymbol{\nu}_1)}{\det(\boldsymbol{I}_k + \boldsymbol{\nu}^{\tau}\boldsymbol{A}^{-1}\boldsymbol{\nu})} = \frac{\det(\boldsymbol{A})}{\det(\boldsymbol{A}_{11})} \cdot \frac{\det(\boldsymbol{A}_{11} + \boldsymbol{\nu}_1\boldsymbol{\nu}_1^{\tau})}{\det(\boldsymbol{A} + \boldsymbol{\nu}\boldsymbol{\nu}^{\tau})}.$$

Let $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{\nu}\boldsymbol{\nu}^{\tau}$ and partition $\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \end{pmatrix}$ such that $\boldsymbol{C}_{11} = \boldsymbol{A}_{11} + \boldsymbol{\nu}_1\boldsymbol{\nu}_1^{\tau}$. If we denote $\boldsymbol{A}_{22.1} = \boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}$ and $\boldsymbol{C}_{22.1} = \boldsymbol{C}_{22} - \boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}$, then (3.6) can be written as

$$\text{(3.7)} \qquad \Lambda_I = \frac{\det(\boldsymbol{A}_{22.1})}{\det(\boldsymbol{C}_{22.1})} = \frac{\det(\boldsymbol{A}_{22.1})}{\det(\boldsymbol{A}_{22.1} + \boldsymbol{W}_{22.1})},$$

where $W_{22.1} = C_{22.1} - A_{22.1}$. Now, let us derive the null distribution of Λ_I from (3.7).

By the definition of $S_{(I)}$, under the null hypothesis H, it is obvious $S_{(I)} \sim W_p(n-k-r, \Sigma)$ so that $A \sim W_p(n-k-r, \Sigma^*)$ where $\Sigma^* = \Gamma_1^\tau \Sigma \Gamma_1$ and $W_p(k, \Sigma)$ is the p-dimensional Wishart distribution with parameters k and Σ. Furthermore,

$$A_{22.1} \sim W_m(n-k-r-p+m, \Sigma^*_{22.1}) \tag{3.8}$$

and is independent of A_{12} and A_{11} (see, e.g., Muirhead (1982), pp. 93–95), where $\Sigma^*_{22.1} = \Sigma^*_{22} - \Sigma^*_{21}\Sigma^{*-1}_{11}\Sigma^*_{12}$ and $\Sigma^* = \begin{pmatrix} \Sigma^*_{11} & \Sigma^*_{12} \\ \Sigma^*_{21} & \Sigma^*_{22} \end{pmatrix}$ in which partition corresponds to that of A. On the other hand, notice that

$$W_{22.1} = \nu_2\nu_2^\tau + A_{12}^\tau A_{11}^{-1} A_{12} - (A_{12} + \nu_1\nu_2^\tau)^\tau (A_{11} + \nu_1\nu_1^\tau)^{-1}(A_{12} + \nu_1\nu_2^\tau), \tag{3.9}$$

which is a function of ν, A_{12} and A_{11}. Since A is independent of ν according to Lemma 3.1, we know $A_{22.1}$ is independent of $W_{22.1}$. Finally, under the null hypothesis H, as $e_I \sim N_{p,k}(0, \Sigma^* \otimes I_k)$ so that $\nu\nu^\tau \sim W_p(k, \Sigma^*)$ and is independent of A. Furthermore, $C = A + \nu\nu^\tau \sim W_p(n-r, \Sigma^*)$ thus $C_{22.1} = A_{22.1} + W_{22.1} \sim W_m(n-r-p+m, \Sigma^*_{22.1})$. This fact and (3.8) imply

$$W_{22.1} \sim W_m(k, \Sigma^*_{22.1}) \tag{3.10}$$

as $A_{22.1}$ is independent of $W_{22.1}$. According to the definition of Wilk's distribution (see, e.g., Muirhead (1982)), we know that under H

$$\Lambda_I \sim \Lambda(m, n-k-r-p+m, k)$$

and the proof is complete. □

Remark 3.1. When $k = 1$ and $I = \{i\}$ $(1 \le i \le n)$, which means to detect whether or not the i-th individual is a single discordant outlier, according to Wilk's distributional property we know under the null hypothesis

$$\frac{n-r-p}{m} \cdot \frac{1-\Lambda_i}{\Lambda_i} \sim F_{m,n-r-p}. \tag{3.11}$$

Therefore, the i-th individual is declared as a single discordant outlier if

$$T_i = 1 + \frac{e_i^\tau S^{-1} X (X^\tau S^{-1} X)^{-1} X^\tau S^{-1} e_i}{1 - p_{ii} - e_i^\tau S^{-1} e_i} \ge 1 + \frac{mC^*_\alpha}{n-r-p}, \tag{3.12}$$

where p_{ii} is the i-th diagonal element of the projection matrix P_{Z^τ}, e_i is the i-th column of the residual e and C^*_α is the upper $100\alpha\%$ critical point of the distribution $F_{m,n-r-p}$.

Remark 3.2. When $k = 2$ and $I = \{i, j\}$ $(1 \le i, j \le n, i \ne j)$, we want to detect whether or not the (i,j)-th individual pair is a discordant outlier pair. According to Wilk's distributional property we know under the null hypothesis

$$\frac{n-r-p-2}{m} \cdot \frac{1-\sqrt{\Lambda_{i,j}}}{\sqrt{\Lambda_{i,j}}} \sim F_{2m,2(n-r-p-2)}. \tag{3.13}$$

Therefore, the (i,j)-th individual is declared as a discordant outlier pair if

$$T_{i,j} \geq \left(1 + \frac{mC_\alpha^{**}}{n-r-p-2}\right)^2, \tag{3.14}$$

where C_α^{**} is the upper $100\alpha\%$ critical point of the distribution $F_{2m,2(n-r-p-2)}$ and $T_{i,j}$ is defined by (2.18).

Remark 3.3. In general, the index subset I could not be given in advance even though the number k is fixed. In this case, a reasonable test statistic

$$\Lambda_{\min}^k = \min_I \{\Lambda_I\} \tag{3.15}$$

is proposed to detect multiple outliers, where I runs over all subsets containing k indexes. The exact null distribution of $\Lambda_{\min}^k$, however, is unknown because Λ_I's are not independent mutually. In this situation, Bonferroni's principle in multiple comparisons is recommended (ref., e.g., Cook and Weisberg (1982), pp. 26–27; Barnett and Lewis (1984), pp. 256–258).

Remark 3.4. Denote $\Lambda_I(\boldsymbol{Y}) = \Lambda_I$ then it is obvious that $\Lambda_I(\alpha\boldsymbol{Y}) \equiv \Lambda_I(\boldsymbol{Y}) \equiv \Lambda_I(\boldsymbol{Y} - \boldsymbol{XBZ})$ for all $\alpha > 0$ and all $m \times r$ matrix $\boldsymbol{B}$. Therefore, the null distribution of Λ_I is distribution-free or distribution-robust in the class of elliptically contoured distributions (see, e.g., Fang and Zhang (1990)). This fact implies that the outlier detection criteria given in Theorem 3.1, (3.12), (3.14) and (3.15) can be extended to elliptically contoured distributions.

4. Illustrative examples

In this section some of the results developed in the preceding sections are applied to two biological data sets analyzed by Rao (1984, 1987) and Lee (1988, 1991). The primary objective is to illustrate the applications of our results. Following Lee (1991), an arbitrary covariance structure $\boldsymbol{\Sigma} > \boldsymbol{0}$ can be assumed in the following analyses.

4.1 *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 (1988, 1991). Dental measurements were made on 11 girls and 16 boys at ages 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 $\boldsymbol{Z}$ can be taken as the following forms, respectively:

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

where $\boldsymbol{1}_s$ is a s-variant vector with components 1's. Table 1 displays some numerical results of the measurements in decreasing order discussed in the preceding

Table 1. Diagnostic statistic for dental data.

Individual No.	p_{ii}	T_i	Individual-pair No.	$T_{i,j}$
24	0.0625	1.9197	(20, 24)	2.6654
15	0.0625	1.4433	(15, 24)	2.6210
21	0.0625	1.2961	(10, 24)	2.5190
10	0.0909	1.2738	(21, 24)	2.4665
20	0.0625	1.2297	(3, 24)	2.2575

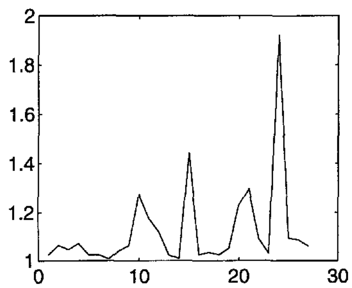

Fig. 1.

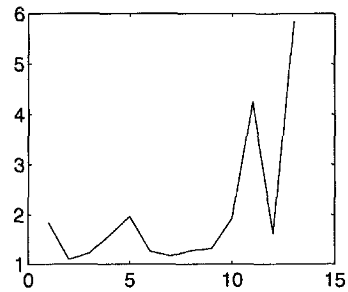

Fig. 2.

sections for detecting a single discordant outlier and outlier pair. Also, the index plot of the diagnostic statistic T_i is showen in Fig. 1.

For detecting a single discordant outlier, it is clear that the individuals 24 and 15 stand out according to the values of T_i in Table 1 and Fig. 1. It seems that the individual 24 is a discordant outlier. In fact, since the right hand side of inequality (3.12) at level $\alpha = 0.01$ is $1 + mC_\alpha^*/(n-r-p) = 1.5505$, there is only the individual 24 such that $T_{24} = 1.9197 > 1.5505$. Therefore, the 24th individual can be declared as a discordant outlier at level $\alpha = 0.01$. The statuses of the 15th and 21th individuals, however, are more questionable and should be treated cautiously.

Since $T_{24} = 1.9197$ and $T_{15} = 1.4433$ are the largest two values of the diagnostic statistic T_i, it seems that the individual pair $(15, 24)$ should be a discordant outlier pair. But the numerical results given in Table 1 show that the maximum of $T_{i,j}$'s value for the individual pair (i, j) is achieved at $(20, 24)$ with $T_{20,24} = 2.6654$. Noticing that the critical value of $T_{i,j}$ at level $\alpha = 0.01$ in (3.14) for detecting outlier pair is $(1 + mC_\alpha^{**}/(n-r-p-2))^2 = 1.9689$, which is smaller than the values of $T_{i,j}$ listed in Table 1, we can conclude that the individual pair $(20, 24)$ is a most discordant outlier pair. Of course the discordance of the individual pair $(15, 24)$ should be noticed and treated carefully.

4.2 *Mouse data*

The data set was analyzed by Rao (1984, 1987) and later by Lee (1988, 1991). It consists of weights of 13 male mice measured at intervals of 3 days from birth to weaning. For this data set, following Rao (1984), a second-degree polynomial in time for the growth function was assumed and hence the design matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ take the following forms

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 4 & 9 & 16 & 25 & 36 & 49 \end{pmatrix}^{\tau}$$

and $\boldsymbol{Z} = \mathbf{1}_{13}^{\tau}$, respectively. Some numerical results of the diagnostic statistics for detecting a single outlier and outlier pair are presented in Table 2 and Fig. 2, respectively.

Table 2. Diagnostic statistic for mouse data.

Individual No.	p_{ii}	T_i	Individual-pair No.	$T_{i,j}$
13	0.0769	5.8416	$(11, 13)$	24.1923
11	0.0769	4.2393	$(5, 13)$	16.9751
5	0.0769	1.9629	$(6, 13)$	13.0252
10	0.0769	1.9183	$(2, 13)$	12.8205
1	0.0769	1.8424	$(10, 13)$	12.7596

For detecting a single discordant outlier, the numerical values of T_i in Table 2 imply that the 13th and 11th individuals stand out. At level $\alpha = 0.1$, the critical value of T_i given in (3.12) is $1 + mC_\alpha^*/(n - r - p) = 3.1720$. From the values of T_i in Table 2, it is obvious that $T_{13} = 5.8416 > T_{11} = 4.2393 > 3.1720$. In other words, there are two individuals, No. 13 and No. 11, such that their T's values are greater than the critical value. Therefore, the 13th individual can be declared as a most discordant outlier at the 10 per cent level. The status of the 11th individual, however, is more questionable and should be investigated cautiously.

On the outlier pair problem, we calculate the values of $T_{i,j}$ and list the largest five ones of $T_{i,j}$ $(1 \leq i, j \leq 13)$ in decreasing order in Table 2. At level $\alpha = 0.10$, the critical value of $T_{i,j}$ given in (3.14) is $(1 + mC_\alpha^{**}/(n - r - p - 2))^2 = 16.4025$, which is smaller than the largest two values of $T_{i,j}$, $T_{11,13} = 24.1923$ and $T_{5,13} = 16.9751$ but greater than the others. This fact implies that the $(11, 13)$-th individual pair can be declared as a most discordant outlier pair at the 10 per cent level. In addition, the status of the individual pair $(5, 13)$ should be investigated carefully.

4.3 *Concluding remarks*

(a) The magnitude of the diagnostic statistic $T_{i,j}$ in detecting a discordant outlier pair does not completely depend upon the values of T_i and T_j. For example, the large values of T_i and T_j do not necessarily guarantee to induce large value of $T_{i,j}$. The results for the Dental Data set illustrate this point well. In this case, the largest two values of T_i are T_{24} and T_{15}, but the largest value of $T_{i,j}$ is achieved at the individual pair $(20, 24)$.

(b) The diagnostic statistics presented in this paper are based on an assumption that the number k of possible outliers is given in advance. In practice, however, the number k is usually unknown and the diagnostic measurements suffer from the so-called masking and swamping effects. Fortunately, some methodologies based on robust statistics have been proposed recently to solve these problems (see, e.g., Rousseeuw and van Zomeren (1990)), and the masking and swamping effects in a GCM can be partially treated by those methods.

Acknowledgements

We are grateful to the Managing Editor, Associate Editor and two anonymous Referees for their very stimulating and helpful comments. Also, discussions with Dr. Dietrich von Rosen have improved significantly this presentation and are especially acknowledged. The authors would like to thank Dr. F. J. Hickernell for his valuable suggestions and comments.

References

Barnett, V. and Lewis, T. (1984). *Outliers in Statistical Data*, Wiley, New York.

Chatterjee, S. and Hadi, A. S. (1988). *Sensitivity Analysis in Linear Regression*, Wiley, New York.

Cook, R. D. (1977). Detection of influential observations in linear regression, *Technometrics*, **19**, 15–18.

Cook, R. D. and Weisberg, S. (1982). *Residual and Influence in Regression*, Chapman and Hall, New York.

Fang, K. T. and Zhang, Y. T. (1990). *Generalized Multivariate Analysis*, Springer, Berlin.

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

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). Applications of growth curve prediction, *Sankhyā Ser. A*, **37**, 239–256.

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.

Nagarsenker, B. N. (1977). On the exact non-null distributions of the LR criterion in a general MANOVA model, *Sankhyā Ser. A*, **39**, 251–263.

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

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. (1967). Least square theory using an estimated dispersion matrix and its applications to measurement of signals, *Proc. 5th Berkeley Symp. on Math. Statist. Probab.*, **1**, 355–372.

Rao, C. R. (1984). Prediction of future observations in polynomial growth curve models, *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Statistics: Applications and New Directions*, 512–520, Indian Statistical Institute, Calcutta.

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

Rousseeuw, P. J. and van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points (with discussion), *J. Amer. Statist. Assoc.*, **85**, 633–651.

Tang, J. and Gupta, A. K. (1986). Exact distributions of certain general test statistics in multivariate analysis, *Austral. J. Statist.*, **28**, 107–114.

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

von Rosen, D. (1990). Moments for a multivariate linear normal models with application to growth curve model, *J. Multivariate Anal.*, **35**, 243–259.

von Rosen, D. (1991). The growth curve model: a review, *Comm. Statist. Theory Methods*, **20**(9), 2791–2822.