Chapter 1

Residual Analysis and Influence Diagnostics for Method Comparison

Model validation and model appraisal are vital parts of the modelling process, yet are too often overlooked. Using a small set of simple measures and methods, such as the AIC and R^2 measures, is insufficient to properly assess the usefulness of a fitted model. In classical linear models model diagnostics are now considered a required part of any statistical analysis, and the methods are commonly available in statistical packages and standard textbooks on applied regression. A full and comprehensive analysis that comprises residual analysis and influence analysis for testing model assumptions, should be carried out. However it has been noted by several papers (**??**) that model diagnostics do not often accompany LME model analyses. Furthermore, a suite of dagnostic procedures designed for method comparison should be adopted.

1.1 Residual Analysis

Model diagnostics techniques determine whether or not the distributional assumptions are satisfied, and to assess the influence of unusual observations, and have been become a required part of any statistical analysis. Well established methods are commonly available in statistical packages and standard textbooks on applied regression. However it has been noted by several papers that model diagnostics do not often accompany LME model analyses.

A residual is simply the difference between an observed value and the corresponding fitted value, as predicted by the model. As with classical models, there are two key techniques for LME models: a residual plot and the normal probability plot. The rationale is that, if the model is properly fitted to the model, then the residuals would approximate the random errors that one should expect. If the residuals behave randomly, with no discernible trend, the model has fitted the data well. Conversely, if some sort of non-random trend is evident in the model, then the model can be considered to be poorly fitted.

The underlying assumptions for LME models are similar to those of classical linear models. However, for LME models the matter of residuals are more complex, both from a theoretical point of view and from the practicalities of implementing a comprehensive analysis using statistical software. **?** discusses residuals for LME model, providing a useful summary of various techniques. Prominent in literature is the taxonomy of residuals for LME Models, distinguishing between condition residuals, marginal residuals and EBLUPS, including **????**.

Statistical software environments, such as the `R` programming language, provides a suite of tests and graphical procedures for appraising a fitted LME model, with several of these procedures analysing the model residuals. Texts such as **???** describe what can be implemented for LME residual analyses with statistical software, such as `R` and `SAS`.

In the context of method comparison, a residual analysis would be carried out just as any other LME model would, testing normality. There is little scope for adding additional insights, other than to say that it is possible to create plots specific to each method. The figures on the next page depict the residual analysis for the *Blood* data, which can be used to indicate which methods disagree with the rest, but these would be a confirmation of something detected previously.

Analysis of the residuals could determine if the methods of measurement disagree systematically, or whether or not erroneous measurements associated with a subset of the cases are the cause of disagreement. The figure depicts residual plot for the Systolic Blood Pressure example used in **?**. Points are labelled by subjects, with cases 67, 68 and 71 being among the prominent cases. Prominent cases warrant further investigation, but an analyst should procede to influence diagnostics beforehand.

The next figure depicts residual plot for the Systolic Blood Pressure example, panelled by the various measurement methods. It serves to confirm agreement between methods J and R, with lack of agreement between those two methods and method S. However, little insight can be gained as to what actually causes lack of agreement here.

1.2 Influence Diagnostics

Model diagnostic techniques can determine whether or not the distributional assumptions are satisfied, but also to assess the influence of unusual observations. Following model specification and estimation, it is of interest to explore the model-data agreement by raising pertinent questions. **?** provide some insight into how to compute and interpret model diagnostic plots for LME models. Unfortunately this aspect of LME theory is not as expansive as the corresponding body of work for Linear Models. Their particular observations will be reverted to shortly. Further to the analysis of residuals, **?** recommends the examination of the following questions:

- Does the model-data agreement support the model assumptions?
- Should model components be refined, and if so, which components? For example, should certain explanatory variables be added or removed, and is the covariance of the observations properly specified?
- Are the results sensitive to model and/or data? Are individual data points or groups of cases particularly influential on the analysis?

The last of these three questions, regarding influential points, is of particular interest in the context of Method Comparison. After fitting an LME model, it is important

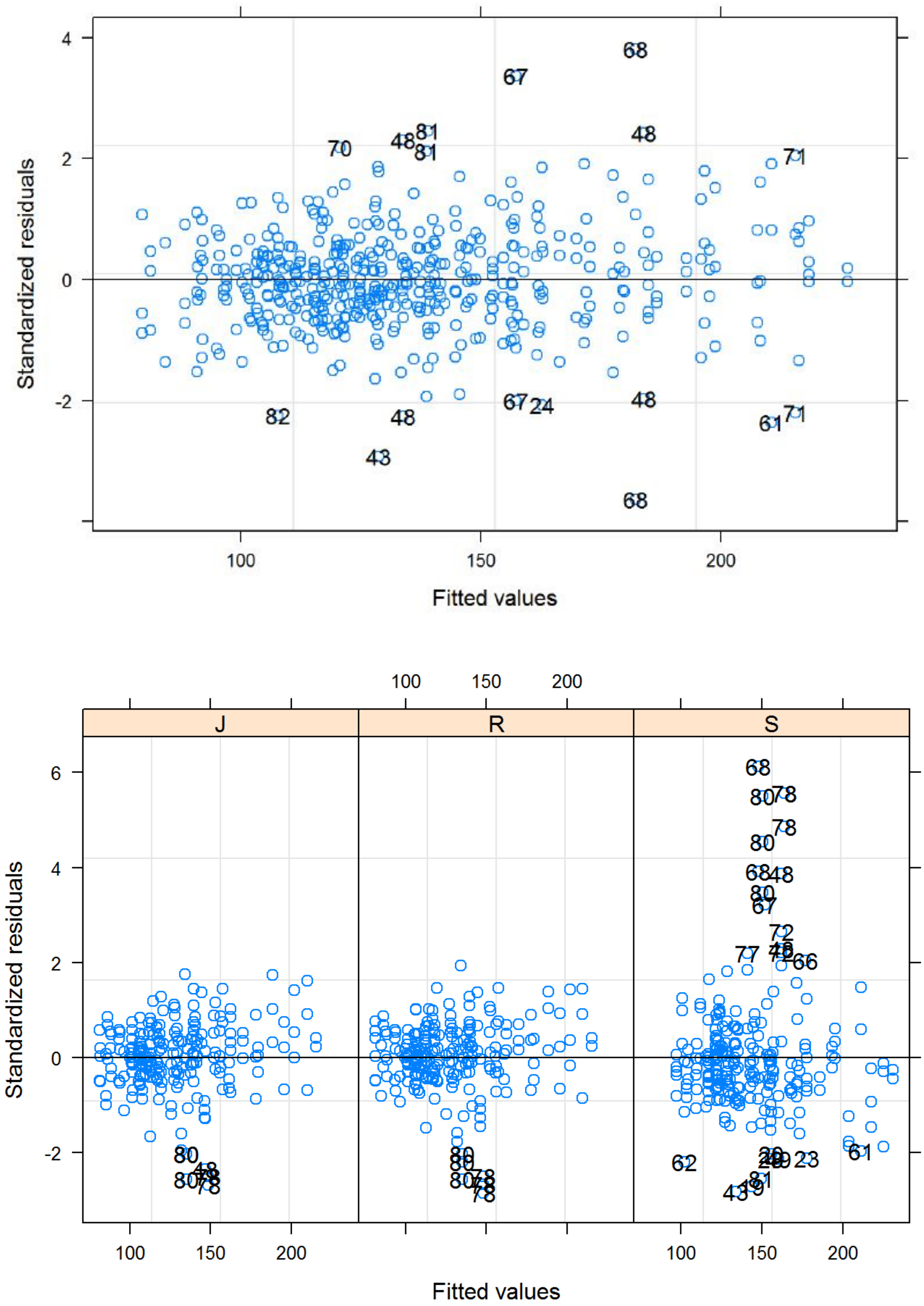

Figure 1.1.1: LME Residuals by Method (Blood Pressure Data)

to carry put model diagnostics to check whether distributional assumptions for the residuals as satisfied and whether the fit the model is sensitive to unusual assumptions. The process of carrying out model diagnostic involves several informal and formal techniques, which will mentioned throughout the chapter.

Influential points have a large influence on the fit of the model. Influential points are a set of one or more observations whose removal would cause a different conclusion in the analysis, e.g. substantially changes the estimate of the regression coefficients. **?** remarks that influence diagnostics play an important role in the interpretation of results, because influential data can negatively influence the statistical model and generalizability of the model. **?** remarks that the concept of critiquing the model-data agreement applies in mixed models in the same way as in linear fixed-effects models. In fact, because of the more complex model structure, you can argue that model and data diagnostics are even more important (**?**).

1.2.1 Analyzing Influence in LME Models

Model diagnostic techniques, well established for classical models, have since been adapted for use with linear mixed effects models. Diagnostic techniques for LME models are inevitably more difficult to implement, due to the increased complexity.

Influence diagnostics are formal techniques allowing for the identification of observations that exert substantial influence on the estimates of fixed effects and variance covariance parameters. While linear models and GLMS can be studied with a wide range of well-established diagnostic technqiues, the choice of methodology is much more restricted for the case of LMEs. However influence diagnostics for LME Models is an area of active research. Research on diagnostic analyses for LME models are presented in **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?** and **??**.

? states that goal of influence analysis is not primarily to mark data points for deletion so that a better model fit can be achieved for the reduced data, although this might be a result of influence analysis. The goal is rather to determine which cases are influential and the manner in which they are important to the analysis.

1.2.2 Measuring of Influence for LME Models

Influence analysis methodologies have been used extensively in classical linear models, and provided the basis for methodologies for use with LME models. Computationally inexpensive diagnostics tools have been developed to examine the issue of influence (?).

? remarks the development of efficient computational formulas is crucial making deletion diagnostics useable, allowing one to obtain the case deletion diagnostics by making use of basic building blocks, computed only once for the full model. A number of approaches to model diagnostics are described, including variance components, dixed effects parameters, prediction of the response variable and of random effects, and the likelihood function. Influence statistics can be grouped by the aspect of estimation that is their primary target:

- **overall measures compare changes in objective functions**: (restricted) likelihood distance (Cook and Weisberg 1982, Ch. 5.2)
- **influence on parameter estimates**: Cook's (Cook 1977, 1979), MDFFITS (Belsley, Kuh, and Welsch 1980, p. 32)
- **influence on precision of estimates**: CovRatio and CovTrace
- **influence on fitted and predicted values**: PRESS residual, PRESS statistic (Allen 1974), DFFITS (Belsley, Kuh, and Welsch 1980, p. 15)
- **outlier properties**: internally and externally studentized residuals, leverage

? lists several established methods of analyzing influence in LME models. These methods include Cook's distance for LME models, likelihood distance, the variance (information) ration, the Cook-Weisberg statistic, and the Andrews-Prebigon statistic.

The subscript (U) is used to denote quantities computed from data with subset of cases U omitted. If the global measure suggests that the points in U are influential, you should next determine the nature of that influence. In particular, the points can affect

- the estimates of fixed effects
- the estimates of the precision of the fixed effects
- the estimates of the covariance parameters
- the estimates of the precision of the covariance parameters
- fitted and predicted values

For example, if observations primarily affect the precision of the covariance parameters without exerting much influence on the fixed effects, then their presence in the data may not distort hypothesis tests or confidence intervals about β. **?** notes that removing observations or sets of observations affects fixed effects and covariance parameter estimates.

1.2.3 Cook's Distance

As previously described, Cooks Distance (D_i) is a diagnostic technique used in classical linear models, that functions as an overall measure of the influence of an observation that is a measure of aggregate impact of each observation on the group of regression coefficients, as well as the group of fitted values. Cook's Distance as a measure of the influence of observations in subset U on a vector of parameter estimates is given below (**?**)

$$\delta_{(U)} = \hat{\beta} - \hat{\beta}_{(U)}.$$

Observations, or sets of observations, that have high Cook's distance usually have high residuals, although this is not necessarily the case. If the predictions are the same with or without the observation in question, then the observation has no influence on the regression model. If the predictions differ greatly when the observation is not included in the analysis, then the observation is influential.

Large values for Cook's distance indicate observations for special attention. Cook's distance can be used in several ways: to indicate data points that are particularly worth checking for validity; to indicate regions of the design space where it would be good to be able to obtain more data points.

Use of threshold values for Cook's Distance is discouraged (**?**). However, informal heuristics do exist for OLS models; Obervations for which Cook's distance is higher than 1 are usually considered as influential. Another informal threshold of $4/n$ or $4/(n-k-1)$, where n is the number of observations and k the number of explanatory variables.

? advises the use of diagnostic plotting and to examine in closer details the points with *"values of D that are substantially larger than the rest"*, and that thresholds should feature only to enhance graphical displays.

The effect on the precision of estimates is separate from the effect on the point estimates. Data points that have a small Cook's distance, for example, can still greatly affect hypothesis tests and confidence intervals, if their influence on the precision of the estimates is large.

? develops case deletion diagnostics, in particular the equivalent of Cook's distance for diagnosing influential observations when estimating the fixed effect parameters and variance components, adapting the Cook's Distance measure for the analysis of LME models. For LME models, two formulations exist; a Cook's distance that examines the change in fixed fixed parameter estimates, and another that examines the change in random effects parameter estimates. The outcome of either Cook's distance is a scaled change in either β or θ. **?** gives a detailed discussion of the various formulation for Cook's distances for LME Models.

Consideration of how leave-U-out diagnostics would work in the context of Method Comparison problems is required. There are several scenarios. **?** describes two type of diagnostics. When the set consists of only one observation, the type is called '*observation-diagnostics*'. For multiple observations, Preisser describes the diagnostics as '*cluster-deletion*' diagnostics. Suppose we have two methods of measurement X

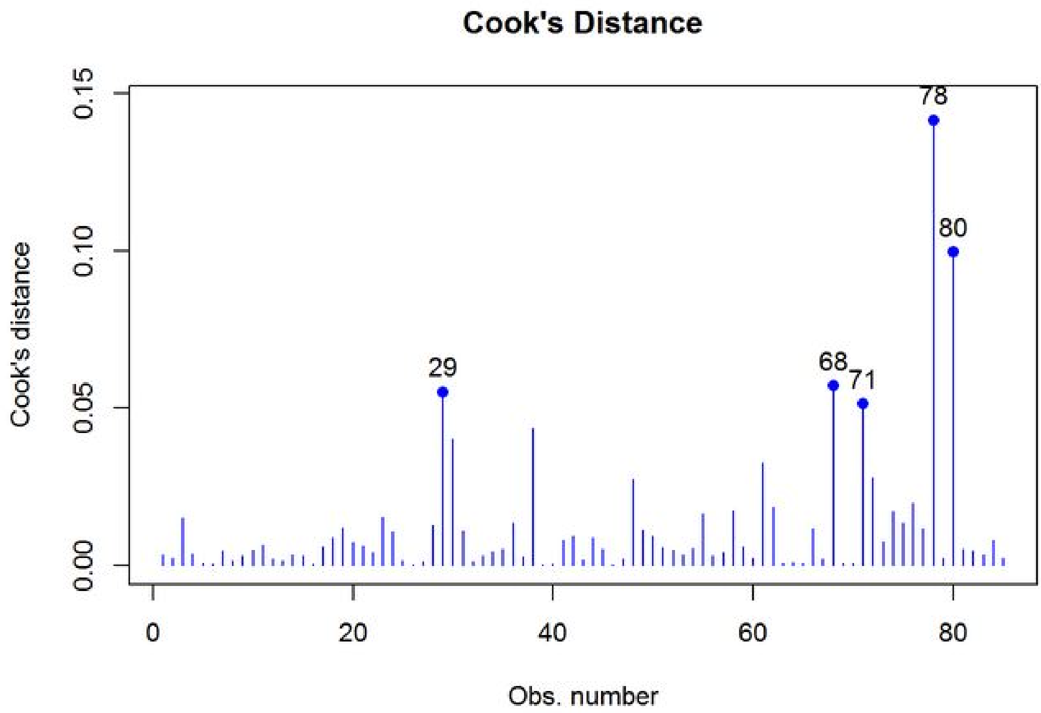

Figure 1.2.2: Cook's Distance Plot for the JS Comparison

and Y, each with three measurements for a specific case: $(x_1, x_2, x_3, y_1, y_2, y_3)$

- Leave One Out - one observation is omitted (e.g. x_1)
- Leave Pair Out - one pair of observation is omitted (e.g. x_1 and y_1)
- Leave Case (or Item or Subject) Out - All observations associated with a particular case or subject are omitted. (e.g. $\{x_1, x_2, x_3, y_1, y_2, y_3\}$)

The natural sampling unit is the item or subject, similar to the example provided by **?**. Hence, the third option, henceforth, referred to as "Leave subject Out" will be the option used.

1.3 Model Diagnostics for Roy's Models

Further to previous work, this section revisits case-deletion and residual diagnostics, and explores how approaches devised by **?** can be used to appraise Roy's model. These authors specifically look at Cook's Distances and Likelihood Distances.

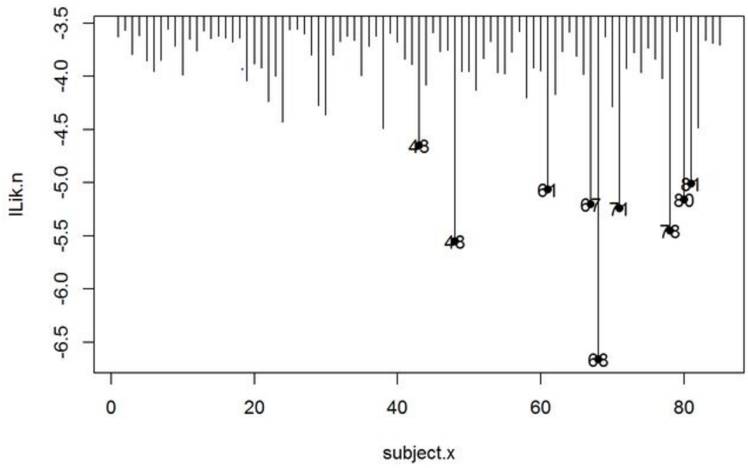

Figure 1.3.3: