Research Report

Department of Statistics

No. 2011:1

Local influence analysis in 2 x 2 cross-over designs

Chengcheng Hao
Tatjana von Rosen
Dietrich von Rosen

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

Local influence analysis in 2×2 cross-over designs

Chengcheng Hao[a], Tatjana von Rosen[a], Dietrich von Rosen[b,c]

[a]*Department of Statistics, Stockholm University, SE-106 91 Stockholm*
[b]*Department of Energy and Technology, Swedish University of Agricultural Sciences, SE-750 07 Uppsala*
[c]*Department of Mathematics, Linköping University, SE-581 83 Linköping*

Abstract

The aim of this work is to develop new methodology to detect influential observations in cross-over design models with random individual effects. Various case-weighted perturbations are performed. We derive the exact solution of influence of the perturbations on each parameter estimate and their dispersion matrix. Closed-form maximum likelihood estimates (MLEs) of variance parameters as well as fixed effect parameters in the cross-over design models are utilised. The work exhibits the possibility to produce closed-form expressions of the influence using the residuals in mixed models. A discussion on restrictions of the case-weighted perturbation schemes is given. Some graphical tools are also presented.

Keywords: Cross-over design modelling, Explicit maximum likelihood estimate, Influential observation, Mixed linear model, Perturbation scheme, Statistical diagnostics

1. Introduction

Assessing influence when perturbing a statistical model is an essential area of statistical research. Statistical inference in one way or another assumes that results are not sensitive to small deviations from the model due to observed data. However, data mining experiences suggest that a dataset often contains several observations with demonstrably larger effects than other observations on estimating or testing an assumed model. According to Belsley *et al.* (2004), such observations are defined as influential observations. In this article, influential observations are of interest under a specific mixed linear model, that is, the cross-over design model.

The cross-over designs, also mentioned in the literature as change-over or repeated measurements designs, are designs in which subjects get different treatments in certain orders with the goal of studying differences between individual treatments (Jones and Kenward, 1989). These designs are often utilised because of their practical simplicity and yet greater power to detect differences of treatments than a parallel study (Grizzle, 1965). They are used by studies in medicine, psychology and agricultural sciences when a few independent subjects are available (examples are, for instance, provided by Afsarinejad, 1990). Modelling cross-over studies has continued to be an active area of statistical research during the past 25 years (Senn, 2006).

Although articles seldom address the issue of influential observations within the framework of cross-over designs, detecting influential observations in statistical models with covariance structure has been explored comprehensively by theoretical studies over the last 30 years, for example, Shi (2006) in a context of multilevel models and Pan and Fang (2002) in the Growth Curve model. More and more researchers, including Hodges (1998), Tan *et al.* (2001) and Banerjee and Frees (1997), have realised that the well-performed Cook's Distance in the homoscedastic or general linear model meets problems when it is extended to multivariate set-ups.

Two methodologies dominate studies on influence analysis. One approach (see e.g. Christensen *et al.*, 1992; Haslett and Dillane, 2004; Shi and Chen, 2008a), suggests to compare the estimates based on the models before and after removing a subset of the data. The ideal outcome is to generate a computationally cheap procedure, which can be expressed or approximated without re-fitting the model. A second approach, namely likelihood-based local influence (Cook, 1986), is based on the curvature of log-likelihood functions. The arguments for applying it are that to omit the whole observation or subject yields a loss of information on the influence, and therefore one should evaluate a minor modification of the original model, which is by Cook (1986) referred to as the perturbation scheme. Beckman *et al.* (1987), Lesaffre and Verbeke (1998) and Tan *et al.* (2001) apply the technique to mixed linear models. Likelihood-based local influence is demonstrated to be a general effective method to discover influential observations, however, without any straightforward explanation of the nature of the influence. Although the likelihood distance is a summary measure which expresses the joint influence on all parameters in the model (Schabenberger, 2005), when particular parameter estimates or tests are of concern, influence analysis relying only on the likelihood distance may make the problem at hand obscure. More explicit outcomes can be achieved in some specific cases, for example outcomes from cross-over design studies which this article discusses.

In this work, a compromise between case-deletion and likelihood-based local influence is presented. The method calculates the changes on the parameter estimates via different case-weighted perturbation schemes. Pregibon (1981) in univariate logistic regression, von Rosen (1995) in the Growth Curve model, Demidenko and Stukel (2005) in mixed linear model, and Shi and Chen (2008b) in multilevel models adopt a similar paradigm. Closed-form maximum likelihood estimates (MLEs) of the parameters in the cross-over designs are utilised. A main distinction of this work from existing results is that it exhibits the possibility to produce closed-form expressions of the influence by using the residuals, even in mixed models with unknown variance components. A discussion on restrictions of the case-weighted perturbation schemes, which to our knowledge has not yet been considered, is also given.

In the next section, we introduce prerequired knowledge on the cross-over design model and influence analysis. Perturbation schemes and objective functions of influence which are used in the coming discussion are also defined. In Sections 3, the obtained results for balanced 2×2 cross-over designs are presented. Section 4 illustrates the main results via a simulation study. Conclusions are presented in Section 5.

2. Preliminaries

Throughout this article, upper case letters with bold face denote matrices, bold lower case letters denote column vectors and non-bold lower case letters with subscripts are used to show elements of matrices or vectors. Let $\boldsymbol{I}_p$, $\mathbf{1}_p$ and $\boldsymbol{J}_p = \mathbf{1}_p \mathbf{1}_p^{\mathrm{T}}$ denote the $p \times p$ identity matrix, the $p \times 1$ vector and the $p \times p$ matrix with elements equal to 1, respectively. The symbol $\otimes$ represents the Kronecker product of matrices. Moreover, the vector space generated by the columns of the $p \times q$ matrix $\boldsymbol{A}$, $\mathcal{C}(\boldsymbol{A})$, is given by $\mathcal{C}(\boldsymbol{A}) = \{\boldsymbol{a} : \boldsymbol{a} = \boldsymbol{A}\boldsymbol{z}, \boldsymbol{z} \in \mathbf{R}^q\}$. The p-dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted $N_p(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$.

2.1. Cross-over design model

Let y_{ijk} denote a response observed during the k-th period on the j-th subject within sequence i of a cross-over design, where $i = 1, 2, \ldots, s$; $j = 1, 2, \ldots, n$; $k = 1, 2, \ldots, p$. The approach throughout this work is to assume a balanced mixed linear model for y_{ijk}. There are various possibilities for model formulation and data analysis, but once the model is set up, the important issue of model validation can be addressed.

Suppose a comparison of two treatments, A and B, is of concern. A natural design to employ is a two-sequence two-period (2×2) cross-over design, that is, $s = 2$ and $p = 2$, whereby subjects are administered with two sequences of treatments, to receive treatment A followed by treatment B (sequence AB) or to receive treatment B followed by treatment A (sequence BA). One widely used model for this design is the one specified in Grizzle (1965) with random effects, which after performing a reparametrization can be presented as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{y} = (y_{111}, y_{112}, y_{121}, \ldots, y_{2n2})^{\mathrm{T}}$ is a $4n \times 1$ vector of responses. The parameter $\boldsymbol{\beta} = (\mu, \pi, \phi, \lambda)^{\mathrm{T}}$ is a vector of unknown mean parameters, in which μ is the general mean, π and ϕ represent the fixed effects for period and treatment, respectively, and the treatment by period interaction, which is referred to as carry-over effect in the literature, is given by λ. The matrix $\boldsymbol{X} = (\mathbf{x}_{111}, \mathbf{x}_{112}, \mathbf{x}_{121}, \ldots, \mathbf{x}_{2n2})^{\mathrm{T}}$ is a $4n \times 4$ known design matrix for $\boldsymbol{\beta}$. Note that the column vector $\mathbf{x}_{ijk}$ is a row of $\boldsymbol{X}$ and is given by

$$\mathbf{x}_{1j1} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}^{\mathrm{T}}, \ \mathbf{x}_{1j2} = \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{4} \end{pmatrix}^{\mathrm{T}},$$
$$\mathbf{x}_{2j1} = \begin{pmatrix} 1 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{4} \end{pmatrix}^{\mathrm{T}}, \ \mathbf{x}_{2j2} = \begin{pmatrix} 1 & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{4} \end{pmatrix}^{\mathrm{T}},$$

for $j = 1, 2, \ldots, n$. The vector $\boldsymbol{\gamma} = (\gamma_{11}, \gamma_{12}, \ldots, \gamma_{2n})^{\mathrm{T}}$ is a random-effects vector representing individual effects which satisfies $\boldsymbol{\gamma} \sim N_{2n}(\mathbf{0},\, \sigma_\gamma^2 \boldsymbol{I})$. The matrix $\boldsymbol{Z} = \boldsymbol{I}_2 \otimes \boldsymbol{I}_n \otimes \mathbf{1}_2$ is the $4n \times 2n$ known incidence matrix for $\boldsymbol{\gamma}$. The vector of random errors $\boldsymbol{\epsilon} \sim N_{4n}(\mathbf{0}, \sigma_e^2 \boldsymbol{I})$ is assumed to be independent of $\boldsymbol{\gamma}$, and the variances σ_γ^2 and σ_e^2 are supposed to be unknown.

2.2. Perturbation schemes

As mentioned in the introduction, two issues on model validation are of importance: to detect influential observations, and to identify their roles on estimates or tests. For these purposes, two central concepts are introduced, the perturbation scheme (Beckman *et al.*, 1987) and the objective functions of influence (Belsley *et al.*, 2004). In this article, the notation $P(\boldsymbol{\omega})$ denotes the perturbation scheme where $\boldsymbol{\omega}$ is the perturbation weight. The objective functions of influence used in this work will be listed in the next subsection.

Case deletion (Cook and Weisberg, 1982) can be regarded as the best-known perturbation scheme, in which the influence is studied by complete removal of the concerned cases from the analysis. But a direct comparison of the estimates according to case deletion may become logically complicated under the cross-over designs. One problem is that, if the whole subject is removed from the design, the individual effect of this subject will no longer be estimable (Banerjee and Frees, 1997). More seriously, when a deletion occurs on the observation level of the data, e.g., one period of a subject is removed from the design, imbalance will be introduced by the deletion and the basic properties of the cross-over design will alter. Therefore, case-weighted perturbation schemes are used in this article.

Definition 2.1. *Suppose that a perturbation scheme $P(\boldsymbol{\omega})$ exists such that the response vector is modified from $\boldsymbol{y}$ to $\boldsymbol{y}_{P(\boldsymbol{\omega})}$, and the design matrix from $\boldsymbol{X}$ to $\boldsymbol{X}_{P(\boldsymbol{\omega})}$. The case-weighted perturbation scheme is a perturbation scheme which satisfies the following two criteria.*

(i) *A subset of the data is removed when $\boldsymbol{\omega} = \mathbf{0}$;*
(ii) *$\boldsymbol{y}_{P(\boldsymbol{\omega}_0)} = \boldsymbol{y}$ and $\boldsymbol{X}_{P(\boldsymbol{\omega}_0)} = \boldsymbol{X}$ for some null perturbation weight $\boldsymbol{\omega}_0$.*

By introducing $P(\boldsymbol{\omega})$, a perturbed model of (1) can be written as

$$\boldsymbol{y}_{P(\boldsymbol{\omega})} = \boldsymbol{X}_{P(\boldsymbol{\omega})}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{2}$$

where $\boldsymbol{\gamma} \sim N_{2n}(\mathbf{0}, \sigma_\gamma^2 \boldsymbol{I})$, $\boldsymbol{\epsilon} \sim N_{4n}(\mathbf{0}, \sigma_e^2 \boldsymbol{I})$ and $Cov(\boldsymbol{\gamma}, \boldsymbol{\epsilon}) = \mathbf{0}$. In further considerations, (1) is referred as unperturbed model and (2) perturbed model.

The MLEs of the perturbed model, which we will mainly focus on, are then expected to be functions of the perturbation weight $\boldsymbol{\omega}$ and the perturbation scheme $P(\boldsymbol{\omega})$. Either a "large" influence on its estimates $\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})$ or a "large" influence on its estimated dispersion matrix $\widehat{D}\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})\right]$ will highlight the risk that the inference is not robust.

It is well recognised by now that an inappropriate perturbation scheme may lead to unreasonable inference about the cause of influence (Chen *et al.*, 2010). For instance, when data are unbalanced, perturbing a subject which has more observations likely produces a larger effect than other subjects (Zhu *et al.*, 2007). Now we present some perturbation schemes which are examined later under the model assumption (1). It should be made clear that this article focuses on case-weighted perturbations, but many other kinds of $P(\omega)$ can be constructed. We refer to Beckman *et al.* (1987) for details.

Case-weighted Perturbation I. Observation level.
Let

$$\boldsymbol{y}_{P_1(\omega)} = (\boldsymbol{y}_{[ijk]}^{\mathrm{T}} : \omega y_{ijk})^{\mathrm{T}} \text{ and } \boldsymbol{X}_{P_1(\omega)} = (\boldsymbol{X}_{[ijk]}^{\mathrm{T}} : \omega \boldsymbol{x}_{ijk})^{\mathrm{T}}, \tag{3}$$

for some non-negative ω belonging to the neighbourhood of 1. The vector or matrix with index $[ijk]$ represents the associated vector or matrix with the k-th period observation in the j-th subject within the i-th sequence removed.

Case-weighted Perturbation II. Subject level.
Let

$$\boldsymbol{y}_{P_2(\omega)} = (\boldsymbol{y}_{[ij]}^{\mathrm{T}} : \omega \boldsymbol{y}_{ij}^{\mathrm{T}})^{\mathrm{T}} \text{ and } \boldsymbol{X}_{P_2(\omega)} = (\boldsymbol{X}_{[ij]}^{\mathrm{T}} : \omega \boldsymbol{X}_{ij}^{\mathrm{T}})^{\mathrm{T}}, \tag{4}$$

for some non-negative ω belonging to the neighbourhood of 1, where $\boldsymbol{y}_{ij} = (y_{ij1},\ y_{ij2})^{\mathrm{T}}$ and the matrix $\boldsymbol{X}_{ij} = (\boldsymbol{x}_{ij1},\ \boldsymbol{x}_{ij2})^{\mathrm{T}}$. The vector or matrix with index $[ij]$ represents the associated vector or matrix with the j-th subject within the i-th sequence removed.

For the extreme case $\omega = 0$, the perturbation $P_1(\omega)$ and $P_2(\omega)$ exclude one observation and one subject, respectively. The next perturbation scheme $P_3(\boldsymbol{W})$ is proposed in order to contain both observation-level and subject-level perturbations. Ouwens *et al.* (2001) emphasize the importance to detect influence on both observation and subject levels in mixed models. However, our discussion of $P_3(\boldsymbol{W})$ will show that when the covariance matrix within subject is assumed to be structural, observation-level perturbations need to be designed carefully. This is especially true for the cross-over designs. Some observation-level perturbations, such as $P_1(\omega)$, may change the structure of the model, so that influential observations become unidentifiable by it. This finding is proven in Theorem 3.4.

Case-weighted Perturbation III. Linear combination.
Let

$$\boldsymbol{y}_{P_3(\boldsymbol{W})} = \left(\boldsymbol{y}_{[ij]}^{\mathrm{T}} :\ \boldsymbol{W}\boldsymbol{y}_{ij}^{\mathrm{T}}\right)^{\mathrm{T}} \text{ and } \boldsymbol{X}_{P_3(\boldsymbol{W})} = \left(\boldsymbol{X}_{[ij]}^{\mathrm{T}} :\ \boldsymbol{W}\boldsymbol{X}_{ij}^{\mathrm{T}}\right)^{\mathrm{T}}, \tag{5}$$

for some 2×2 matrix

$$\boldsymbol{W} = \begin{pmatrix} \omega_1 & \omega_2 \\ \omega_3 & \omega_4 \end{pmatrix},$$

where $\omega_1, \omega_4 \geq 0,\ \omega_1 \geq \omega_3,$ and $\omega_2 \in \mathbb{R}$.

The null perturbation of $P_3(\boldsymbol{W})$, which is mentioned in Definition 2.1, exists at $\boldsymbol{W}_0 = \boldsymbol{I}_2$. When $\boldsymbol{W} = \boldsymbol{0}$, $P_3(\boldsymbol{W})$ is analogous to the removal of the ij-th subject. When

$$\boldsymbol{W} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

$P_3(\boldsymbol{W})$ is analogous to the removal of the observation on the second period in the ij-th subject. When

$$\boldsymbol{W} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix},$$

$P_3(\boldsymbol{W})$ is analogous to the replacement of the observation on the second period in the ij-th subject with its correlated observation on the first period.

2.3. Objective functions of influence

Because the perturbed model (2) is the same as the unperturbed model (1) when $\omega = \omega_0$, the task of assessing the influence of a perturbation can be formulated by studying local departures in (2) at ω_0. To accomplish it, an appropriate objective function to measure such departures, such as the likelihood displacement function in Cook's likelihood-based approach (Cook, 1986), or the delta-beta statistic (Belsley *et al.*, 2004) on which this article focuses, need to be defined according to the inferential interests.

Definition 2.2. *Let $l(\boldsymbol{\theta})$ and $l_{\boldsymbol{\omega}}(\boldsymbol{\theta})$ denote the log-likelihood functions under the unperturbed model (1) and the perturbed model (2), respectively, where $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^{\mathrm{T}}, \sigma_\gamma^2, \sigma_e^2\right)^{\mathrm{T}}$. Let $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ denote the maximum likelihood estimators (MLEs) of the two models. The likelihood displacement is defined as follows (Cook, 1986):*

$$LD(\boldsymbol{\omega}) = 2\{l(\widehat{\boldsymbol{\theta}}) - l(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\}.$$

Alternative objective functions for summarizing the influence of a perturbation are the above mentioned delta-beta statistic and special variance ratios.

Definition 2.3. *Let $\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})$ be the MLE of $\boldsymbol{\beta}$ and $D\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})\right]$ be the associated dispersion matrix under the perturbed model (2).*
(i) *The delta-beta statistic with respect to a perturbation $P(\boldsymbol{\omega})$ on the ij-th subject is defined by*

$$\Delta_{ij}\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\boldsymbol{\omega}) - \widehat{\boldsymbol{\beta}}(\boldsymbol{\omega}_0). \tag{6}$$

(ii) *The influence on the estimation precision of mean parameters on the ij-th subject with respect to a perturbation $P(\boldsymbol{\omega})$ is calculated by*

$$\Delta_{ij}D\left[\widehat{\boldsymbol{\beta}}\right] = \widehat{D}\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})\right] - \widehat{D}\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega}_0)\right], \tag{7}$$

where $\widehat{D}\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})\right]$ and $\widehat{D}\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega}_0)\right]$ are the estimators of $D\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})\right]$ and $D\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})\right]$, respectively, when the MLEs of σ_γ^2 and σ_e^2 are inserted.

Definition 2.4. *Let $\sigma_\gamma^2(\boldsymbol{\omega})$ and $\sigma_e^2(\boldsymbol{\omega})$ be the MLEs of the variance parameters in (2). The variance ratio for random errors (VRE) and the variance ratio for random effects (VRR) with respect to the perturbation $P(\boldsymbol{\omega})$ on the ij-th subject are defined by*

$$\mathrm{VRR}_{ij} = \frac{\widehat{\sigma}_\gamma^2(\boldsymbol{\omega})}{\widehat{\sigma}_\gamma^2(\boldsymbol{\omega}_0)}, \tag{8}$$

$$\mathrm{VRE}_{ij} = \frac{\widehat{\sigma}_e^2(\boldsymbol{\omega})}{\widehat{\sigma}_e^2(\boldsymbol{\omega}_0)}. \tag{9}$$

Later, because of the existence of explicit MLEs in (1), we show that, by selecting appropriate perturbation schemes, the above defined objective functions are convenient for analysis and yield closed-form expressions which, in our opinion, is an important property.

The result presented in the following auxiliary lemma for the homoscedastic linear model will be used in the subsequent.

Lemma 2.1. *Let us consider the following linear model*

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma_e^2 \boldsymbol{I}_n),$$

where $\boldsymbol{\beta}$ and σ_e^2 are unknown. Suppose that the i-th observation is perturbed so that $\boldsymbol{y}_{P(\omega)} = (\boldsymbol{y}_{[i]}^{\mathrm{T}} : \omega y_i)^{\mathrm{T}}$, $\boldsymbol{X}_{P(\omega)} = (\boldsymbol{X}_{[i]}^{\mathrm{T}} : \omega \mathbf{x}_i)^{\mathrm{T}}$ for some ω belonging to the neighborhood of 1. Then, in the perturbed model, the functions defined by (6) and (7) equal

$$\begin{aligned}
\Delta_i \widehat{\boldsymbol{\beta}} &= \frac{(\omega^2-1)r_i}{(\omega^2-1)h_{ii}+1}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\mathbf{x}_i,\\
\Delta_i D\left[\widehat{\boldsymbol{\beta}}\right] &= \frac{(1-\omega^2)\widehat{\sigma}^2}{(\omega^2-1)h_{ii}+1}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}.
\end{aligned}$$

and the variance ratio of random errors after and before perturbation defined in (9) is

$$VRE_i = 1 + \frac{\omega^2-1}{(\omega^2-1)h_{ii}+1}\frac{r_i^2}{\sum_{i=1}^n r_i^2},$$

where $h_{ij} = \mathbf{x}_i^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\mathbf{x}_j$ denotes the element in the i-th row and the j-th column of the hat matrix $\boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}$ for the unperturbed model, and $r_i = y_i - \widehat{y}_i$ denotes the residual of the i-th observation.

Proof. If the matrices $\boldsymbol{A}$, $\boldsymbol{A}+\boldsymbol{BCD}$ and $\boldsymbol{C}$ are non-singular, according to the inverse binomial theorem, we get

$$(\boldsymbol{A}+\boldsymbol{BCD})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{DA}^{-1}\boldsymbol{B}+\boldsymbol{C}^{-1})^{-1}\boldsymbol{DA}^{-1}.$$

Thus,

$$\begin{aligned}
&\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + (\omega^2-1)\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\right)^{-1}\\
&= (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1} + \frac{1-\omega^2}{(\omega^2-1)h_{ii}+1}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1},
\end{aligned}$$

and the MLE of $\boldsymbol{\beta}$ after the perturbation equals

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}(\omega) &= \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + (\omega^2-1)\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} + (\omega^2-1)y_i\mathbf{x}_i\right)\\
&= \widehat{\boldsymbol{\beta}}(1) + \frac{(\omega^2-1)\left(y_i - \widehat{y}_i\right)}{(\omega^2-1)h_{ii}+1}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\mathbf{x}_i,
\end{aligned}$$

with dispersion matrix

$$
\begin{aligned}
D\left[\widehat{\boldsymbol{\beta}}(\omega)\right] &= \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + (\omega-1)\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}\right)^{-1}\sigma_e^2 \\
&= D\left[\widehat{\boldsymbol{\beta}}(1)\right] + \frac{(1-\omega^2)\sigma_e^2}{(\omega^2-1)h_{ii}+1}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}.
\end{aligned}
$$

The MLE of σ^2 after the perturbation equals

$$
\begin{aligned}
\widehat{\sigma}_e^2(\omega) =& \frac{1}{n}\left(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} + (\omega^2-1)y_i^2\right) - \frac{1}{n}\left(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{X} + (\omega^2-1)y_i\mathbf{x}_i^{\mathrm{T}}\right) \\
& \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + (\omega^2-1)\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} + (\omega^2-1)y_i\mathbf{x}_i\right) \\
=& \widehat{\sigma}_e^2(1) + \frac{(\omega^2-1)r_i^2/n}{(\omega^2-1)h_{ii}+1}.
\end{aligned}
$$

■

When ω approaches 0, the above outcome is identical with the well-known result for case deletion.

2.4. Estimation in cross-over designs

The interest in (1) is usually connected to the mean estimator, $\widehat{\boldsymbol{\beta}}$, especially concerning the treatment effect $\widehat{\phi}$, when σ_γ^2 and σ_e^2 are assumed to be unknown. One important feature of the cross-over design model is that it can be represented as two randomly independent homoscedastic linear models with independent sets of parameters. Therefore, both the maximum likelihood estimators and the dispersion matrices of the estimators can be obtained explicitly. The next lemma presents the explicit MLE of $\boldsymbol{\beta}$ as well as those of σ_γ^2 and σ_e^2 in (1). The model transformation in the proof will be repeated later when the objective functions of influence are calculated in Section 3.

Lemma 2.2. (i) *The MLE of $\boldsymbol{\beta}$ in (1) is given by*

$$
\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \frac{1}{4}(\overline{y}_{1\cdot 1} + \overline{y}_{1\cdot 2}) + \frac{1}{4}(\overline{y}_{2\cdot 1} + \overline{y}_{2\cdot 2}) \\ \frac{1}{2}(\overline{y}_{1\cdot 1} - \overline{y}_{1\cdot 2}) + \frac{1}{2}(\overline{y}_{2\cdot 1} - \overline{y}_{2\cdot 2}) \\ \frac{1}{2}(\overline{y}_{1\cdot 1} - \overline{y}_{1\cdot 2}) - \frac{1}{2}(\overline{y}_{2\cdot 1} - \overline{y}_{2\cdot 2}) \\ (\overline{y}_{1\cdot 1} + \overline{y}_{1\cdot 2}) - (\overline{y}_{2\cdot 1} + \overline{y}_{2\cdot 2}) \end{pmatrix}.
$$

The MLEs of σ_γ^2 and σ_e^2 in (1) equal

$$
\widehat{\sigma}_\gamma^2 = \frac{1}{2n}\sum_{ij} r_{ij1}r_{ij2}, \quad \widehat{\sigma}_e^2 = \frac{1}{4n}\sum_{ij}(r_{ij1} - r_{ij2})^2,
$$

where $\overline{y}_{i\cdot k} = \dfrac{1}{n}\sum_{j=1}^{n} y_{ijk}$ *and* $r_{ijk} = y_{ijk} - \mathbf{x}_{ijk}\widehat{\boldsymbol{\beta}} = y_{ijk} - \overline{y}_{i\cdot k}$ *is the residual, for* $i, k = 1, 2,\ j = 1, 2, \ldots, n.$

(ii) *The dispersion matrix of* $\widehat{\boldsymbol{\beta}}$ *equals*

$$D\left[\widehat{\boldsymbol{\beta}}\right] = \frac{1}{n}\begin{pmatrix} \frac{1}{4}(2\sigma_\gamma^2+\sigma_e^2) & 0 & 0 & 0 \\ 0 & \sigma_e^2 & 0 & 0 \\ 0 & 0 & \sigma_e^2 & 0 \\ 0 & 0 & 0 & 4(2\sigma_\gamma^2+\sigma_e^2) \end{pmatrix}.$$

Proof. The traditional way of how to obtain the estimators of $\boldsymbol{\beta}$, σ_γ^2 and σ_e^2 is demonstrated in Laird *et al.* (1992). We present a proof based on linear models theory. If a full-rank matrix

$$\boldsymbol{T} = \boldsymbol{I}_{2n} \otimes \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{10}$$

is pre-multiplied to both sides of model (1), the following two independent models appear,

$$\begin{cases} \boldsymbol{y}_s = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\eta}_1, \\ \boldsymbol{y}_d = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\eta}_2, \end{cases} \tag{11}$$

where $\boldsymbol{y}_s = (y_{111}+y_{112},\ldots,y_{2n1}+y_{2n2})^{\mathrm{T}}$, $\boldsymbol{y}_d = (y_{111}-y_{112},\ldots,y_{2n1}-y_{2n2})^{\mathrm{T}}$, $\boldsymbol{\beta}_1 = (\mu,\ \lambda)^{\mathrm{T}}$, $\boldsymbol{\beta}_2 = (\pi,\ \phi)^{\mathrm{T}}$, and

$$\boldsymbol{\eta}_1 \sim N_{2n}(\boldsymbol{0}, \sigma_1^2\boldsymbol{I}_{2n}), \quad \boldsymbol{\eta}_2 \sim N_{2n}(\boldsymbol{0}, \sigma_2^2\boldsymbol{I}_{2n}), \quad Cov(\boldsymbol{\eta}_1,\ \boldsymbol{\eta}_2) = \boldsymbol{0},$$

with

$$\sigma_1^2 = 2\sigma_e^2 + 4\sigma_\gamma^2, \quad \sigma_2^2 = 2\sigma_e^2.$$

The design matrices are as follows:

$$\boldsymbol{X}_1 = \begin{pmatrix} 2 & 2 & \cdots & 2 & 2 \\ 1/2 & 1/2 & \cdots & -1/2 & -1/2 \end{pmatrix}^{\mathrm{T}}, \quad \boldsymbol{X}_2 = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & \cdots & -1 & -1 \end{pmatrix}^{\mathrm{T}}.$$

The whole sample obtained from the 2×2 cross-over design can be analyzed by these two independent models with functionally independent means and variance parameters. We have

$$\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{X}_1 = 2n\begin{pmatrix} 4 & 0 \\ 0 & \frac{1}{4} \end{pmatrix}, \quad \left(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{X}_1\right)^{-1} = \frac{1}{2n}\begin{pmatrix} \frac{1}{4} & 0 \\ 0 & 4 \end{pmatrix},$$

$$\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{y}_s = 2n\begin{pmatrix} (\overline{y}_{1\cdot 1}+\overline{y}_{1\cdot 2}) + (\overline{y}_{2\cdot 1}+\overline{y}_{2\cdot 2}) \\ \frac{1}{4}(\overline{y}_{1\cdot 1}+\overline{y}_{1\cdot 2}) - \frac{1}{4}(\overline{y}_{2\cdot 1}+\overline{y}_{2\cdot 2}) \end{pmatrix},$$

and

$$\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{X}_2 = 2n\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \left(\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{X}_2\right)^{-1} = \frac{1}{2n}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{y}_d = 2n\begin{pmatrix} \frac{1}{2}(\overline{y}_{1\cdot 1} - \overline{y}_{1\cdot 2}) + \frac{1}{2}(\overline{y}_{2\cdot 1} - \overline{y}_{2\cdot 2}) \\ \frac{1}{2}(\overline{y}_{1\cdot 1} - \overline{y}_{1\cdot 2}) - \frac{1}{2}(\overline{y}_{2\cdot 1} - \overline{y}_{2\cdot 2}) \end{pmatrix}.$$

For each model, the homoscedastic setup is satisfied. Then, the MLEs of the mean parameters in (11) equal

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_1 &= \left(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{y}_s = \begin{pmatrix} \frac{1}{4}(\overline{y}_{1\cdot 1} + \overline{y}_{1\cdot 2}) + \frac{1}{4}(\overline{y}_{2\cdot 1} + \overline{y}_{2\cdot 2}) \\ (\overline{y}_{1\cdot 1} + \overline{y}_{1\cdot 2}) - \ (\overline{y}_{2\cdot 1} + \overline{y}_{2\cdot 2}) \end{pmatrix}, \\
\widehat{\boldsymbol{\beta}}_2 &= \left(\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{y}_d = \begin{pmatrix} \frac{1}{2}(\overline{y}_{1\cdot 1} - \overline{y}_{1\cdot 2}) + \frac{1}{2}(\overline{y}_{2\cdot 1} - \overline{y}_{2\cdot 2}) \\ \frac{1}{2}(\overline{y}_{1\cdot 1} - \overline{y}_{1\cdot 2}) - \frac{1}{2}(\overline{y}_{2\cdot 1} - \overline{y}_{2\cdot 2}) \end{pmatrix},
\end{aligned}$$

with dispersion matrices

$$\begin{aligned}
D\left[\widehat{\boldsymbol{\beta}}_1\right] &= \left(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{X}_1\right)^{-1}\sigma_1^2 = \frac{2\sigma_\gamma^2 + \sigma_e^2}{n}\begin{pmatrix} \frac{1}{4} & 0 \\ 0 & 4 \end{pmatrix}, \\
D\left[\widehat{\boldsymbol{\beta}}_2\right] &= \left(\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{X}_2\right)^{-1}\sigma_2^2 = \frac{\sigma_e^2}{n}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.
\end{aligned}$$

The MLEs of the variance components in (11) are given by

$$\begin{aligned}
\widehat{\sigma}_1^2 &= \frac{1}{2n}\boldsymbol{y}_s^{\mathrm{T}}\left(\boldsymbol{I}_{2n} - \boldsymbol{X}_1\left(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1^{\mathrm{T}}\right)\boldsymbol{y}_s, \\
\widehat{\sigma}_2^2 &= \frac{1}{2n}\boldsymbol{y}_d^{\mathrm{T}}\left(\boldsymbol{I}_{2n} - \boldsymbol{X}_2\left(\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2^{\mathrm{T}}\right)\boldsymbol{y}_d.
\end{aligned}$$

Since

$$\boldsymbol{X}_1\left(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1^{\mathrm{T}} = \boldsymbol{X}_2\left(\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2^{\mathrm{T}} = \frac{1}{n}\left(\boldsymbol{I}_2 \otimes \boldsymbol{J}_n\right),$$

the MLEs $\widehat{\sigma}_\gamma^2$ and $\widehat{\sigma}_e^2$ in (1) equal

$$\begin{aligned}
\widehat{\sigma}_\gamma^2 = \frac{1}{4}\left(\widehat{\sigma}_1^2 - \widehat{\sigma}_2^2\right) &= \frac{1}{8n}\left(\left(\boldsymbol{y}_s + \boldsymbol{y}_d\right)^{\mathrm{T}}\left(\boldsymbol{I}_{2n} - \frac{1}{n}\left(\boldsymbol{I}_2 \otimes \boldsymbol{J}_n\right)\right)\left(\boldsymbol{y}_s - \boldsymbol{y}_d\right)\right) \\
&= \frac{1}{2n}\sum_{i=1}^{2}\left(\sum_{j=1}^{n} y_{ij1}y_{ij2} - \frac{1}{n}\sum_{j=1}^{n} y_{ij1}\sum_{j=1}^{n} y_{ij2}\right) \\
&= \frac{1}{2n}\sum_{i=1}^{2}\left(\sum_{j=1}^{n} y_{ij1}y_{ij2} - n\overline{Y}_{i.1}\overline{Y}_{i.2}\right) \\
&= \frac{1}{2n}\sum_{i=1}^{2}\sum_{j=1}^{n}\left(y_{ij1} - \overline{Y}_{i.1}\right)\left(y_{ij2} - \overline{Y}_{i.2}\right) \\
&= \frac{1}{2n}\sum_{i=1}^{2}\sum_{j=1}^{n} r_{ij1}r_{ij2},
\end{aligned}$$

and

$$
\begin{aligned}
\widehat{\sigma}_e^2 = \frac{1}{2}\widehat{\sigma}_2^2 &= \frac{1}{4n}\boldsymbol{y}_d^{\mathrm{T}}\left(\boldsymbol{I} - \frac{1}{n}\left(\boldsymbol{I}_2 \otimes \boldsymbol{J}_n\right)\right)\boldsymbol{y}_d \\
&= \frac{1}{4n}\sum_{i=1}^{2}\left(\sum_{j=1}^{n}(y_{ij1} - y_{ij2})^2 - \frac{1}{n}\left(\sum_{j=1}^{n}(y_{ij1} - y_{ij2})\right)^2\right) \\
&= \frac{1}{4n}\sum_{i=1}^{2}\left(\sum_{j=1}^{n}(y_{ij1} - y_{ij2})^2 - n\left(\overline{Y}_{i.1} - \overline{Y}_{i.2}\right)^2\right) \\
&= \frac{1}{4n}\sum_{i=1}^{2}\sum_{j=1}^{n}\left(\left(y_{ij1} - \overline{Y}_{i.1}\right)^2 - \left(y_{ij2} - \overline{Y}_{i.2}\right)^2\right) \\
&= \frac{1}{4n}\sum_{i=1}^{2}\sum_{j=1}^{n}\left(r_{ij1} - r_{ij2}\right)^2 .
\end{aligned}
$$

■

3. Main Results

The algebraic basis of the early work on influential diagnostic, for example in the context of ordinary regression, is that exact formulae are available for the deletion quantity of mean and variance estimates. But such a nice property does not longer hold if there exists dependence within subjects. In most mixed models, analogous formulae can only be obtained by approximations.

However, the cross-over design models can be reformulated in a form where influential quantities for deletion or perturbation have closed-form solutions. The influence of the ij-th subject on the estimates of the fixed effects and the variance parameters is presented in Theorem 3.1.

Theorem 3.1. *Assume that the subject-level case-weighted perturbation scheme $P_2(\omega)$ in (4) is applied to the ij-th subject in model (1). The functions defined by (6) to (9) are in this case as follows:*

(i) *When $i = 1$, i.e., the perturbed subject comes from sequence AB,*

$$
\Delta_{ij}\widehat{\boldsymbol{\beta}} = \frac{\omega^2 - 1}{\omega^2 + n - 1}\begin{pmatrix} \frac{1}{4}(r_{ij1} + r_{ij2}) \\ \frac{1}{2}(r_{ij1} - r_{ij2}) \\ \frac{1}{2}(r_{ij1} - r_{ij2}) \\ (r_{ij1} + r_{ij2}) \end{pmatrix}. \tag{12}
$$

When $i = 2$, i.e., the perturbed subject comes from sequence BA,

$$
\Delta_{ij}\widehat{\boldsymbol{\beta}} = \frac{\omega^2 - 1}{\omega^2 + n - 1}\begin{pmatrix} \frac{1}{4}(r_{ij1} + r_{ij2}) \\ \frac{1}{2}(r_{ij1} - r_{ij2}) \\ -\frac{1}{2}(r_{ij1} - r_{ij2}) \\ -\ (r_{ij1} + r_{ij2}) \end{pmatrix}. \tag{13}
$$

(ii) *When $i=1$, i.e., the perturbed subject comes from sequence AB,*

$$\Delta_{ij} D\left[\widehat{\boldsymbol{\beta}}\right] = \frac{1-\omega^2}{2n(\omega^2+n-1)} \begin{pmatrix} \frac{1}{4}\left(\sigma_e^2+2\sigma_w^2\right) & 0 & 0 & \left(\sigma_e^2+2\sigma_w^2\right) \\ 0 & \sigma_e^2 & \sigma_e^2 & 0 \\ 0 & \sigma_e^2 & \sigma_e^2 & 0 \\ \left(\sigma_e^2+2\sigma_w^2\right) & 0 & 0 & 4\left(\sigma_e^2+2\sigma_w^2\right) \end{pmatrix}. \tag{14}$$

When $i=2$, i.e., the perturbed subject comes from sequence BA,

$$\Delta_{ij} D\left[\widehat{\boldsymbol{\beta}}\right] = \frac{1-\omega^2}{2n(\omega^2+n-1)} \begin{pmatrix} \frac{1}{4}\left(\sigma_e^2+2\sigma_w^2\right) & 0 & 0 & -\left(\sigma_e^2+2\sigma_w^2\right) \\ 0 & \sigma_e^2 & -\sigma_e^2 & 0 \\ 0 & -\sigma_e^2 & \sigma_e^2 & 0 \\ -\left(\sigma_e^2+2\sigma_w^2\right) & 0 & 0 & 4\left(\sigma_e^2+2\sigma_w^2\right) \end{pmatrix}. \tag{15}$$

(iii)

$$\mathrm{VRR}_{ij} = 1 + \frac{(\omega^2-1)n}{\omega^2+n-1} \frac{r_{ij1} r_{ij2}}{\sum_{i=1}^{2}\sum_{j=1}^{n} r_{ij1} r_{ij2}}, \tag{16}$$

$$\mathrm{VRE}_{ij} = 1 + \frac{(\omega^2-1)n}{\omega^2+n-1} \frac{(r_{ij1}-r_{ij2})^2}{\sum_{i=1}^{2}\sum_{j=1}^{n} (r_{ij1}-r_{ij2})^2}, \tag{17}$$

where $r_{ijk} = y_{ijk} - \overline{y}_{i\cdot k}$ is the residual for y_{ijk} in (1), $i=1,2$, $j=1,\ldots,n$, $k=1,2$, and $(r_{ij1}+r_{ij2})$ is independent from $(r_{ij1}-r_{ij2})$, $i=1,2$, $j=1,\ldots,n$.

Proof. Use the matrix $\boldsymbol{T}$ defined in (10) to pre-multiply both sides of the perturbed model. The restriction on the perturbation scheme that the observations within each subject are scaled by the same perturbation weight enables the perturbed model to be splitted into the following two models

$$\begin{cases} \boldsymbol{y}_{s,P_1(\omega)} = \boldsymbol{X}_{1,P_1(\omega)} \boldsymbol{\beta}_1 + \boldsymbol{\eta}_1, \\ \boldsymbol{y}_{d,P_1(\omega)} = \boldsymbol{X}_{2,P_1(\omega)} \boldsymbol{\beta}_2 + \boldsymbol{\eta}_2, \end{cases} \tag{18}$$

where the parameters $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, the random terms $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ have the same meaning as those given in (11). Let us denote the perturbed response variables

$$\boldsymbol{y}_{s,P_1(\omega)} = (\boldsymbol{y}_{s,[ij]}^{\mathrm{T}} : \omega y_{s,ij})^{\mathrm{T}} \text{ and } \boldsymbol{y}_{d,P_1(\omega)} = (\boldsymbol{y}_{d,[ij]}^{\mathrm{T}} : \omega y_{d,ij})^{\mathrm{T}},$$

and the perturbed design matrices

$$\boldsymbol{X}_{1,P_1(\omega)} = (\boldsymbol{X}_{1,[ij]}^{\mathrm{T}} : \omega \mathbf{x}_{1,ij})^{\mathrm{T}} \text{ and } \boldsymbol{X}_{2,P_1(\omega)} = (\boldsymbol{X}_{2,[ij]}^{\mathrm{T}} : \omega \mathbf{x}_{2,ij})^{\mathrm{T}}.$$

It can be shown that the diagonal elements in the hat matrices of the unperturbed models in (11) equal

$$\mathbf{x}_{1,ij}^{\mathrm{T}} (\boldsymbol{X}_1^{\mathrm{T}} \boldsymbol{X}_1)^{-1} \mathbf{x}_{1,ij} = \mathbf{x}_{2,ij}^{\mathrm{T}} (\boldsymbol{X}_2^{\mathrm{T}} \boldsymbol{X}_2)^{-1} \mathbf{x}_{2,ij} = \frac{1}{n}, \text{ for } i=1,2,\ j=1,\ldots,n,$$

according to Lemma 2.1, the MLEs in (18) equal

$$\widehat{\boldsymbol{\beta}}_1(\omega) = \widehat{\boldsymbol{\beta}}_1(1) + \frac{(\omega^2-1)n}{\omega^2+n-1}(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{X}_1)^{-1}\mathbf{x}_{1,ij}\left(y_{s,ij} - \widehat{y}_{s,ij}\right), \tag{19}$$

$$\widehat{\boldsymbol{\beta}}_2(\omega) = \widehat{\boldsymbol{\beta}}_2(1) + \frac{(\omega^2-1)n}{\omega^2+n-1}(\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{X}_2)^{-1}\mathbf{x}_{2,ij}\left(y_{d,ij} - \widehat{y}_{d,ij}\right), \tag{20}$$

with the corresponding dispersion matrices

$$D\left[\widehat{\boldsymbol{\beta}}_1(\omega)\right]=D\left[\widehat{\boldsymbol{\beta}}_1(1)\right]+\frac{(1-\omega^2)n}{\omega^2+n-1}(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{X}_1)^{-1}\mathbf{x}_{1,ij}\mathbf{x}_{1,ij}^{\mathrm{T}}(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{X}_1)^{-1}\sigma_1^2, \tag{21}$$

$$D\left[\widehat{\boldsymbol{\beta}}_2(\omega)\right]=D\left[\widehat{\boldsymbol{\beta}}_2(1)\right]+\frac{(1-\omega^2)n}{\omega^2+n-1}(\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{X}_2)^{-1}\mathbf{x}_{2,ij}\mathbf{x}_{2,ij}^{\mathrm{T}}(\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{X}_2)^{-1}\sigma_2^2, \tag{22}$$

and

$$\widehat{\sigma}_1^2(\omega) = \widehat{\sigma}_1^2(1) + \frac{(\omega^2-1)n}{\omega^2+n-1}\frac{\left(y_{s,ij} - \widehat{y}_{s,ij}\right)^2}{2n}, \tag{23}$$

$$\widehat{\sigma}_2^2(\omega) = \widehat{\sigma}_2^2(1) + \frac{(\omega^2-1)n}{\omega^2+n-1}\frac{\left(y_{d,ij} - \widehat{y}_{d,ij}\right)^2}{2n}, \tag{24}$$

where $\widehat{y}_{s,ij} = \overline{y}_{i\cdot 1} + \overline{y}_{i\cdot 2}$ and $\widehat{y}_{d,ij} = \overline{y}_{i\cdot 1} - \overline{y}_{i\cdot 2}$ are the predictions of $y_{s,ij}$ and $y_{d,ij}$ for the unperturbed model in the given order, $j = 1, 2, \ldots, n$. It should be noted that $r_{ij1} + r_{ij2} = y_{s,ij} - \widehat{y}_{s,ij}$ and $r_{ij1} - r_{ij2} = y_{d,ij} - \widehat{y}_{d,ij}$ can be interpreted as the residuals in the terms of the two independent models in (11).

Subtracting (24) from (23) and using that $\widehat{\sigma}_1^2 = 4\widehat{\sigma}_\gamma^2 + 2\widehat{\sigma}_e^2$ and $\widehat{\sigma}_2^2 = 2\widehat{\sigma}_e^2$, we get

$$\begin{aligned}
\widehat{\sigma}_\gamma^2(\omega) &= \widehat{\sigma}_\gamma^2(1) + \frac{\omega^2-1}{8(\omega^2+n-1)}\left((r_{ij1}+r_{ij2})^2 - (r_{ij1}-r_{ij2})^2\right)\\
&= \widehat{\sigma}_\gamma^2(1) + \frac{\omega^2-1}{2(\omega^2+n-1)}r_{ij1}r_{ij2},\\
\widehat{\sigma}_e^2(\omega) &= \widehat{\sigma}_e^2(1) + \frac{\omega^2-1}{4(\omega^2+n-1)}(r_{ij1}-r_{ij2})^2.
\end{aligned}$$

According to Lemma 2.2,

$$\widehat{\sigma}_\gamma^2(1) = \frac{1}{2n}\sum_{i=1}^{2}\sum_{j=1}^{n} r_{ij1}r_{ij2} \text{ and } \widehat{\sigma}_e^2(1) = \frac{1}{4n}\sum_{i=1}^{2}\sum_{j=1}^{n}(r_{ij1}-r_{ij2})^2,$$

explicit solutions of VRR_{ij} and VRE_{ij} are obtained. Next, replacing the vectors in (19) to (22) with

$$(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{X}_1)^{-1}\mathbf{x}_{1,1j} = \frac{1}{2n}\begin{pmatrix}\frac{1}{2} & 2\end{pmatrix}^{\mathrm{T}}, \quad (\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{X}_1)^{-1}\mathbf{x}_{1,2j} = \frac{1}{2n}\begin{pmatrix}\frac{1}{2} & -2\end{pmatrix}^{\mathrm{T}},$$

$$(\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{X}_2)^{-1}\mathbf{x}_{2,1j} = \frac{1}{2n}\begin{pmatrix}1 & 1\end{pmatrix}^{\mathrm{T}}, \quad (\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{X}_2)^{-1}\mathbf{x}_{2,2j} = \frac{1}{2n}\begin{pmatrix}1 & -1\end{pmatrix}^{\mathrm{T}},$$

the proof is completed. ■

Theorem 3.1 (i) and (iii) show that influence of each subject on the MLEs of the parameters is decided only by its sequence and the residual of the unperturbed model. However, when random individual effects are assumed, the influence of subjects is not presented by $\{r_{ijk}\}$ but by its two independent decomposition $\left\{\frac{1}{2}(r_{ij1}+r_{ij2})\right\}$ and $\left\{\frac{1}{2}(r_{ij1}-r_{ij2})\right\}$. The subject with largest $\{r_{ij1}+r_{ij2}\}$ has greatest influence on the estimation of μ and λ, while the subject with largest $\{r_{ij1}-r_{ij2}\}$ has greatest influence on the estimation of π and ϕ. The influences on σ_γ^2 and σ_e^2 are also given. Additionally, Theorem 3.1 (ii) indicates that, when the variance parameters are known, the influences on the dispersion matrix of the MLEs of the mean parameters are identical for all the subjects within the same sequence. When variance parameters are unknown, vaguely speaking, large influence on the estimation of the dispersion matrix only happens when estimates of the variance parameters change much according to the perturbation.

Theorem 3.2. *Under the perturbed model (2), the delta-beta statistic $\Delta_{ij}\widehat{\boldsymbol{\beta}}$ obtained in Theorem 3.1 follows a multivariate normal distribution with expectation*

$$E\left[\Delta_{ij}\widehat{\boldsymbol{\beta}}\right]=\mathbf{0}.$$

The dispersion matrix of $\Delta_{ij}\widehat{\boldsymbol{\beta}}$ is singular of rank 2. If the perturbed subject belongs to sequence AB, then

$$D\left[\Delta_{ij}\widehat{\boldsymbol{\beta}}\right]=\frac{(n-1)(\omega^2-1)^2}{2n^2\omega^2(\omega^2+n-1)}\begin{pmatrix}\frac{1}{4}\left(\sigma_e^2+2\sigma_w^2\right) & 0 & 0 & \left(\sigma_e^2+2\sigma_w^2\right)\\ 0 & \sigma_e^2 & \sigma_e^2 & 0\\ 0 & \sigma_e^2 & \sigma_e^2 & 0\\ \left(\sigma_e^2+2\sigma_w^2\right) & 0 & 0 & 4\left(\sigma_e^2+2\sigma_w^2\right)\end{pmatrix},$$

and if the perturbed subject belongs to sequence BA, then

$$D\left[\Delta_{ij}\widehat{\boldsymbol{\beta}}\right]=\frac{(n-1)(\omega^2-1)^2}{2n^2\omega^2(\omega^2+n-1)}\begin{pmatrix}\frac{1}{4}\left(\sigma_e^2+2\sigma_w^2\right) & 0 & 0 & -\left(\sigma_e^2+2\sigma_w^2\right)\\ 0 & \sigma_e^2 & -\sigma_e^2 & 0\\ 0 & -\sigma_e^2 & \sigma_e^2 & 0\\ -\left(\sigma_e^2+2\sigma_w^2\right) & 0 & 0 & 4\left(\sigma_e^2+2\sigma_w^2\right)\end{pmatrix}.$$

Proof. It was proven in Theorem 3.1 that

$$\Delta_{ij}\widehat{\boldsymbol{\beta}}=\frac{\omega^2-1}{\omega^2+n-1}\boldsymbol{L}\begin{pmatrix} r_{ij1}+r_{ij2}\\ r_{ij1}-r_{ij2}\end{pmatrix},$$

where

$$\boldsymbol{L}=\begin{pmatrix}\frac{1}{4} & 0 & 0 & 1\\ 0 & \frac{1}{2} & \frac{1}{2} & 0\end{pmatrix}^T,\ \text{if } i=1,$$

$$\boldsymbol{L}=\begin{pmatrix}\frac{1}{4} & 0 & 0 & -1\\ 0 & \frac{1}{2} & -\frac{1}{2} & 0\end{pmatrix}^T,\ \text{if } i=2.$$

The statements of this theorem can be easily obtained by the distributions of the two independent parts $(r_{ij1} - r_{ij2})$ and $(r_{ij1} - r_{ij2})$. Since $r_{ijk} = y_{ijk} - \overline{y}_{i\cdot k}$,

$$r_{ij1} + r_{ij2} = \frac{n-1}{n}(y_{ij1} + y_{ij2}) - \frac{1}{n} \sum_{\substack{1 \le l \le n \\ l \ne j}} (y_{il1} + y_{il2}).$$

The perturbation $P_2(\omega)$ with respect to the ij-th subject implies the following distributions under model assumption (2),

$$(y_{il1} + y_{il2}) \sim \begin{cases} N\left(\mathbf{x}_{ij1}\boldsymbol{\beta} + \mathbf{x}_{ij2}\boldsymbol{\beta}, \dfrac{2\sigma_e^2 + 4\sigma_\gamma^2}{\omega^2}\right), & \text{for } l = j, \\ N\left(\mathbf{x}_{ij1}\boldsymbol{\beta} + \mathbf{x}_{ij2}\boldsymbol{\beta},\ 2\sigma_e^2 + 4\sigma_\gamma^2\right), & \text{for } l \ne j. \end{cases}$$

Since $(y_{i11} + y_{i12}), (y_{i21} + y_{i22}), \ldots, (y_{in1} + y_{in2})$ are independent, it follows that $(r_{i11} + r_{i12})$ is normally distributed with expectation

$$E(r_{ij1} + r_{ij2}) = \frac{n-1}{n}(\mathbf{x}_{ij1}\boldsymbol{\beta} + \mathbf{x}_{ij2}\boldsymbol{\beta}) - \frac{1}{n} \sum_{\substack{1 \le l \le n \\ l \ne j}} (\mathbf{x}_{ij1}\boldsymbol{\beta} + \mathbf{x}_{ij2}\boldsymbol{\beta}) = 0,$$

and variance

$$\begin{aligned} Var(r_{ij1} + r_{ij2}) &= \left(\frac{n-1}{n}\right)^2 \frac{2\sigma_e^2 + 4\sigma_\gamma^2}{\omega^2} + \frac{1}{n} \sum_{\substack{1 \le l \le n \\ l \ne j}} (2\sigma_e^2 + 4\sigma_\gamma^2) \\ &= \frac{2(n-1)(\omega^2 + n - 1)(\sigma_e^2 + 2\sigma_\gamma^2)}{n^2\omega^2}. \end{aligned}$$

Similarly, we can find that the distribution of $(r_{ij1} - r_{ij2})$ is giving by

$$(r_{ij1} - r_{ij2}) \sim N\left(0, \frac{2(n-1)(\omega^2 + n - 1)\sigma_e^2}{n^2\omega^2}\right),$$

which completes the proof of the theorem. ■

It is interesting to observe that, in the 2×2 cross-over designs, the matrices $\Delta_{ij} D\left[\widehat{\boldsymbol{\beta}}\right]$ and $D\left[\Delta_{ij}\widehat{\boldsymbol{\beta}}\right]$ have the same structure, which is stated in the next corollary.

Corollary 3.3. *The influence on the dispersion matrix of the mean estimate and the dispersion matrix of the influence on the mean estimator are proportional, according to the perturbation $P_2(\omega)$:*

$$D\left[\Delta_{ij}\widehat{\boldsymbol{\beta}}\right] = c_n c_\omega \Delta_{ij} D\left[\widehat{\boldsymbol{\beta}}\right],$$

where $c_n = (n-1)/n$ and $c_\omega = (1-\omega^2)/\omega^2$ are two scalars.

Until now, the subject-level perturbation $P_2(\omega)$ defined by (4) was made so that the observations within the same subject were perturbed with the same scalar. As mentioned in the proof of Theorem 3.1, one important benefit by doing this is that $P_2(\omega)$ keeps basic estimation properties of the unperturbed cross-over design models to their perturbed versions. In other words, model (2) can be transformed into two independent homoscedastic linear models as well, which then also results in explicit MLEs. However, such a perturbation scheme is not the only way to have explicit estimations. In the following two theorems, the perturbation $P_3(\boldsymbol{W})$ defined by (5) is considered. The response vectors $\boldsymbol{y}_s$ and $\boldsymbol{y}_d$ in (11) are perturbed via a scaling with different quantities, which suggests an alternative perturbation scheme for cross-over designs.

Theorem 3.4. *Suppose that the linear-combination perturbation scheme $P_3(\boldsymbol{W})$ in (5) with perturbation weight*

$$\boldsymbol{W} = \begin{pmatrix} w_1 & w_2 \\ w_3 & w_4 \end{pmatrix}$$

is applied to the ij-th subject in (1). The perturbed model (2) can be splitted into two independent homoscedastic linear models if and only if $\omega_3 = \omega_2$ and $\omega_4 = \omega_1$. A null perturbation exists at $\mathbf{W}_0 = \boldsymbol{I}_2$.

Proof. We prove sufficiency and necessity separately.

Sufficiency. Let $\omega_3 = \omega_2,\ \omega_4 = \omega_1$. By pre-multiplying with the same matrix (10), the perturbed model is obviously equivalent to two independent homoscedastic linear models given by

$$\begin{cases} \boldsymbol{y}_{s,P_1(\delta_1)} = \mathbf{X}_{1,P_1(\delta_1)}\boldsymbol{\beta}_1 + \boldsymbol{\eta}_1, \\ \boldsymbol{y}_{d,P_1(\delta_2)} = \mathbf{X}_{2,P_1(\delta_2)}\boldsymbol{\beta}_2 + \boldsymbol{\eta}_2, \end{cases} \tag{25}$$

where

$$\boldsymbol{y}_{s,P_1(\delta_1)} = (\boldsymbol{y}_{s,[ij]}^{\mathrm{T}} : \delta_1 y_{s,ij})^{\mathrm{T}}, \qquad \mathbf{X}_{1,P_1(\delta_1)} = (\boldsymbol{X}_{1,[ij]}^{\mathrm{T}} : \delta_1 \mathbf{x}_{1,ij})^{\mathrm{T}},$$
$$\boldsymbol{y}_{d,P_1(\delta_2)} = (\boldsymbol{y}_{d,[ij]}^{\mathrm{T}} : \delta_2 y_{d,ij})^{\mathrm{T}}, \qquad \mathbf{X}_{2,P_1(\delta_2)} = (\boldsymbol{X}_{2,[ij]}^{\mathrm{T}} : \delta_2 \mathbf{x}_{2,ij})^{\mathrm{T}},$$

and $\delta_1 = \omega_1 + \omega_2,\ \delta_2 = \omega_1 - \omega_2$.

Necessity. If the perturbed model (2) can be splitted into two independent homoscedastic linear models, the MLEs of (2) have explicit representations.

Let $\mathbf{\Sigma} = \sigma_\gamma^2 \boldsymbol{I}_{2n} \otimes \boldsymbol{J}_2 + \sigma_e^2 \boldsymbol{I}_{4n}$ denote the covariance matrix of the vector of the response variable in (2). According to Theorem 2 and 5 in Szatrowski (1980), the explicit MLEs exist only if the column space $\mathcal{C}\left(\boldsymbol{X}_{P(\mathbf{W})}\right)$ is $\mathbf{\Sigma}$-invariant, i.e.,

$$\mathcal{C}\left(\boldsymbol{X}_{P(\mathbf{W})}\right) = \mathcal{C}\left(\mathbf{\Sigma}\boldsymbol{X}_{P(\mathbf{W})}\right).$$

Note that the matrix $\boldsymbol{T}$ in (10) is constructed by orthogonal eigenvectors of $\mathbf{\Sigma}$ corresponding to the two distinct eigenvalues $\lambda_1 = \sigma_e^2 + 2\sigma_\gamma^2$ and $\lambda_2 = \sigma_e^2$ of multiplicity $2n$. The following matrix equation holds for $\mathbf{\Sigma}$,

$$\boldsymbol{T}\mathbf{\Sigma} = \mathbf{\Lambda}\boldsymbol{T},$$

where

$$\boldsymbol{T} = \boldsymbol{I}_{2n} \otimes \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \text{ and } \boldsymbol{\Lambda} = \boldsymbol{I}_{2n} \otimes \begin{pmatrix} \sigma_e^2 + 2\sigma_\gamma^2 & 0 \\ 0 & \sigma_e^2 \end{pmatrix}.$$

Therefore, the perturbed model always satisfies

$$\mathcal{C}(\boldsymbol{T}\boldsymbol{X}_{P(\boldsymbol{W})}) = \mathcal{C}(\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{X}_{P(\boldsymbol{W})}) = \mathcal{C}(\boldsymbol{\Lambda}\boldsymbol{T}\boldsymbol{X}_{P(\boldsymbol{W})}),$$

where $\boldsymbol{T}\boldsymbol{X}_{P(\boldsymbol{W})}$ and $\boldsymbol{\Lambda}\boldsymbol{T}\boldsymbol{X}_{P(\boldsymbol{W})}$ are $4n \times 4$ matrices. The matrix $\boldsymbol{T}\boldsymbol{X}_{P(\boldsymbol{W})}$ is $\boldsymbol{\Lambda}$-invariant. Since the perturbation $P_3(\boldsymbol{W})$ in (5) is applied to the 2×2 cross-over design, when $i = 1$,

$$\boldsymbol{T}\boldsymbol{X}_{P_3(\boldsymbol{W})} = \begin{pmatrix} 2 & 0 & \cdots & 2 & 0 & \omega_1+\omega_2+\omega_3+\omega_4 & \omega_1+\omega_2-\omega_3-\omega_4 \\ 0 & 1 & \cdots & 0 & 1 & \frac{1}{2}(\omega_1-\omega_2+\omega_3-\omega_4) & \frac{1}{2}(\omega_1-\omega_2-\omega_3+\omega_4) \\ 0 & 1 & \cdots & 0 & -1 & \frac{1}{2}(\omega_1-\omega_2+\omega_3-\omega_4) & \frac{1}{2}(\omega_1-\omega_2-\omega_3+\omega_4) \\ \frac{1}{2} & 0 & \cdots & -\frac{1}{2} & 0 & \frac{1}{4}(\omega_1+\omega_2+\omega_3+\omega_4) & \frac{1}{4}(\omega_1+\omega_2-\omega_3-\omega_4) \end{pmatrix}^{\mathrm{T}},$$

and

$$\boldsymbol{\Lambda}\boldsymbol{T}\boldsymbol{X}_{P_3(\boldsymbol{W})} = \begin{pmatrix} 2\lambda_1 & 0 \cdots & 2\lambda_1 & 0 & \lambda_1(\omega_1+\omega_2+\omega_3+\omega_4) & \lambda_2(\omega_1+\omega_2-\omega_3-\omega_4) \\ 0 & \lambda_2 \cdots & 0 & \lambda_2 & \frac{\lambda_1}{2}(\omega_1-\omega_2+\omega_3-\omega_4) & \frac{\lambda_2}{2}(\omega_1-\omega_2-\omega_3+\omega_4) \\ 0 & \lambda_2 \cdots & 0 & -\lambda_2 & \frac{\lambda_1}{2}(\omega_1-\omega_2+\omega_3-\omega_4) & \frac{\lambda_2}{2}(\omega_1-\omega_2-\omega_3+\omega_4) \\ \frac{\lambda_1}{2} & 0 \cdots & -\frac{\lambda_1}{2} & 0 & \frac{\lambda_1}{4}(\omega_1+\omega_2+\omega_3+\omega_4) & \frac{\lambda_2}{4}(\omega_1+\omega_2-\omega_3-\omega_4) \end{pmatrix}^{\mathrm{T}},$$

The equality of column spaces holds if and only if

$$\begin{cases} \omega_1+\omega_2-\omega_3-\omega_4 = 0, \\ \omega_1-\omega_2+\omega_3-\omega_4 = 0, \end{cases}$$

that is, $\omega_3 = \omega_2$ and $\omega_4 = \omega_1$. The same restriction is obtained when $i = 2$. ■

Theorem 3.5. *Assume that the perturbation scheme $P_3(\boldsymbol{W})$ under the restriction in Theorem 3.4 holds. Denote*

$$c_1 = \frac{\omega_1^2 + 2\omega_1\omega_2 + \omega_2^2 - 1}{\omega_1^2 + 2\omega_1\omega_2 + \omega_2^2 + n - 1} \text{ and } c_2 = \frac{\omega_1^2 - 2\omega_1\omega_2 + \omega_2^2 - 1}{\omega_1^2 - 2\omega_1\omega_2 + \omega_2^2 + n - 1}.$$

The functions defined by (6) to (9) are given as follows.

(i) *When i=1, i.e., the perturbed subject comes from sequence AB,*

$$\Delta_{ij}\widehat{\boldsymbol{\beta}} = c_1(r_{ij1} + r_{ij2})\begin{pmatrix} \frac{1}{4} \\ 0 \\ 0 \\ 1 \end{pmatrix} + c_2(r_{ij1} - r_{ij2})\begin{pmatrix} 0 \\ \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{pmatrix}.$$

When i=2, i.e., the perturbed subject comes from sequence BA,

$$\Delta_{ij}\widehat{\boldsymbol{\beta}} = c_1(r_{ij1}+r_{ij2})\begin{pmatrix} \frac{1}{4} \\ 0 \\ 0 \\ -1 \end{pmatrix} + c_2(r_{ij1}-r_{ij2})\begin{pmatrix} 0 \\ \frac{1}{2} \\ -\frac{1}{2} \\ 0 \end{pmatrix}.$$

(ii) *When $i = 1$, i.e., the perturbed subject comes from sequence AB,*

$$\Delta_{ij}D\left[\widehat{\boldsymbol{\beta}}\right] = -\frac{c_1\left(\sigma_e^2+2\sigma_\gamma^2\right)}{2n}\begin{pmatrix} \frac{1}{4} & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 4 \end{pmatrix} - \frac{c_2\sigma_e^2}{2n}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

When $i = 2$, i.e., the perturbed subject comes from sequence BA,

$$\Delta_{ij}D\left[\widehat{\boldsymbol{\beta}}\right] = -\frac{c_1\left(\sigma_e^2+2\sigma_\gamma^2\right)}{2n}\begin{pmatrix} \frac{1}{4} & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 4 \end{pmatrix} - \frac{c_2\sigma_e^2}{2n}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

(iii)

$$\mathrm{VRR}_{ij} = 1 + \frac{c_1 n(r_{ij1}+r_{ij2})^2}{4\sum_{i=1}^{2}\sum_{j=1}^{n} r_{ij1}r_{ij2}} - \frac{c_2 n(r_{ij1}-r_{ij2})^2}{4\sum_{i=1}^{2}\sum_{j=1}^{n} r_{ij1}r_{ij2}},$$

$$\mathrm{VRE}_{ij} = 1 + \frac{c_2 n(r_{ij1}-r_{ij2})^2}{\sum_{i=1}^{2}\sum_{j=1}^{n}(r_{ij1}-r_{ij2})^2}.$$

Proof. The proof is similar to that of Theorem 3.1. ■

4. Illustrative example

Theorem 3.1 and 3.5 suggest that, when the 2×2 cross-over design modelling with individual random effect is assumed, the residuals $\{r_{ij1}+r_{ij2}\}$ and $\{r_{ij1}-r_{ij2}\}$ contain all the information to detect influential cases. We illustrate the obtained result through a simulation study. This analysis uses the delta-beta-based local influence and is compared with the case-deletion diagnostics computed by the MIXED Procedure in SAS/STAT® software. For details about the influential diagnostics in the MIXED Procedure, see Littell *et al.* (2006).

An artificial dataset for the 2×2 cross-over design is generated based on (1) with the following true values of parameters,

$$\boldsymbol{\beta} = \begin{pmatrix} \mu \\ \pi \\ \phi \\ \lambda \end{pmatrix} = \begin{pmatrix} 40 \\ -5 \\ -10 \\ 20 \end{pmatrix} \text{ and } \boldsymbol{\theta} = \begin{pmatrix} \sigma_\gamma^2 \\ \sigma_e^2 \end{pmatrix} = \begin{pmatrix} 100 \\ 50 \end{pmatrix}.$$

The dataset contains 10 subjects for both sequence AB and BA and thus, there are 40 observations in total. The likelihood distance (Cook and Weisberg, 1982) suggests that model (1) has no obvious influential observations in the original simulated dataset. In order to verify whether our suggested residuals $\{r_{ij1}+r_{ij2}\}$ and $\{r_{ij1}-r_{ij2}\}$ can detect influential cases effectively, we contaminate three subjects.

The contaminated subjects are selected at random and constructed in the following different ways. Firstly, subject 4 within sequence AB is contaminated by adding 30 to both of its responses, $y_{1,4,1}$ and $y_{1,4,2}$. Secondly, 30 and -30 are added to the responses on the first and the second period of subject 5 within sequence BA, $y_{2,5,1}$ and $y_{2,5,2}$, respectively. Thirdly, we add 60 to the first-period response of subject 10 within sequence AB, $y_{1,10,1}$. The next step is to figure out whether the case-deletion diagnostics and the delta-beta-based local influence are able to identify them.

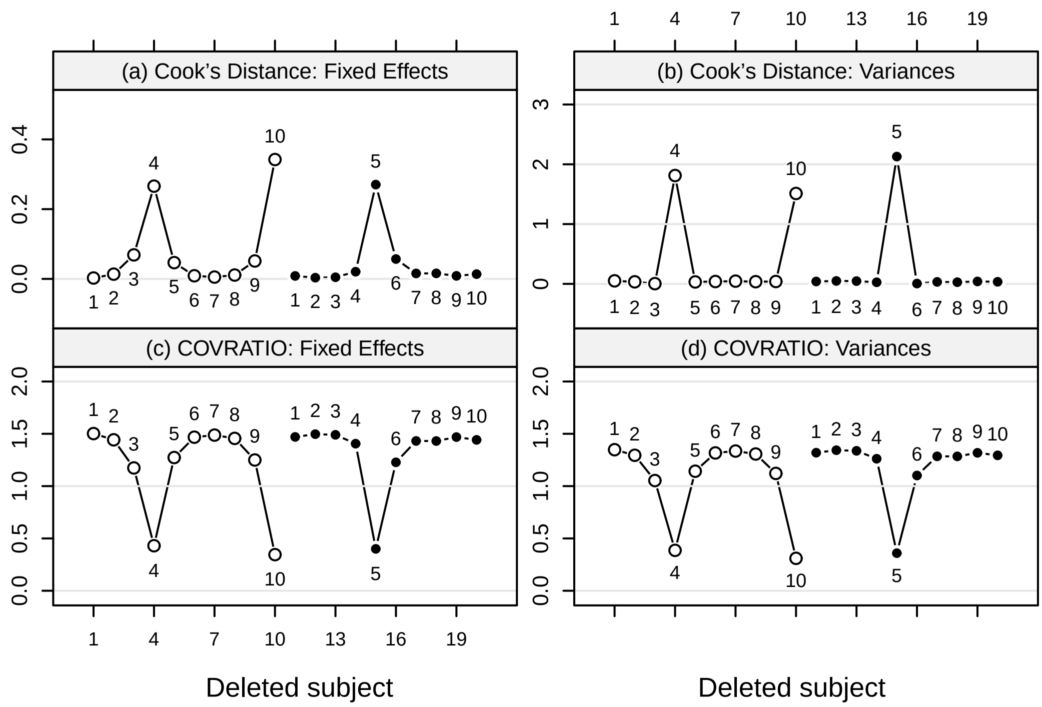

Figure 1: *The subject-level case-deletion influence diagnostics. In each graph, the hollow and solid circles represent subjects from sequence AB and sequence BA, respectively. (a) and (b) Cook's Distances for fixed effects and variance parameters measure influences on the corresponding parameter estimates. (c) and (d) The COVRATIO statistics for fixed effects and variance parameters measure influences on the dispersion matrices of the corresponding parameter estimates.*

Figure 1 illustrates main influence diagnostics with subject-level case deletions. Cook's Distance (Cook, 1977) and the COVRATIO statistic (Belsley *et al.*, 2004) are available by recursive algorithms or approximation methods, e.g. the ridge-stabilised Newton-Raphson algorithm in the MIXED Procedure Littell *et al.* (2006)

or the one-step approximation suggested by Christensen *et al.* (1992). Clearly, Figure 1 indicates that subjects 4 and 10 within sequence AB and subject 5 within sequence BA have larger influences on the estimates and the precision of estimates of both the mean and the variance parameters. The case-deletion approach correctly discovers all of the contaminated subjects. However, more detailed influence of the subjects on each parameter are masked from the case-deletion measurements and requires the recursive algorithms to evaluate the MLEs in the case-deleted model.

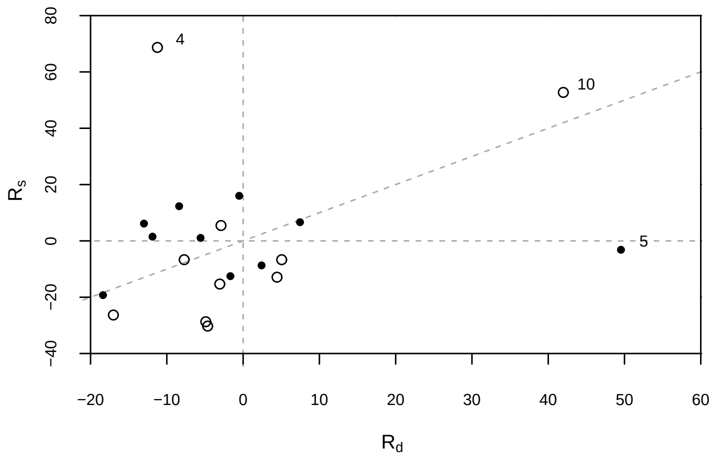

Figure 2: *The sums of the residuals within the subjects* $R_s = \{r_{ij1} + r_{ij2}\}$ *versus the differences* $R_d = \{r_{ij1} - r_{ij2}\}$ *in the unperturbed model. Each hollow or solid circle represents one subject from sequence AB or from sequence BA, respectively.*

By contrast, the delta-beta-based local influence is computationally cheap and provides exact solutions. The conclusion of Section 3 implies that the location of the subjects in Figure 2 indicates their distinct features on the parameter estimation. Theorem 3.1 and 3.5 yield that

(i) The subjects deviating along the "sum axis" R_s, i.e. subject 4 and subject 10 within sequence AB in this case, present great influence on the general mean μ and the carry-over effect λ.

(ii) The subjects deviating along the "difference axis" R_d, i.e. subject 10 within sequence AB and subject 5 within sequence BA, present great influence on the period effect π, the treatment effect ϕ and the variance of errors σ_e^2.

(iii) The subjects close to the diagonal of the sum and difference axes have small influence on the variance of the random effects, σ_γ^2. So although subject 10 in sequence AB is influential with respect to the joint parameters as case deletion methods suggests, its influence on σ_γ^2 is limited.

Let us do one more step. We calculate and plot the proposed $\Delta_{ij}\widehat{\boldsymbol{\beta}}$, VRR_{ij} and VRE_{ij}, given in (12)-(17). Figure 3 displays the statistic $\Delta_{ij}\widehat{\boldsymbol{\beta}}$, which presents the influence on the each mean parameter estimates. Figure 4 is the plot of the

VRR and VRE influences, which contains the information of the influence on σ_γ^2 and σ_e^2. The outcome confirms the three statements observed from the residuals. Consequently, the delta-beta-based local influence approach provides an effective method to detect influential observations, and to identify their roles on influence in the cross-over models with random individual effects.

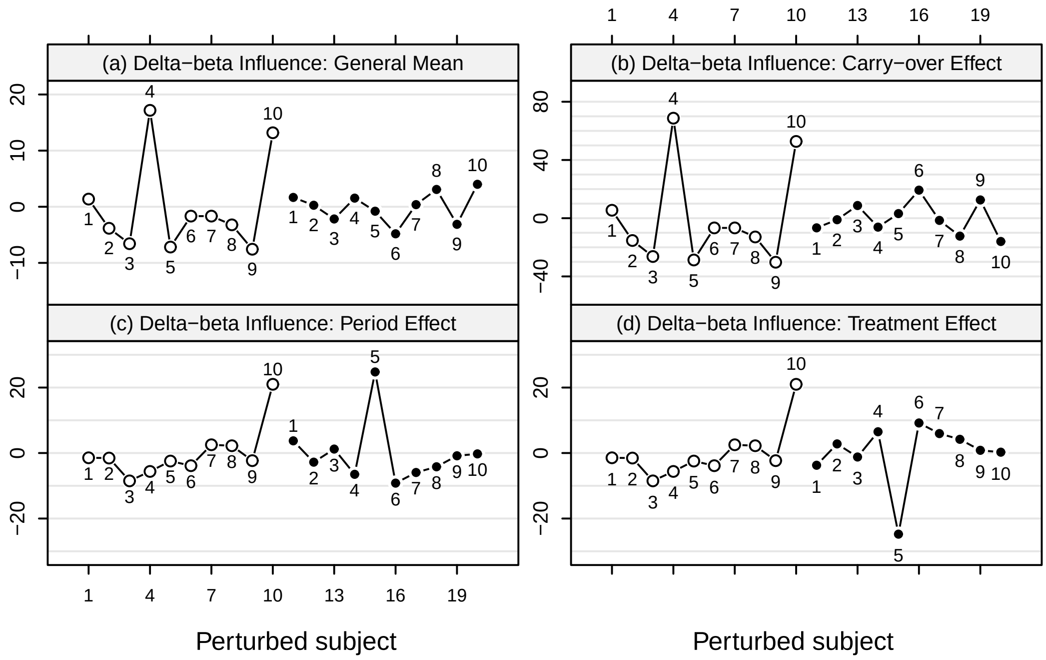

Figure 3: *Influence of the case-weighted perturbations on the mean parameters. In each graph, the hollow and solid circles represent subjects from sequence AB and sequence BA, respectively. The perturbation weight* ω *is* 0.9.

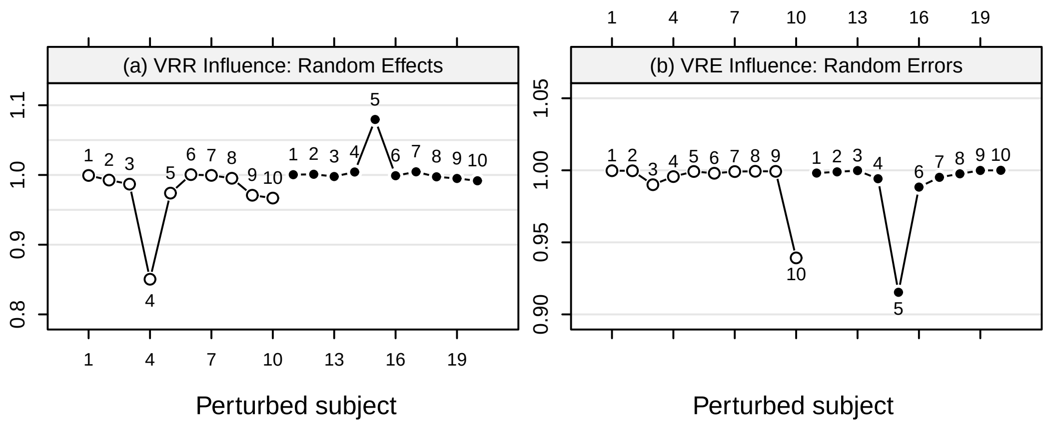

Figure 4: *Influence of the case-weighted perturbations on the variance parameters. In each graph, the hollow and solid circles represent subjects from sequence AB and sequence BA, respectively. The perturbation weight* ω *is* 0.9.

5. Conclusions and Remarks

Statistical models play an important role when making practical decisions in a wide number of fields. As a consequence, it follows an urgent task to investigate sensitivity of the used model formulations. Our work have studied sensitivity of the mixed linear model for the balanced 2×2 cross-over design. Influences of observations on the MLEs and the estimating precisions are surveyed, based on simple but powerful diagnostic tools, and $\Delta_{ij}\widehat{\boldsymbol{\beta}}$ under case-weighted perturbations has been derived. The results suggest data users to survey the within-subject sums and differences of residuals of the unperturbed model, i.e, $\{r_{ij1} + r_{ij2}\}$ and $\{r_{ij1} - r_{ij2}\}$ in (1), to identify influential observations for parameters of interest.

For several reasons, this article does not provide any result on diagnostic criteria. One main reason is that there is no precise or operational definition of influential observation in the existing literature on robustness and diagnostics. The choice of what to identify as an influential observation should depend on what character of data and problems we are looking at, not on a universal numerical rule. Moreover, as Shi (2006) argues, the true distributions of influence quantities are often difficult to know. Although Theorem 3.2 presents the distribution of $\Delta_{ij}\widehat{\boldsymbol{\beta}}$, it assumes the perturbed cases normal distributed. Normal assumption may not be satisfied, particularly, if the perturbed case is a prospective influential observation.

Although more complex and generalised diagnostic tools are available, our explicit approach have direct interpretations in term of the effects on interested parameters, see Figure 3 and 4. More important, we believe that our proposed approach provides a theoretical interpretation of the structures of influence in the cross-over design models. The mixed linear model has explicit MLEs in cross-over design, and our delta-beta-based local influence approach has exploited this fact.

The proposed approach can quickly be extended to other mixed models with explicit MLEs. It directs our further attention to complicated cross-over design models, for example, with different covariance structures or with repeated measurements. Influence on maximum likelihood predictions of random effects and alternative case-weighted perturbations, which have not been covered by this work, are also points to further studies.

References

Afsarinejad, K. (1990). Repeated measurements designs - a review. *Communications in Statistics: Theory and Methods*, **19**, 3985–4028.

Banerjee, M. and Frees, E. W. (1997). Influence diagnostics for linear longitudinal models. *Journal of the American Statistical Association*, **92**, 999–1005.

Beckman, R. J., Nachtsheim, C. and Cook, R. D. (1987). Diagnostics for mixed-model analysis of variance. *Technometrics*, **29**, 413–426.

Belsley, D. A., Kuh, E. and Welsch, R. E. (2004). *Regression diagnostics: Identifying influential data and sources of collinearity.* John Wiley & Sons, Inc., Hoboken, New Jersey.

Chen, F., Zhu, H.-T., Song, X.-Y. and Lee, S.-Y. (2010). Perturbation selection and local influence analysis for generalized linear mixed models. *Journal of Computational and Graphical Statistics*, **19**, 826–842.

Christensen, R., Pearson, L. M. and Johnson, W. (1992). Case-deletion diagnostics for mixed models. *Technometrics*, **34**, 38–45.

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

Cook, R. D. (1986). Assessment of local influence. *Jounal of the Royal Statistical Society: Series B*, **48**, 133–169.

Cook, R. D. and Weisberg, S. (1982). *Residuals and influence in regression.* Chapman & Hall, New York.

Demidenko, E. and Stukel, T. A. (2005). Influence analysis for linear mixed-effects models. *Statistics in Medicine*, **24**, 893–909.

Grizzle, J. E. (1965). The two-period change-over design and its use in clinical trials. *Biometrics*, **21**, 467–480.

Haslett, J. and Dillane, D. (2004). Application of "delete = replace" to deletion diagnostics for variance component estimation in the linear mixed model. *Journal of the Royal Statistical Society: Series B*, **66**, 131–143.

Hodges, J. S. (1998). Some algebra and geometry for hierarchical models, applied to diagnostics. *Journal of the Royal Statistical Society: Series B*, **60**, 497–536.

Jones, B. and Kenward, M. G. (1989). *Design and analysis of cross-over trials.* Chapman & Hall, London and New York.

Laird, N. M., Skinner, J. and Kenward, M. (1992). An analysis of two-period crossover designs with carry-over effects. *Statistics in Medicine*, **11**, 1967–1979.

Lesaffre, E. and Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, **54**, 570–582.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D. and Schabenberber, O. (2006). *SAS for mixed models.* SAS Institute Inc., Cary, NC, USA, 2nd edn.

Ouwens, M. J. N., Tan, F. E. S. and Berger, M. P. F. (2001). Local influence to detect influential data structures for generalized linear mixed models. *Biometrics*, **57**, 1166–1172.

Pan, J.-X. and Fang, K.-T. (2002). *Growth curve models and statistical diagnostics.* Springer-Verlag, New York.

Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics*, **9**, 705–724.

von Rosen, D. (1995). Influential observations in multivariate linear models. *Scandinavian Journal of Statistics*, **22**, 207–222.

Schabenberger, O. (2005). Mixed model influence diagnostics. In *Proceedings of the 29th annual sas users group international conference*, Paper 189-29.

Senn, S. (2006). Cross-over trials in statistics in medicine: The first 25 years. *Statistics in Medicine*, **25**, 3430–3442.

Shi, L. (2006). *Multilevel diagnostics for mixed and hierarchical linear model.* Ph.D. thesis, Department of Mathematics and Statistics, University of Calgary.

Shi, L. and Chen, G. (2008a). Case deletion diagnostics in multilevel models. *Journal of Multivariate Analysis*, **99**, 1860–1877.

Shi, L. and Chen, G. (2008b). Local influence in multilevel models. *Canadian Journal of Statistics*, **36**, 259–275.

Szatrowski, T. H. (1980). Necessary and sufficient conditions for explicit solutions in the multivariate normal estimation problem for patterned means and covariances. *The Annals of Statistics*, **8**, 802–810.

Tan, F. E. S., Ouwens, M. J. N. and Berger, M. P. F. (2001). Detection of influential observations in longitudinal mixed effects regression models. *Journal of the Royal Statistical Society: Series D*, **50**, 271–284.

Zhu, H., Ibrahim, J. G., Lee, S. and Zhang, H. (2007). Perturbation selection and influence measures in local influence analysis. *The Annals of Statistics*, **35**, 2565–2588.