

*schabenberger* examines the use and implementation of influence measures in LME models.

Influence is understood to be the ability of a single or multiple data points, through their presences or absence in the data, to alter important aspects of the analysis, yield qualitatively different inferences, or violate assumptions of the statistical model (*schabenberger*).

Outliers are the most noteworthy data points in an analysis, and an objective of influence analysis is how influential they are, and the manner in which they are influential.

*schabenberger* describes a simple procedure for quantifying influence. Firstly a model should be fitted to the data, and estimates of the parameters should be obtained. The second step is that either single or multiple data points, specifically outliers, should be omitted from the analysis, with the original parameter estimates being updated. This is known as ‘leave one out’ or ‘leave k out’ analysis. The final step of the procedure is comparing the sets of estimates computed from the entire and reduced data sets to determine whether the absence of observations changed the analysis.

A residual is the difference between an observed quantity and its estimated or predicted value. In LME models, there are two types of residuals, marginal residuals and conditional residuals. A marginal residual is the difference between the observed data and the estimated marginal mean. A conditional residual is the difference between the observed data and the predicted value of the observation. In a model without random effects, both sets of residuals coincide.

*schabenberger* notes that it is not always possible to derive influence statistics necessary for comparing full- and reduced-data parameter estimates.

## **Abstract**

This paper reviews the use of diagnostic measures for LME models in SAS. This text has been widely cited by texts that don't deal with SAS implementations.

## Schabenberger: Summary and Conclusions

- Standard residual and influence diagnostics for linear models can be extended to linear mixed models. The dependence of fixed-effects solutions on the covariance parameter estimates has important ramifications in perturbation analysis.
- To gauge the full impact of a set of observations on the analysis, covariance parameters need to be updated, which requires refitting of the model.
- The experimental INFLUENCE option of the MODEL statement in the MIXED procedure (SAS 9.1) enables you to perform iterative and noniterative influence analysis for individual observations and sets of observations.
- The conditional (subject-specific) and marginal (population-averaged) formulations in the linear mixed model enable you to consider conditional residuals that use the estimated BLUPs of the random effects, and marginal residuals which are deviations from the overall mean.
- Residuals using the BLUPs are useful to diagnose whether the random effects components in the model are specified correctly, marginal residuals are useful to diagnose the fixed-effects components.
- Both types of residuals are available in SAS 9.1 as an experimental option of the MODEL statement in the MIXED procedure.
- It is important to note that influence analyses are performed under the assumption that the chosen model is correct. Changing the model structure can alter the conclusions. Many other variance models have been fit to the data presented in the repeated measures example. You need to see the conclusions about which model component is affected in light of the model being fit.
- For example, modeling these data with a random intercept and random slope for each child or an unstructured covariance matrix will affect your conclusions about which children are influential on the analysis and how this influence manifests itself.

## 0.1 Iterative and non-iterative influence analysis

? highlights some of the issue regarding implementing mixed model diagnostics.

A measure of total influence requires updates of all model parameters.

however, this doesn't increase the procedures execution time by the same degree.

### 0.1.1 Iterative Influence Analysis

For linear models, the implementation of influence analysis is straightforward. However, for LME models, the process is more complex. Update formulas for the fixed effects are available only when the covariance parameters are assumed to be known. A measure of total influence requires updates of all model parameters. This can only be achieved in general is by omitting observations, then refitting the model.

? describes the choice between iterative influence analysis and non-iterative influence analysis.

### 0.1.2 Summary of Paper

Standard residual and influence diagnostics for linear models can be extended to LME models. The dependence of the fixed effects solutions on the covariance parameters has important ramifications on the perturbation analysis. Calculating the studentized residuals-And influence statistics whereas each software procedure can calculate both conditional and marginal raw residuals, only SAS Proc Mixed is currently the only program that provide studentized residuals Which are preferred for model diagnostics. The conditional Raw residuals are not well suited to detecting outliers as are the studentized conditional residuals. (schabenbege r)

LME are flexible tools for the analysis of clustered and repeated measurement data. LME extend the capabilities of standard linear models by allowing unbalanced and missing data, as long as the missing data are MAR. Structured covariance matrices for both the random effects  $G$  and the residuals  $R$ . missing at Random.

A conditional residual is the difference between the observed value and the predicted value of a dependent variable- Influence diagnostics are formal techniques that allow the identification of observations that heavily influence estimates of parameters. To alleviate the problems with the interpretation of conditional residuals that may have unequal variances, we consider scaling. Residuals obtained in this manner are called studentized residuals.

### 0.1.3 ITERATIVE VS. NONITERATIVE INFLUENCE ANALYSIS

While the basic idea of influence analysis is straightforward, the implementation in mixed models can be tricky. For example, update formulas for the fixed effects are available only when the covariance parameters are assumed to be known. At most the profiled residual variance can be updated without refitting the model. A measure of total influence requires updates of all model parameters, and the only way that this can be achieved in general is by removing the observations in question and refitting the model. Because this brute-force method involves iterative reestimation of the covariance parameters, it is termed iterative influence analysis. Reliance on closed-form update formulas for the fixed effects without updating the (un-profiled) covariance parameters is termed a noniterative influence analysis. An iterative analysis seems like a costly, computationally intensive enterprise. If you compute iterative influence diagnostics for all  $n$  observations, then a total of  $n + 1$  mixed models are fit iteratively. This does not imply, of course, that the procedure's execution time increases  $n$ -fold. Keep in mind that

- iterative reestimation always starts at the converged full-data estimates. If a data point is not influential, then its removal will have little effect on the objective function and parameter estimates. Within one or two iterations, the process should arrive at the reduced-data estimates.
- if complete reestimation does require many iterations, then this is important

information in itself. The likelihood surface has probably changed drastically, and the reduced-data estimates are moving away from the full-data estimates.

## SUMMARY AND CONCLUSIONS

Standard residual and influence diagnostics for linear models can be extended to linear mixed models. The dependence of fixed-effects solutions on the covariance parameter estimates has important ramifications in perturbation analysis. To gauge the full impact of a set of observations on the analysis, covariance parameters need to be updated, which requires refitting of the model.

The experimental INFLUENCE option of the MODEL statement in the MIXED procedure (SAS 9.1) enables you to perform iterative and noniterative influence analysis for individual observations and sets of observations. The conditional (subject-specific) and marginal (population-averaged) formulations in the linear mixed model enable you to consider conditional residuals that use the estimated BLUPs of the random effects, and marginal residuals which are deviations from the overall mean. Residuals using the BLUPs are useful to diagnose whether the random effects components in the model are specified correctly, marginal residuals are useful to diagnose the fixed-effects components. Both types of residuals are available in SAS 9.1 as an experimental option of the MODEL statement in the MIXED procedure. It is important to note that influence analyses are performed under the assumption that the chosen model is correct. Changing the model structure can alter the conclusions. Many other variance models have been fit to the data presented in the repeated measures example. You need to see the conclusions about which model component is affected in light of the model being fit. For example, modeling these data with a random intercept and random slope for each child or an unstructured covariance matrix will affect your conclusions about which children are influential on the analysis and how this influence manifests itself.