



Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

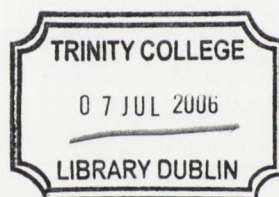
Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.



THOSIS
7942.

Deletion Diagnostics for the Linear Mixed Model

Dominic Mark Dillane

Thesis submitted for the degree of Doctor of Philosophy

Trinity College Dublin

Department of Statistics

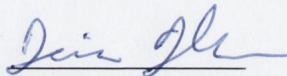
September 2005

Declaration

This thesis has not been submitted as an exercise for a degree at any other university. It is entirely my own work

The Library of Trinity College Dublin may lend or copy this thesis.

Signed


Dominic Dillane

Summary

Modeling data is an integral element of modern statistical analysis. Methodological developments combined with the explosion in computing power over the past ten to fifteen years have greatly enhanced statisticians' ability to model situations and phenomena. The need to assess a model's validity and suitability is an integral element of the model building process. Model criticism is central to this thesis and specifically model criticism for one of the most frequently utilised models in statistical analyses, the Linear Mixed Model with normally distributed errors. Particular emphasis is given to deletion diagnostics for both the fixed and for the covariance structure parameters.

Model estimation techniques are sensitive to unusual observations. The data analyst must validate as carefully as possible the assumptions underlying the application of such models and also identify observations influential on the results of the analysis. Such data may be outlying and removed from the analysis, may be entirely appropriate and retained in the analysis, or may suggest that the model is inadequate. Whatever the ultimate nature of such cases it is imperative that they be identified to ensure that intelligent subject-matter-based inferences be drawn from the data analysis.

Deletion diagnostics for the LMM are the focus of this thesis. Such measures have always been available to anyone who was willing to invest the time and money to refit the model with each observation or arbitrary subset deleted. The key to making deletion diagnostics useful in practice is the development of computationally efficient approaches. New, fast and approximate computational procedures, arising as a by-product of the fitting process are developed in this thesis. These new approaches enable influence analysis to be conducted at both stages of the fitting process for the LMM. These are at

the estimation of the covariance parameters and at the subsequent estimation of the fixed effects. The accuracy of these approaches are assessed using a number of datasets.

Acknowledgements

I would like to express my sincere gratitude and thanks to my supervisor Professor John Haslett who has always been most generous with his time, direction and advice. His constant encouragement, support, expertise and insight were invaluable in securing the completion of this thesis.

A special thanks to the staff in the Department of Statistics. In particular, I would like to thank Kris, Myra, Eamonn, Michael, Simon, Aideen, Olivia and Lydia who were always very helpful and supportive.

I wish to thank the Dublin Institute of Technology for enabling me to undertake the research. In particular, I wish to thank the Head of School of Hospitality Management and Tourism at the DIT, Dr. Noel O'Connor for his encouragement and for facilitating me in every way possible throughout my research.

The interest, unfailing support, advice and encouragement of my parents, brothers and close friends, was a constant source of inspiration and is greatly appreciated.

Glossary and Notation

ANOVA	Analysis Of Variance
AO	Additive Outlier
BLUE	Best Linear Unbiased Estimator
BLUP	Best Linear Unbiased Predictor
CVR	Covariance Ratio
GLM	General Linear Model
IO	Innovative Outlier
LM	Linear Model
LMM	Linear Mixed Model
MINQUE	Minimum Norm Quadratic Unbiased Estimation
ML	Maximum Likelihood
MM	Method of Moments Estimation
OLS	Ordinary Least Squares
REML	Restricted Maximum Likelihood
RSS	Residual Sum of Squares
RVC	Relative Variance Change
Y	Stacked vector of observed data
X	Design matrix for fixed effects
Z	Design matrix for random effects.
V	$COV(Y)$
$Var(A)$	Variance of A
$\hat{e}$	Stacked vector of marginal residuals
$\tilde{e}_{(a)}$	The leave- a -out conditional residual.
$\tilde{e}_{(\bullet)}$	Stacked vector of leave-1-out conditional residuals
Q	$V^{-1} - H$
H	$V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}$

Q_{aa}	Block of Q corresponding to subset a
V^{aa}	Block of V^{-1} corresponding to subset a
$\hat{V}$	Estimate of V based on all the data Y
$\hat{V}_{(a)}$	Estimate of V based on the data excluding subset a
$ A $	The determinant of the matrix A
$F_{(a)}$	Any deletion diagnostic computed based on $\hat{V}$ and on the exclusion of subset a
$F_{((a))}$	Any deletion diagnostic computed based on $\hat{V}_{(a)}$ and on the exclusion of subset a
$F_{(a)}^{Tayl}$	Deletion diagnostic computed using Taylor series approximation

TABLE OF CONTENTS

	Page number
TITLE PAGE	1
DECLARATION	2
SUMMARY	3
ACKNOWLEDGEMENTS	5
GLOSSARY AND NOTATION	6
CONTENTS	8
LIST OF FIGURES	11
LIST OF TABLES	13
CHAPTER 1	INTRODUCTION
	14
1.0	Introduction
	14
1.1	Diagnostic analyses and the statistical modeling process
	16
1.2	The Linear Model
	12
1.3	Model Criticism for the LM
	19
1.4	Conclusions
	33
CHAPTER 2	DIAGNOSTICS FOR THE LMM
	35
2.0	Introduction
	35
2.1	The Linear Mixed Model
	36
2.2	Methods of estimation
	34
2.3	Diagnostics for the LMM
	42
2.4	Influence diagnostics for the covariance structure
	48
2.5	Conclusions
	50

CHAPTER 3	VARIANCE COMPONENT DIAGNOSTICS	53
3.0	Introduction	53
3.1	Subset Deletion for components of variance in the LMM	54
3.2	Evaluation of proposals	58
3.3	Comparison of $\hat{\sigma}_{(a)}$ with $\tilde{\sigma}_{(a)}$	78
3.4	Summary and Discussion	85
CHAPTER 4	AUTOCORRELATION PARAMETER DIAGNOSTICS	88
4.0	Introduction	88
4.1	LMMs for autocorrelated data	89
4.2	Evaluation of Proposals	93
4.3	A further application of these proposals	100
4.4	Discussion	105
CHAPTER 5	MEAN STRUCTURE DIAGNOSTICS	107
5.0	Introduction	107
5.1	Comparison of diagnostics based on $\hat{V}$ and $\hat{V}_{(a)}$	108
5.2	Approximations based on Taylor Series	118
5.3	Evaluation of proposals	120
5.4	Discussion	126
CHAPTER 6	CONCLUSIONS AND DISCUSSION	127
6.0	Introduction	127
6.1	Thesis Summary and Achievements	128
6.2	Deficiencies of thesis and further research opportunities	130

References 132

APPENDICES

Appendix 1 Derivation of results used to compute proposed diagnostics of chapters three and four. 139

Appendix 2 Derivation of results used in Taylor expansion approach outlined in Chapter 5 141

List of Figures

Chapter 3	Page Number
3.1	Deletion plots for nicotine study 60
3.2a	Deletion plots for corn data 67
3.2b	Deletion plots for soybeans data 69
3.3	Plot of rats' weight over time 70
3.4	Deletion plots for rats data 68
3.5	Deletion plots for the AMC data 75
3.6(a)-(b)	$\text{Log}(Cov(\tilde{e}_{(a)}, \tilde{e}_{(a)}))$ and $\text{Log}(Cov(\hat{e}_a, \tilde{e}_{(a)}))$ for each observation 79
3.6(c)-(d)	Plots of $\frac{ X_{(a)}^T V_{bb}^{-1} X_{(a)} }{ X^T V^{-1} X }$ for each subset (rat) 80
3.7	Plot of modified rats dataset 82
3.8(a)-(b)	$\text{Log}(Cov(\tilde{e}_{(a)}, \tilde{e}_{(a)}))$ and $\text{Log}(Cov(\hat{e}_a, \tilde{e}_{(a)}))$ for each observation 82
3.8(c)-(d)	Plots of the differences between $Cov(\tilde{e}_{(a)}, \tilde{e}_{(a)})$ and $Cov(\hat{e}_a, \tilde{e}_{(a)})$ 83
3.8(e)-(f)	Plots of $\frac{ X_{(a)}^T V_{bb}^{-1} X_{(a)} }{ X^T V^{-1} X }$ for each rat. 83
3.8(g)-(h)	Plot of the variance of the error against $\frac{ X_{(a)}^T V_{bb}^{-1} X_{(a)} }{ X^T V^{-1} X }$ 84
Chapter 4	
4.1	Longitudinal plot of simulated data 93
4.2	Plot of $\frac{\hat{\rho}_{(a)}}{\hat{\rho}} - 1$ using REML and proposed method 94
4.3	Plot of ice cream consumption against temperature 96
4.3.1	Deletion plots for ice cream data 97
4.4	Deletion plots for autocorrelation parameter for rats data 98
4.5	Deletion plots for modified rats dataset 99
4.6	Example 1- Analysis of simulated data 101
4.7	Example 2- Analysis of simulated data 103

Chapter 5

5.1(a)-(h)	Singleton deletion scatterplots of $F_{(a)}$ versus $F_{((a))}$ for rats data	108
5.1(i)-(p)	Subset (rat) deletion scatterplots of $F_{(a)}$ versus $F_{((a))}$ for rats data	110
5.2	Plots of Cook's Distance for each rat	111
5.3a(a)-(f)	Singleton deletion plots for corn data	112
5.3a(g)-(l)	Singleton deletion plots for soybeans data	113
5.3b(a)-(f)	Subset (counties) deletion plots for corn data	114
5.3(g)-(l)	Subset (counties) deletion plots for soybeans data	115
5.4a	Plot of Cook's Distance for each observation for the corn data	116
5.4b	Plot of Cook's Distance for each county for the corn data	116
5.5	Scatterplots of $F_{(a)}^{Tayl}$ versus $F_{((a))}$ for rats data	120
5.6	Scatterplots of $F_{(a)}^{Tayl}$ versus $F_{((a))}$ for corn and soybeans data	121
5.7	Plots of leverage versus $\frac{F_{(a)}^{Tayl}}{F_{((a))}} - 1$ for the rats data	123
5.8	Plots of $Var(F_{((a))} - F_{(a)}^{Tayl})$ versus leverage	124

List of Tables

Chapter 3	Page Number
3.1 RVC using the full REML refit and the three approaches	61
3.2 Survey and Satellite Data for Corn and Soybeans in 12 Iowa Counties	63
3.3 Analytical Chemistry Data	73
3.4 RVC using the three approaches for increasingly large deletions	81

Chapter 1

Introduction

1.0 Introduction

Model criticism for the linear model (LM) is the concern of this chapter. The focus of the thesis is on diagnostics for the linear mixed model (LMM), which is an important subset of the LM family of models. The main diagnostic approaches in the literature stem from consideration of the LM. The LM is introduced in this chapter and its main variants considered.

Constructing appropriate models for various types of data is an integral element of modern statistical analysis. The statistical model building process is well documented in the literature. The need to assess a model's validity and suitability is an integral element of this process. There is a substantial literature on model criticism and diagnostic methodologies for statistical models. The principal diagnostic approaches are introduced and considered in this chapter.

In statistics, the most widely used models revolve around linear structures. Diagnostics are best developed for the LM with constant variance and independent errors. The principal diagnostic measures have been generalised for analysis of data described using a more general form of the LM and are introduced in this chapter. An introduction to the main concepts and issues considered throughout the thesis is provided within the chapter.

1.1 Diagnostic analyses and the statistical modeling process

A diagnostic analysis is a key element of statistical modeling. The structured statistical model-building strategy espoused so well by Box and Jenkins (1976) consists of three main steps, model identification, model estimation and model validation.

At the model identification stage, models are selected that may be appropriate for the dataset of interest. The model will inevitably involve one or more parameters whose values must be estimated from the data. Model estimation consists of finding the best possible estimates of those unknown parameters within a given model. The final step in the model building process is concerned with analysing the quality of the model that we have specified and estimated. This involves considering how well the model fits the data and how the data influences the model identification and estimation processes. This step is critical to the success of the modeling process. If no inadequacies are found, the modeling may be assumed to be complete. Otherwise another model is chosen in light of the inadequacies and the process is repeated until an acceptable model is found.

Model diagnostics are primarily concerned with two interrelated questions. Firstly, diagnostics are concerned with checking how well the fitted model resembles the observed data. Model validation approaches focus on residual analysis (Weisberg, 1985). The second element of model criticism is influence analysis. Influence diagnostics are used to analyse the effect of perturbing the model formulation or data. It is essential that the data analyst identify influential observations that affect the major statistical inference of the study.

Belsley, Kuh and Welsch (1980, pg 15) define an influential observation as “one which either individually or together with several other observations has a demonstrably larger

impact on the calculated values of various estimates than is the case for most of the other observations". Influential data can arise due to a mis-specified model or due to anomalous data. However, influential data may arise within a correctly specified model with satisfactory data. Their presence simply indicates the need for further analysis.

Numerous books and articles consider residual and influence analysis. These include the seminal papers of Cook (1977,1986), Pregibon (1981) and Chatterjee and Hadi (1986) and the books by Belsley *et al*, (1980), Chatterjee and Hadi (1988) and Cook and Weisberg (1982). These focus on diagnostic methodologies in the context of the LM with constant variance and independent errors. They have been developed further for application to more general forms of the LM. In section 1.3 the principal diagnostic approaches for the LM from the literature are considered. Firstly, the LM and its main variants are introduced in section 1.2 below.

1.2 The Linear Model

The defining characteristic of the LM is that the model is linear in the parameters. McCulloch and Searle (2001) outline a general form of the linear model that may be adapted to describe its main variants. Let Y be an $N \times 1$ vector with N data values and mean μ . V is defined as the variance covariance matrix of Y so

$$Y \sim (\mu, V) \quad (1.1)$$

The variant of the linear model can be defined by specifying μ and V appropriate to the nature of the data being studied. For example, the forms of μ and V for the LM with constant variance and independent errors are

$$\mu = X\beta \quad (1.2)$$

where β is a $p \times 1$ vector of unknown fixed parameters and X is a known $N \times p$ matrix and

$$V = \sigma^2 I \quad (1.3)$$

with I being the identity matrix of size $N \times N$. The principal diagnostic approaches were developed in the context of this model with the data deemed to have come from a normal distribution. The General Linear Model (GLM) with normal errors is defined as

$$Y \sim N(X\beta, V) \quad (1.4)$$

This incorporates many variants of the LM principal among these being the LMM. The LMM is an important member of the LM family, which while assuming normally distributed errors permits heterogeneity of variance. The LMM is a LM where some of the parameters are treated not as constants but as realisations of random variables. The LMM is defined as

$$Y = X\beta + Z\gamma + \varepsilon \quad (1.5)$$

where X and $Z \equiv (Z_1, Z_2, \dots, Z_r)$ are known matrices, β is a vector of fixed effects and γ is a vector of random effects with $E(\gamma) = 0$, $Cov(\gamma) = D$ and $(Cov(\gamma, \varepsilon)) = 0$.

V will then be of the form $ZDZ^T + A$ where Z^T denotes the transpose of Z and $A = Var(\varepsilon)$. The LMM is considered in more detail in chapter 2.

A more general form of the LM has been developed to accommodate non-normal data. Nelder and Wedderburn (1972) recognised that non-normal models such as log-linear

models, probit and logit models could be generalised to unify an entire collection of models. The generalised linear model has been extended to incorporate generalised linear mixed models. Generalised linear models are of the form

$$E(Y) = \mu \quad (1.6)$$

and

$$g(\mu) = X\beta \quad (1.7)$$

where $g(\cdot)$ is a known function called the link function. For generalised linear models Y is assumed to follow a distribution from the exponential family of distributions. For Ordinary Least Squares (OLS) linear regression or for the GLM the link function is the identity. There is an extensive literature on generalised linear models with McCullagh and Nelder (1989) and McCulloch and Searle (2001) being the principal texts in the area. In this thesis the focus is on diagnostics for the LMM with normal errors and so generalised linear models are not considered in detail.

The development of diagnostic methodologies has largely stemmed from the LM with independent errors and constant variance. There are a multitude of measures proposed for this situation. Many of these generalize to other members of the LM family. The principal diagnostic methodologies and measures are outlined in section 1.3 below in the context of the GLM with a single component of variance. They are reviewed in the context of the LMM in chapter 2.

1.3 Model Criticism for the LM.

Cook's seminal work (Cook, 1977) in the context of the linear model with constant variance and uncorrelated errors has spawned an entire literature on residuals and influence. The key to this development was Cook's demonstration that the effects on estimates $\hat{\beta}$ of β of deleting each observation in turn could be cheaply computed as a by-product of the fitting procedure. The impact of the deletion of each observation can be computed *en bloc*. This ease of computation has led to these deletion diagnostics becoming a feature of statistical software for this OLS situation.

Various papers have been written generalising these diagnostic measures to the GLM with the variance covariance matrix defined as $\sigma^2 V$. These papers include Martin (1992), Haslett and Hayes (1998), Haslett (1999) and Baade and Pettitt (2000). The formulae derived to compute the diagnostics are based on the key assumption that $\hat{V}$ is unchanged when observations are deleted. The effect of this assumption on the accuracy of the various diagnostic measures has not been explored in detail in the literature. This issue is considered in chapter 5. The various diagnostic approaches from the literature for the GLM with a single component of variance are outlined below.

1.3.1 Residual Analysis

The computation and analysis of residuals are key elements of any diagnostic analysis. The classic or marginal residuals are calculated by subtracting the estimate of the expected value of Y , $E(Y) = X\hat{\beta}$ from the observed Y with $\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$. The marginal residuals are thus defined as

$$\hat{e} = Y - X\hat{\beta} \tag{1.8}$$

If we define $Q = V^{-1} - H$ where $H = V^{-1}XB$ and $B = (X^T V^{-1} X)^{-1} X^T V^{-1}$ then it can be easily shown that $\hat{e} = VQY$ and thus $\text{Var}(\hat{e}) = VQV$. The marginal residuals are by-products of the fitting process and are easily calculated. For the case of OLS linear regression, graphical plots of the residuals are used to validate model assumptions. However, for the GLM the marginal residuals are not independent even asymptotically (unless V is diagonal). One general approach has been to rotate the residuals by $(\text{var}(\hat{e}))^{-1/2}$ when ordinary least squares applies. Theil (1965) provides an example of this approach. The principal disadvantage of diagnostics based on this approach is the difficulty in interpreting them.

While the marginal residuals based on the best linear unbiased estimator (BLUE) of Y for the GLM are well defined, residuals based on the best linear unbiased predictor (BLUP) of Y are degenerate. This is the case because the BLUP of an observed datum y_i given the observed data Y is y_i . This has led to the study of conditional residuals referred to by some authors as deletion prediction residuals. Let Y^T be partitioned as $\{Y_a^T, Y_b^T\}$ so that Y is a stacked vector. The conditional residual or deletion prediction residual has been defined (Martin, 1992) as the difference between Y_a and $\tilde{Y}_a(Y_b)$ the BLUP of Y_a given Y_b and given an estimate of $(\hat{V})$ the matrix V . Therefore $\tilde{Y}_a(Y_b)$ is an approximation as we still use $\hat{V}$ rather than $\hat{V}_{(a)}$. For notational convenience $\hat{V}$ is referred to as V . Thus the conditional or deletion prediction residual can be defined as

$$\tilde{e}_{(a)} = Y_a - (X_a \hat{\beta}(Y_b) + V_{ab} V_{bb}^{-1} (Y_b - X_b \hat{\beta}(Y_b))) \quad (1.9)$$

The subscript $_{(a)}$ denotes "computed without the use of a subset Y_a of the data". A very useful expression for deriving the conditional residuals from the marginal residuals is given in Haslett and Hayes (1998) with

$$\tilde{e} = \Delta V^{-1} \hat{e}_{(\bullet)} \quad (1.10)$$

where $\tilde{e}_{(\bullet)}$ is a vector containing the leave-1-out conditional residuals and $\Delta = 1/\text{diag}(Q)$.

This result is extended to subset deletion for arbitrary subsets in Haslett and Hayes (1998) and Martin (1992) with $\tilde{e}_{(a)} = Q_{aa}^{-1} (V^{-1} \hat{e})_a$. Haslett (1999) and Haslett and Dillane (2004) highlight the importance of the conditional residual $\tilde{e}_{(a)}$ in the computation of the various deletion diagnostics. They show that many of these statistics are simple functions of $\tilde{e}_{(a)}$.

Martin (1992) proposes alternative conditional residuals called “prediction residuals”.

These are defined as

$$\bar{s}_{(a)} = Y_a - (X_a \hat{\beta}(Y) + V_{ab} V_{bb}^{-1} (Y_b - X_b \hat{\beta}(Y))) \quad (1.11)$$

These two types of conditional residual depend on which estimate of β is used in their calculation. An estimate of β based on all the data produces what Martin (1992) terms a prediction residual. The deletion prediction residuals are based on β estimated from Y_b . Both measures are computed using $\hat{V}$ and are thus approximations. Martin’s ‘prediction’ residuals are related to the marginal residuals by

$$\bar{s}_{(a)} = -V_{ab} V_{bb}^{-1} \hat{e} \quad (1.12)$$

However, they are somewhat counter intuitive as the estimate of β used in their computation is based on all the data.

The third category of residuals considered is recursive residuals. Recursive residuals which arise principally in the time series literature given the time ordered nature of such

data are applicable to some LMM applications. They are directly connected to the conditional residuals. The recursive residuals $\hat{r}_t$ are a contrast between Y_t and an estimate of $E[Y_t|Y_s; s < t]$ in which the unknown parameter β is estimated from $Y_{<t} \equiv (Y_s; s < t)$. Haslett and Haslett (2005) show that $\hat{r}_t = V^{-1/2} X(\hat{\beta}(Y) - \hat{\beta}(Y_{<t})) = B \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix}$ with $\tilde{e}_{(a)} = \tilde{e}_{(\geq t)}$ the conditional residual for the subset $(s = t, \dots, N)$. The past values on which the prediction of Y_t is based are clearly defined for time series data. However, the use of recursive residuals in other areas has produced quite imaginative definition of the "past". Pickford and Haslett (1999) order data by measure of size and consider both a "backward" and "forward" interpretation in computing the recursive residuals. A review by Kianaford and Swallow (1996) point out that the recursive residuals are not unique as the data can be ordered in any number of ways and are thus problematic.

1.3.2 Contributions to the lack of fit statistic

The lack-of-fit statistic is defined as $S(Y) = \hat{e}^T V^{-1} \hat{e} = Y^T Q Y$. It is central to the analysis of the linear model reflecting the distance of the data from the best fitting linear model. By using appropriate decompositions of $S(Y)$ or for simplicity S the impact of individual observations or groups of observations may be assessed. Using the earlier result $\tilde{e} = \Delta V^{-1} \hat{e}$ S may be rewritten as

$$S = \hat{e}^T \Delta^{-1} \tilde{e} = \sum_{i=1}^n q_{ii} \hat{e}_i \tilde{e}_{(i)} = \sum_{i=1}^n C_i \quad (1.13)$$

Thus C_i , the contribution to the lack-of-fit statistic can be written as a weighted sum of the marginal residuals and the leave-1-out prediction residuals for each case. For an observation to have a large value for C_i each of the residuals $\hat{e}_i$ and $\tilde{e}_{(i)}$ must be large and simultaneously have a large q_{ii} . In interpreting this statistic it is worth considering

what these two types of residuals are measuring. A marginal residual measures the difference between the data and the underlying trend of the model, a global feature that has a certain structural permanence throughout all values of the design matrix X . The conditional residuals are measures of the difference between the data and the expected value of the data ‘given the rest’. Thus the conditional residual can be viewed as a measure of the local aspects of the fit. Haslett and Hayes (1998) use plots of $\hat{e}_i$ versus $\tilde{e}_{(i)}$ to explore both aspects of the model fit.

1.3.3 Deletion Diagnostics for Influence Analysis

Deletion diagnostics are established tools for influence analysis. As Chatterjee and Hadi (1986) observe “a bewilderingly large number of deletion diagnostic measures have been proposed for the OLS regression model”. Martin (1992) and Haslett and Hayes (1998) provide derivations of these diagnostics for the GLM. These derivations rely heavily on detailed identities associated with the inverses of partitioned matrices. Haslett (1999) outlined the “delete = replace” approach which provided a general simple and intuitive derivation of deletion diagnostics. Intuitively, it may be stated thus: there is an equivalence between (a) the reduction of a model by the deletion of a subset of observations and (b) the replacement of the subset by its BLUP given the remaining observations. Its value is that the reduced model may be analysed in terms of vectors and matrices of the original dimensions, considerably simplifying comparisons with the original fit.

The principal mean structure diagnostics are *DFBETA* and Cook’s Distance. The “delete=replace” approach is described in the context of deriving *DFBETA*. We form $Y_{\bar{a}} = (\tilde{Y}_a(Y_b), Y_b)$. Haslett (1999) shows that $\hat{\beta}(Y_{\bar{a}}) = \hat{\beta}(Y_b)$. $DFBETA_a$ is defined as $\hat{\beta}(Y) - \hat{\beta}(Y_b)$ and may be expressed as $B(Y - Y_{\bar{a}}) = B \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix}$. Thus this approach

provides a very simple derivation of the $DFBETA_a$ statistic for the GLM case. The “delete = replace” approach is exploited in chapters 3 and 4 for developing covariance structure diagnostics.

The $DFBETA_a$ statistic is a measure of the impact on the mean parameter estimates of deleting an arbitrary subset of observations a . An extension of $DFBETA_a$ is its standardised form $DFBETAS_a$ which is the $DFBETA_a$ statistic standardised with respect to the standard errors of the components of $\hat{\beta}(Y) = \hat{\sigma}^2 (X^T V^{-1} X)^{-1}$. A large value of $DFBETAS_a$ indicates that the subset a has a sizeable impact on the regression coefficient. As a result, the analyst must observe $N \times p$ statistics in assessing influence on the mean parameter estimates.

In addition, associated with each data point or subset of data points is a single essentially a composite measure of the influence on the set of coefficients. This quantity or statistic is Cook's Distance, and is defined as

$$DFBETA_a^T (X^T V^{-1} X) DFBETA_a / p \sigma^2 \quad (1.14)$$

Cook's Distance is a measure, representing the standardised distance between the vector of estimated coefficients $\hat{\beta}$ and $\hat{\beta}_{(a)}$. A large value for Cook's Distance implies that the subset a exerts undue influence on the set of coefficients. To determine which specific coefficients are affected one must direct attention to the $DFBETAS_a$ statistics. Martin (1992) describes some variants of Cook's Distance due to different standardisations. These include $\hat{\sigma}_{(a)}^2 (X^T V^{-1} X)^{-1}$ and $\hat{\sigma}_{(a)}^2 (X_{(a)}^T V_{bb}^{-1} X_{(a)})^{-1}$.

In order to externally studentise the diagnostic measures we need an updating formula for $\sigma_{(a)}^2$. The lack-of-fit statistic was defined above as $\hat{e}^T V^{-1} \hat{e} = Y^T Q Y = S(Y)$ and

$\hat{\sigma}^2 = Y^T Q Y / (N - p)$ with $\hat{\sigma}_{(a)}^2 = Y_b^T Q_{(a)} Y_b / (N - p - o)$ where o denotes the number of observations in subset a . It can be easily shown that

$$Y^T Q Y - Y_b^T Q_{(a)} Y_b = \tilde{e}_{(a)}^T Q_{aa} \tilde{e}_{(a)} \quad (1.15)$$

This $M_a = \tilde{e}_{(a)}^T Q_{aa} \tilde{e}_{(a)} / \hat{\sigma}^2$ is the Mahalanobis distance of Y_a from Y_b . It is easy to show that

$$\hat{\sigma}_{(a)}^2 = \frac{N - p - M_a}{N - p - o} \hat{\sigma}^2 \quad (1.16)$$

It is thus easy to ‘studentize’ externally with variances based on $\hat{\sigma}_{(a)}^2$ as for the two variants of Cook’s Distance above. An influence measure based on the above formula is the Relative Variance Change (RVC) which is defined as

$$RVC_{(a)} = \frac{\hat{\sigma}_{(a)}^2}{\hat{\sigma}^2} - 1 \quad (1.17)$$

which from above equals

$$\frac{N - p - M_a}{N - p - o} - 1 \quad (1.18)$$

Outlying values in the Y space will have a large effect on the residual sum of squares. Thus the RVC is a diagnostic measure appropriate for outlier detection in the Y space.

A number of measures based on the impact of deletion on confidence ellipsoids have been proposed. These include the Covariance Ratio (CVR) and the Andrews Pregibon (AP) Statistic.

The Covariance Ratio is defined in equation 1.19 below

$$\text{CVR}_a = \frac{|\hat{\sigma}_{(a)}^2 (\mathbf{X}_{(a)}^T \mathbf{V}_{bb}^{-1} \mathbf{X}_{(a)})^{-1}|}{|\hat{\sigma}^2 (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}|} = \left(\frac{\hat{\sigma}_{(a)}^2}{\hat{\sigma}^2} \right) \frac{|\mathbf{V}^{aa}|}{|\mathbf{Q}_{aa}|} \quad (1.19)$$

where $\mathbf{V}^{aa} = (\mathbf{V}_{aa})^{-1}$. This statistic has no standard error type scaling (Belsley *et al*, 1980). However, it is clear that a value exceeding 1 implies that the subset deleted provides an improvement i.e. a reduction in the estimated generalised variance of the coefficient over what would be produced without the subset of data. A similar statistic in the literature is the AP statistic which is proportional to the reciprocal of the CVR.

$$\text{AP}_a = \frac{(N-p-o)\hat{\sigma}_{(a)}^2 |\mathbf{X}_{(a)}^T \mathbf{V}_{bb}^{-1} \mathbf{X}_{(a)}|}{(n-p)\hat{\sigma}^2 |\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}|} = \frac{(N-p-o)\hat{\sigma}_{(a)}^2 |\mathbf{Q}_{aa}|}{(N-p)\hat{\sigma}^2 |\mathbf{V}^{aa}|} \quad (1.20)$$

As we observed when considering the RVC the deletion of an observation which is outlying in the Y space will result in a marked reduction in the residual sum of squares (RSS) or a large difference between $(N-p)\hat{\sigma}^2$ and $(n-p-o)\hat{\sigma}_{(a)}^2$. An observation which is outlying in the X and V senses will manifest itself in the change in $|\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}|$. Thus the AP statistic is designed to detect observations influential in both the X, V and Y sense (Cook and Weisberg, 1982).

1.3.4 Subsets and Singletons

The various diagnostic measures outlined above are attractive because of their ease of computation for singleton deletion. The entire set of singleton deletions may be computed *en bloc*. In this section we consider subsets involving more than a single observation. Consider a partition P of the indices of Y into blocks a_j ; $j = 1, \dots, k$. It is more challenging to compute $F_{(P)}$ when $k \neq N$, where the subsets are not all singletons. However, the conditional residuals and results in Appendix 1 allow computation of some of these diagnostic measures *en bloc*. Let $D_a = \text{Var}(\hat{e}_{(a)}) = Q_{aa}^{-1}$. It can be easily shown that $\tilde{e}_{(P)} = D_P QY$ where D_P is block diagonal with blocks D_{a_j} , $j = 1, \dots, k$. The entire set of conditional residuals $\tilde{e}_{(P)}$ may also be computed from the following important interrelationships.

$$D_P^{-1} \tilde{e}_P = \Delta^{-1} \tilde{e}_{(\bullet)} = V^{-1} \hat{e} \tag{1.21}$$

The $DFBETA_a$ statistic can be computed from $DFBETA_a = B \begin{pmatrix} \tilde{e}_a \\ 0 \end{pmatrix}$. The $p \times k$ matrix $DFBETA_P$ may be computed from $DFBETA_P = BE$ where E is an $N \times k$ matrix with the j th column of E containing $\tilde{e}_{(a_j)}$ and 0. $DFBETA_P$ will be a $p \times k$ matrix with the j th column containing $DFBETA_{a_j}$. An alternative expression from equation 1.14 for Cook's Distance is given in equation 1.22 below

$$\text{Cook's Distance}_a = \frac{\tilde{e}_{(a)}^T H_{aa} \tilde{e}_{(a)}}{p \hat{\sigma}^2} \tag{1.22}$$

This expression provides easy computation of Cook's Distance $_P$ *en bloc* from

$$\text{Cook's Distance}_p = \tilde{e}_{(p)}^T G E / p \hat{\sigma}^2 \quad (1.23)$$

where G is block diagonal with elements H_{aa_j} and where $H = V^{-1}XB$. An alternative approach to compute Cook's Distance_p, which doesn't involve forming the matrix E is also available. Let the operator $\square_p$ be defined such that for any two vectors f, g of dimension N and a partition P , $f \square_p g$ is a row vector of length k whose elements are $\sum_{i \in a_j} f_i g_i$ (Haslett and Haslett, 2005). Using this notation Cook's Distance_p may be computed as $\tilde{e}_{(p)}^T G \square_p \tilde{e}_{(p)} / p \hat{\sigma}^2$.

The issue of external studentisation was considered in section 1.3.3. and the need for an updating formula for $\hat{\sigma}_{(a)}^2$ highlighted. This was defined as $\hat{\sigma}_{(a)}^2 = \frac{Y^T Q Y - \tilde{e}_{(a)}^T Q_{aa} \tilde{e}_{(a)}}{N - p - o}$.

The computation of $\hat{\sigma}_{(p)}^2$ is quite straightforward. Let S denote the $k \times 1$ matrix with elements $Y^T Q Y$ and $S_{(p)}$ the $k \times 1$ matrix with elements $Y_b^T Q_{(a)} Y_b$. $S_{(p)}$ is computed from $S - \tilde{e}_{(p)}^T D_p^{-1} \square_p \tilde{e}_{(p)}$. Thus $\hat{\sigma}_{(p)}^2 = S_{(p)} / s$ where s is a $k \times 1$ matrix with elements $N - p - o_j$. The $RVC_{(p)}$ is easily computed using these results. It was not possible to develop methods to compute the AP and CVR statistics easily *en bloc* as both statistics involve computing the determinants of various matrices.

1.3.5 Leverage

In the OLS case leverage can be adequately described in terms of the remoteness of an observation in the X space (Chatterjee and Hadi, 1988, pg 95). High leverage cases are identified through the diagonal elements of the H matrix where H is defined as $X(X^T X)^{-1} X^T$ and X is full rank. If H is of full rank then $tr(H) = p$. The average size of a diagonal element is p/n . The values $2p/n$ and $3p/n$ are used as benchmarks with which to compare the elements of H . Draper and John (1981) suggest using the statistic defined in equation 1.24 to identify high leverage subsets for the OLS case. The

statistic is analogous to Wilk's statistic (Cook and Weisberg, 1982, pg 129) which is used to detect a single outlier in multivariate data.

$$\frac{|X_{(a)}^T X_{(a)}|}{|X^T X|} \quad (1.24)$$

Draper and John (1981) advocate this statistic as a generalized measure of leverage that could be used for both subset and singleton deletion. High leverage subsets are identified by small values of this measure. However, they also point out that high leverage cases identified from this measure need not correspond to those identified from H .

In the GLM context the issue of leverage becomes more complex. Schull and Dunne (1988) and consider transforming the data to the OLS situation with $KY = KX\beta + K\varepsilon$ and $\text{Var}(K\varepsilon) = \sigma^2 I$ where K is such that $K^T K = V^{-1}$. The associated H matrix will be $KX(X^T V^{-1} X)^{-1} X^T K^T$. Schull and Dunne (1988) use the eigenvalue and eigenvector decomposition of V to obtain K whilst Puterman (1988) takes K as lower triangular. However, other K are possible with examples considered in the context of the LMM in section 2.3.3. Martin (1992) shows that the most fundamental quantity, which might be considered the generalization of leverage to dependent data for observations is q_{ii} the diagonal elements of Q . An observation demonstrating high leverage will have a small q_{ii} causing $1/q_{ii}$ to be large. The q_{ii} may be scaled by dividing by the corresponding diagonal element of V^{-1} .

The issue of high leverage subsets with $\rho > 1$ receives little attention in the literature for either the case when $V = I$ or for the GLM situations. However, a recent paper by Demidenko and Stukel (2005) outline an approach for identifying high leverage subsets in the context of the LMM. This is considered in chapter 2. The approach may also be applied to the GLM with a single component of variance. Consider

$H_a = X_a(X_a^T V_{aa}^{-1} X_a)^{-1} X_a^T V_{aa}^{-1}$ the H matrix associated with subset a . Then $\sum_{j=1}^k tr(H_{a_j}) = p$ and thus the average value of the $tr(H_{a_j}) = p/k$. They propose $tr(H_{a_j})$ as a leverage measure. A larger than average value of $tr(H_{a_j})$ points to an influential subset j .

The statistic defined in equation 1.2.4 is easily applied to the GLM situation. Using the transformation to OLS with $X^* = V^{-1/2} X$ the corresponding expression for the GLM in equation 1.25 is easily derived.

$$\frac{|X_{(a)}^T V_{bb}^{-1} X_{(a)}|}{|X^T V^{-1} X|} \quad (1.25)$$

Again as for the OLS case high leverage cases identified from this measure need not correspond to those identified from the other methods outlined above.

1.3.6 Local Influence and the Influence Graph

Cook (1986) developed a general method for assessing the local influence of minor perturbations of a statistical model. For a given set of data let $L(\Theta)$ denote the log likelihood corresponding to the model where Θ is a vector of unknown parameters. Perturbations are represented by a matrix ω . The likelihood displacement is defined as

$$LD(\omega) = 2[L(\hat{\Theta}) - L(\hat{\Theta}_\omega)] \quad (1.26)$$

The 'influence graph' of $LD(\omega)$ versus ω as ω varies in Ω , the range of possible perturbations, shows the sensitivity of variables to perturbation. This technique may be used to assess the impact of perturbing the data, the parameters or individual covariates. The method relies on a well behaved likelihood and requires a distribution assumption.

1.3.7 Infinitesimal Influence

Pregibon (1981) introduced the concept of infinitesimal (infinitely small) influence in which the sensitivity of a statistic to a small perturbation of data or model may be considered. As the focus is on infinitesimal change, the derivative is the natural measure of influence. The infinitesimal influence of the i th observation y_i on some function of Y denoted as $F(Y)$ is defined as

$$\frac{dF(Y)}{d(y_i)} \quad (1.27)$$

The influence analysis based on the derivative in equation 1.27 is referred to as infinitesimal data influence analysis. The partial derivative is evaluated at the current data so no re-estimation is needed.

This approach also allows consideration of small perturbations in the model or model influence. Let $l(\theta)$ be the log-likelihood of the postulated model. This model is nested in a more general or “parent” model that is dependent on an additional parameter ω . Let $t(\omega)$ be any statistic or characteristic of interest as a function of ω , e.g., the Maximum Likelihood Estimate (MLE) which maximizes the log-likelihood $l(\theta/\omega)$. Demidenko (2004) defines the influence of t with respect to a possible departure from the postulated model as

$$\left. \frac{dt}{d\omega} \right|_{\omega=0} \quad (1.28)$$

As was the case for data influence the partial derivative is evaluated at the current data so no re-estimation is required.

This approach is similar to Cook's local influence approach outlined above. The difference between the approaches is that Cook (1986) took the likelihood as the measure of model departure. This approach expresses model departure in terms of the characteristic of interest t . Obviously if t is taken as the likelihood displacement both approaches are identical.

The infinitesimal approach is attractive as it is a general tool and can be applied to any statistic to assess the influence of data or model perturbation. It is also easy to compute as the analysis is based on the current estimate and does not require additional computations. It also has the advantage that it does not require a distribution assumption.

1.3.8 Fixed effects joint and conditional influence

There are situations when an observation is not influential individually but taken in a group with other observation may be highly influential. On the other hand an individual observation may be influential on its own but when taken with another observation it may have little influence. These effects are described as "swamping" and "masking" in the literature. In the time series literature it is referred to as "smearing" though the correlated nature of the data worsens the problem. Successive deletions have been used (Lawrence, 1995), (Baade and Pettitt, 2000) to identify conditional influence in the fixed effects.

Haslett (1999) has used the "delete=replace" approach to derive conditional deletion diagnostics. Interest focuses on the deletion of Y_i given that Y_j has already been deleted.

The results extend from partitioning Y as (Y_a, Y_b, Y_c) which may also be written as (Y_{ab}, Y_c) with $\hat{\beta}_{(ab)} = \hat{\beta}(Y_c)$ and $\tilde{e}_{(ab)} = Y_{ab} - \tilde{Y}_{ab}(Y_c)$. $\hat{\beta}_{(ab)}$ is then contrasted with $\hat{\beta}_{(b)}$ and this is a measure of the effect on $\hat{\beta}$ of deleting Y_a given that Y_b had already been deleted. Following from this

$$\hat{\beta}_{(b)} - \hat{\beta}_{(ab)} = (\hat{\beta} - \hat{\beta}_{(ab)}) - (\hat{\beta} - \hat{\beta}_{(b)}) = B_{[ab]} \tilde{e}_{(ab)} - B_{[b]} \tilde{e}_{(b)} = B_{[ab]} \tilde{e}_{(a|b)} \quad (1.29)$$

where $B_{[ab]}$ denotes the columns of B corresponding to Y_a and to Y_b and $\tilde{e}_{(a|b)}$ is defined as $\tilde{e}_{(ab)} - (0_a, \tilde{e}_{(b)})$. The conditional Cook's Distance is defined as

$$D_{a|b} = \tilde{e}_{(a|b)}^T H_{[ab,ab]} \tilde{e}_{(a|b)} \quad (1.30)$$

where $H_{[ab,ab]}$ refers to the block of H associated with the $[ab]$ rows and columns. The conditional deletion of subsets and their impact on the variance parameter $\hat{\sigma}^2$ can be also be derived. $S(Y_{a|b}) = S(Y_{(b)}) - S(Y_{(ab)}) = S(Y) - S(Y_{(b)}) - S(Y) - S(Y_{(ab)})$. Therefore

$$\hat{\sigma}_{a|b}^2 = \frac{\tilde{e}_{(ab)}^T Q_{[ab,ab]} \tilde{e}_{(ab)} - \tilde{e}_{(b)}^T Q_{bb} \tilde{e}_{(b)}}{N - p - m} \quad (1.31)$$

where m is the size of subset (ab) .

1.4 Conclusions

Model criticism is an essential element of the model building process. There is an extensive literature on diagnostic measures for the LM. These diagnostics help the data analyst to test model assumptions and identify outlying and influential observations. The methodologies fall into two broad categories, residual analysis and influence diagnostic techniques.

There is a substantial literature on the use of residuals for checking model assumptions and exploring model fit for the simple linear model with $\text{Var}(Y) \propto I$. However, for the GLM the literature is sparse. McCulloch and Searle (2001) contains no reference to residuals or model checking. Christensen (1987) devotes a chapter (Ch 13) to residuals.

However, it focuses on the simple case of $\text{Var}(Y) \propto I$ only. Whittaker (1990) in considering graphical models for multivariate data and Cox and Wermuth (1998) also in the multivariate context mention residuals only in the OLS setting.

A large number of deletion diagnostic measures have been proposed to study influential observations in OLS linear regression analysis. Many of these measures have been extended to the GLM and are summarised by Martin (1992) and Haslett and Hayes (1998). These extensions are based on the assumption that the covariance matrix V is unchanged following deletion. The assumption allows the various measures to be easily derived. However, the impact of this assumption on the accuracy of the diagnostics is not clear from the literature.

The extent to which the various diagnostic methodologies proposed in the literature may be applied to the LM family of models varies. The focus of this thesis is deletion diagnostics for the LMM. The LMM may have more than one variance component. This fact makes diagnostic analysis for the LMM more complex. The various diagnostic measures are considered in the context of the LMM in chapter 2.

Chapter 2

Diagnostics for the LMM

2.0 Introduction

This chapter focuses on the LMM. The model is an important member of the LM family and the focus of this thesis is on deletion diagnostics for the LMM. In chapter one, the various diagnostic methodologies and measures were outlined in the context of the GLM with a single component of variance. The LMM may have more than one component of variance, which makes a diagnostic analysis more complex for this model. In this chapter, the diagnostics are reviewed in the context of the LMM.

Diagnostic analyses arise at two stages in the LMM. These are at the estimation of the covariance structure parameters and at the subsequent estimation of the mean structure parameters. Mean and covariance structure diagnostics for the model from the current literature are reviewed and the need for such methods highlighted. Inadequacies and shortcomings in the diagnostics proposed to date in the literature are clearly enunciated and the issues arising from this chapter act as the motivation for work in the following chapters.

2.1 The Linear Mixed Model

LMMs appear under a variety of titles in diverse literatures. In sociological research, they are often referred to as multilevel linear mode (Mason *et al*, 1983). Lindley and Smith (1972) introduced the term hierarchical linear models. This was adopted later by Bryk and Raudenbush (1992) and by many others since. In the biometrics literature the terms mixed effects models and random-effects models are common. Examples include Laird and Ware (1982) and Diggle (1990). They are also referred to as random-coefficient regression models in the econometrics literature while in some statistical literature they are referred to as covariance component models (Dempster *et al*, 1981), (Longford, 1987). The many areas of application of this model suggest that simply computable diagnostics would have wide appeal.

Developments in numerical approaches to covariance component estimation allied with advances in statistical software have allowed mixed and random-effects models to be applied extensively. Longitudinal studies are an active area of application for the LMM. Hand and Crowder (1996), Lindsey (1993) and Diggle *et al* (1994) all contain chapters describing the application of LMMs in the analysis of longitudinal data. In such studies a response is measured repeatedly over time on a number of experimental units or subjects. The focus of any diagnostic analysis for these studies is likely to centre on the experimental units as well as on the individual observations.

It is clear that any proposed diagnostics must be easily computable for both observation and subset (other than singleton) deletion, as a thorough diagnostic analysis should involve both. In the context of longitudinal data, Ouwens (2001) demonstrates the necessity to use observation influence measures in addition to subject based influence measures as existing subject influence measures may fail to detect influential subjects due to the relative position of the observations within and across subjects.

The LMM was defined in 1.5. with the variance covariance matrix $V = ZDZ^T + A$ where $A = \text{Var}(\varepsilon)$. An alternative expression for $V = A + \sum_{i=1}^r \sigma_i^2 Z_i Z_i^T$ with σ_i^2 denoting the variance of the random parameters γ_i and Z_i is a design matrix relating γ_i to Y . Models for $A = \text{Var}(\varepsilon)$ describe the auto-covariance structure in the data. The matrix A is often of the form $\sigma_0^2 I$. However, when there is autocorrelation in the data A will include a correlation parameter ρ . This is common in time series and longitudinal data. Various models for A have been proposed in the literature. Hand and Crowder (1996, chapter 6) and Diggle Liang and Zeger (1994, chapter 5) outline the principal parametric models for the auto-covariance structure.

There are two stages in the fitting of such models; these may be carried out simultaneously or sequentially. The first is the modelling and estimation of the covariance structure V . The second stage in fitting such models involves the estimation of the fixed parameters β . The estimation of these involves the use of the BLUE of β which is well known from the classical theory to be $\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$. The standard estimation methods used at both stages of the fitting process are sensitive to unusual observations. Therefore any diagnostic analysis needs to be conducted at both stages for LMMs.

The wide usage of the LMM as reflected in the literature indicates it is likely that easily computable diagnostics for the model (the focus of this thesis) would be appealing and have wide usage. In this thesis diagnostics for both stages in the fitting process are considered. In chapter 3 cheaply computable diagnostics for the variance components are proposed. Chapter 4 develops approaches for assessing the influence of observations on autoregressive (AR) and moving average (MA) parameters. These models are commonly

used to model the autocovariance structure in data. These methods are also applicable for the important compound symmetry case. The focus in Chapter 5 is on mean structure diagnostics for the LMM. The methods developed in the thesis exploit elements of the various model fitting and criticism procedures in the literature and are considered in sections 2.2 and 2.3 respectively.

2.2 Methods of estimation

Estimation methods for the LMM are detailed in many books and papers, the principal approaches being ANOVA, maximum likelihood (ML), restricted maximum likelihood (REML) and minimum norm quadratic unbiased estimation (MINQUE). ANOVA and REML methods are used in chapters 3 and 4 to develop diagnostic methods for the variance and correlation parameters in LMMs. This consideration of estimation methods is confined to these approaches.

2.2.1 ANOVA methods and Henderson's Method III

Various ANOVA approaches have been developed to estimate variance components for random and mixed models. For balanced designs, they produce unique, computationally inexpensive and intuitive estimates. Indeed for the balanced case the solutions to the REML equations are identical to ANOVA estimators (Searle *et al.*, 1992). However, for unbalanced data there is no unique set of sums of squares that can be used. Early extensions of ANOVA methods to unbalanced data focused on the 1-way classification. These included work by Cochran (1939) and Winsor and Clarke (1940).

Henderson (1953) extended the approach to higher order classifications motivated by his interest in estimating variance components in a genetics setting where available data can be voluminous but severely unbalanced. The approaches outlined in this paper became known as Henderson's Methods I, II and III. Method III is applicable to all mixed

models unlike the other methods, which are restricted to the class of models that they may be applied to. Method I is applicable only to random models while Method II cannot be used if there are interactions between the fixed and random effects. The motivation for outlining Method III below stems from its usage in section 3.1.1 to develop diagnostics for the variance components in the LMM

In the formulation of the LMM in equation 1.5 there was a clear distinction between the fixed and random effects. In considering Method III the fixed effects β and random effects γ are combined into a single vector η and the model may be re-expressed as $Y = M_r \eta + \varepsilon$ with M_r consisting of $(X, Z_1, Z_2, \dots, Z_r)$ and having dimension $N \times (p + q)$. Similarly M_{r-1} consists of $(X, Z_1, Z_2, \dots, Z_{r-1})$. Let $Q_r = I - M_r (M_r^T M_r)^{-1} M_r^T$. The variance components may be estimated from taking expected values of the differences in the "sums of squares" $S_r = Y^T Q_r Y$. To estimate σ_r^2 we consider the expected value of the difference between S_r and S_{r-1}

$$E[S_{r-1} - S_r] = \text{tr}[(Q_{r-1} - Q_r)V] + \beta^T X^T (Q_{r-1} - Q_r) X \beta \quad (2.1)$$

It can be easily shown that $Q_{r-1} X$ and $Q_r X = 0$ and therefore $\beta^T X^T (Q_{r-1} - Q_r) X \beta = 0$

Re-expressing V as $\sigma_0^2 + \sum_{i=1}^r \sigma_i^2 Z_i Z_i^T$ we can express $(Q_{r-1} - Q_r)V$ as

$$(Q_{r-1} - Q_r)\sigma_0^2 + \sum_{i=1}^r \sigma_i^2 (Q_{r-1} - Q_r) Z_i Z_i^T = (Q_{r-1} - Q_r)\sigma_0^2 + \sigma_r^2 (Q_{r-1} - Q_r) Z_r Z_r^T$$

because $(Q_{r-1} - Q_r) Z_i Z_i^T = 0$ for $i = 1, \dots, r-1$. Given these results

$$\begin{aligned} E(S_{r-1} - S_r) &= \text{tr}[(Q_{r-1} - Q_r)\sigma_0^2 + (Q_{r-1} - Q_r) Z_r Z_r^T \sigma_r^2] \\ &= \sigma_0^2 \text{tr}[(Q_{r-1} - Q_r)] + \sigma_r^2 \text{tr}[(Q_{r-1} - Q_r) Z_r Z_r^T] \end{aligned} \quad (2.2)$$

A similar argument shows that

$$E(S_r) = tr(Q_r V) = \sigma_0^2 tr[Q_r]$$

An unbiased estimate of σ_0^2 is thus $\hat{\sigma}_0^2 = S_r / tr(Q_r)$

An unbiased estimate of σ_r^2 is

$$\hat{\sigma}_r^2 = [S_{r-1} - S_r - \hat{\sigma}_0^2 tr(Q_{r-1} - Q_r)] / tr[(Q_{r-1} - Q_r) Z_r Z_r^T] \quad (2.3)$$

In a similar fashion we can estimate σ_{r-1}^2 by taking expectations of sums of squares based on M_{r-1} and M_{r-2} which gives

$$\begin{aligned} E[Y^T (Q_{r-2} - Q_{r-1}) Y] &= tr[(Q_{r-2} - Q_{r-1}) V] \\ &= tr[(Q_{r-2} - Q_{r-1}) \sigma_0^2 + (Q_{r-2} - Q_{r-1}) Z_{r-1} Z_{r-1}^T \sigma_{r-1}^2 + (Q_{r-2} - Q_{r-1}) Z_r Z_r^T \sigma_r^2] \end{aligned}$$

An unbiased estimate σ_{r-1}^2 is thus

$$[S_{r-2} - S_{r-1} - \sigma_0^2 tr(Q_{r-2} - Q_{r-1}) - \sigma_r^2 tr((Q_{r-2} - Q_{r-1}) Z_r Z_r^T)] / tr[(Q_{r-2} - Q_{r-1}) Z_{r-1} Z_{r-1}^T] \quad (2.4)$$

The remaining variance parameters $\sigma_1^2, \sigma_2^2, \dots, \sigma_{r-1}^2$ can be estimated in a similar fashion.

This approach is developed in section 3.1.1 to provide estimates of the change in the variance components following subset deletion.

2.2.2 REML Estimation

The other estimation procedure exploited in chapter 3 is REML. The REML method was introduced by Patterson and Thompson (1971) as a way of estimating variance components for the GLM. The impetus for this approach to parameter estimation, in this case arises from the inability of standard ML estimation to produce unbiased estimates. This is well known in the case of the LM with independent errors. The maximum likelihood estimate of σ^2 is RSS/N where N is the number of data values. This is a biased estimate of σ^2 whereas the REML estimator of σ^2 is the unbiased estimator $\hat{\sigma}^2 = RSS/N - p$. Thus REML estimation gives an unbiased estimate for σ^2 . It can be defined as a ML estimator based on a linearly transformed set of data $Y^* = AY$ such that

the distribution of Y^* does not depend on β . One approach to achieve this is by taking A to be the matrix which converts Y to ordinary least square residuals $A = I - (X^T X)^{-1} X^T$. Then Y^* has a multivariate normal distribution with mean zero whatever the value of β . Any full rank matrix with the property $E[Y^*] = 0$ for all β will give the same answer.

The equations for solving iteratively the REML estimates are

$$\sum_{j=1}^r \sigma_j^2 \text{tr}[Z_j Z_j^T Q Z_i Z_i^T Q] = Y^T Q Z_i Z_i^T Q Y \quad (2.5)$$

for $i = 0, \dots, r$ and where Q is as defined in section 1.3.1 (Christensen, 1987, pg.237).

REML is routinely used to estimate the variance components in LMMs. Various studies have considered its merits relative to ML. These may be summarized by Diggle *et al*'s (1994, pg. 68) observation that in many cases both will give very similar results but when they do differ substantially REML estimates are preferable.

When computing deletion diagnostics the data will often be unbalanced after deleting a subset. There is agreement in the literature that for unbalanced data REML is preferable to ANOVA approaches for estimating the variance components. Historically ANOVA methods were preferred as REML was impractical because of its computing requirements. However, this is no longer an issue. Nevertheless, ANOVA methods are likely to continue being used (Searle *et al*, 1992, pg.168). This is because many researchers are comfortable with the closed form ANOVA estimators and find ML or REML approaches too theoretic and are overawed by the mathematics involved. Therefore in chapter 3 diagnostic methods based on both ANOVA and REML are proposed.

2.3 Diagnostics for the LMM

As outlined in chapter 1, there are two principal strands to any diagnostic analysis. These are residual analysis and influence analysis. For LMMs with more than one variance component, these analyses become more complex. The consideration of these issues is relatively sparse in the literature. For example, McCulloch and Searle (2001) contains no reference to diagnostics in the entire text and Christensen (1987) focuses his discussion exclusively on the LM with independent errors.

2.3.1 Residual analysis for the LMM

The marginal residuals $\hat{e}$ defined in equation 1.8 are estimates of the true model errors $\varepsilon = Y - X\beta$. For the LMM these errors may be defined by a number of independent underlying variables. The estimates of these underlying variables are termed innovation residuals (Haslett and Haslett, 2005). In time series the innovations are the white noise processes that drive the model and the term is drawn from the time series literature. The marginal residuals for the LMM defined in equation 1.5 incorporate the innovation residuals with $\hat{e} = Z\hat{\gamma} + \hat{\varepsilon}$. The ε are "pure error" and the r terms γ_j have interpretations as independent random deviations from an overall mean structure for subgroups of the observations Y . The innovation residuals are estimated as $Z\hat{\gamma}$ and $\hat{e} - Z\hat{\gamma}$ with the random effects parameters estimated by $\hat{\gamma} = DZ^T V^{-1} \hat{e} = DZ^T QY$ (Searle *et al*, 1992). They are related to the conditional residuals with $\hat{\gamma} = DZ^T \Delta^{-1} \tilde{e}_{(\cdot)}$. The term 'innovation' for these residuals is only appropriate if $P \propto I$ and D is diagonal.

Several papers have considered residuals for the LMM defined in equation 1.5 with $P \propto I$ and D being diagonal. Many distinguish between $\hat{\varepsilon}$ and $Z\hat{\gamma}$. Lange and Ryan (1989,pg628) and Oman (1995) refer to the $\hat{\varepsilon}$ as "residuals". Oman (1995) uses the term "model" residuals for the $Z\hat{\gamma}$. Longford (1993, pg66) calls the $\hat{\varepsilon}$ and the $\hat{\gamma}$ "elementary level" and "cluster level" residuals respectively. An interesting point worth noting is the

natural interpretation of residuals associated with individual elements of γ . Particularly for longitudinal data, subsets comprising multiple observations are referred to as "subjects" or "cases". Weiss and Lazaro (1992) in the context of longitudinal data analysis describe a useful exploratory approach for model specification and for the identification of anomalous data based on plotting these residuals against time. Waternaux *et al* (1989) and Oman (1995) use residuals based on the γ to check model assumptions and identify outlying observations.

The standardized conditional residual $\tilde{e}_{(a)}$ of equation 1.9 is defined as $\tilde{e}_{(a)} / \text{Var}(\tilde{e}_{(a)})^{1/2}$.

This statistic may be externally or internally studentised. The denominators for the internally and externally studentised versions are $(\hat{\sigma}^2(Q_{aa})^{-1})^{1/2}$ and $(\hat{\sigma}_{(a)}^2(Q_{aa})^{-1})^{1/2}$ respectively. For the single component of variance case either version is applicable. However, for the LMM with more than one variance component external studentisation is problematic. There is a need to update each variance component following deletion. There is no cheap computational method for doing this. Externally studentised conditional residuals are not computationally feasible in this situation. Methods for estimating the change in the variance components following deletion are proposed in chapter 3.

The literature on residual analysis for the LMM is relatively sparse. McCulloch and Searle (2001) contains no reference to residuals or model checking. Christensen (1996) devotes a chapter (Ch 13) to residuals though it focuses on the simple case of $\text{Var}(Y) \propto I$ only. It is clear that for the LMM residual analysis is more challenging than for the simple OLS case or indeed for the GLM with a single component of variance. Currently residual analysis for the LMM with more than a single source of variation is not routinely used and is not an element in the principal statistical packages.

2.3.2 Deletion Diagnostics for the LMM

Within the LMM deletion diagnostics are not routinely used despite huge advances in computing. Diagnostic issues arise at two stages in the LMM. These are (a) the estimation of V by some $\hat{V}$ and (b) the subsequent estimation of the regression coefficients β and γ given $\hat{V}$. Christensen *et al* (1992) propose that influence analysis should begin at the first stage. However, the majority of the research on diagnostics for the LMM focuses on (b). Hodges (1998) and its discussion reviewed much of this literature under the heading of the hierarchical model. Atkinson (1998) in the discussion of Hodges (1998) remarks on the fact that $\hat{V}$ is taken as fixed and observes that this is a weakness in such research. Related work by Tan *et al* (2001), Ouwens *et al* (2001), Bannerjee (1998) and Bannerjee and Frees (1997) all assume $\hat{V}$ is fixed following deletion. The impact of this assumption is explored in section 5.1 of chapter 5. The various deletion diagnostics outlined in section 1.3.3 are reviewed in the context of the LMM below.

The principal mean structure diagnostics are *DFBETA* and Cook's Distance. These diagnostics are used to identify data influential on the fixed parameter estimates. In many studies involving LMMs the random parameters may be the focus of interest. An analogous diagnostic based on *DFBETA* for the random parameters *DFGAMMA_a* is $\hat{\gamma} - \hat{\gamma}_{(a)}$. Haslett and Dillane (2004) show that this may be estimated as $\hat{\gamma} - \hat{\gamma}_{(a)} = DZQ \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix}$. This may also be standardized with respect to the standard error of the components of $\gamma = DZQZ^T D$. The standardized version would correspond to the *DFBETAS* statistic. Using results in section 1.3.4 $DFGAMMA_p = DZQE$ and thus *DFGAMMA_p* can be easily computed *en bloc*.

The most popular influence measure in the regression literature is Cook's Distance (Bannerjee, 1998). However, a number of authors have demonstrated the limited effectiveness of the diagnostic in the LMM context. A general criticism of Cook's Distance is as a composite measure, an observation with a moderate influence on all regression coefficients may be judged more influential than one with a large influence on one coefficient and negligible influence on all others. This has led to the development of the idea of partial influence. Partial influence involves examining the influence on individual parameters or on a given subset of parameters. Cook and Weisberg (1982) defined the partial influence of the i th observation on $\hat{\Theta}$ where $\hat{\Theta} = L\beta$ that is a linear combination of the mean parameters β as

$$C_i(L) = (\hat{\Theta} - \hat{\Theta}_{(i)})^T (L^T X^T V^{-1} X L) (\hat{\Theta} - \hat{\Theta}_{(i)}) \quad (2.6)$$

Bannerjee and Frees (1997) and Bannerjee (1998) use the partial influence approach to explore the impact on subject specific parameters and on the population parameters of deletion. Tan *et al* (2001) demonstrate that influential observations having a large effect on the subject specific parameters are not always detected by Cook's Distance owing to large between subject variation. They propose a conditional version of Cook's Distance by conditioning on the subjects. The statistic is decomposed to measure the impact on the fixed parameters and on the random parameters. It is defined as

$$C_{cond} = \frac{(\hat{\beta} - \hat{\beta}_{(a)})^T X^T X (\hat{\beta} - \hat{\beta}_{(a)})}{k\sigma^2} + \frac{(\hat{\gamma} - \hat{\gamma}_{(a)})^T Z^T Z (\hat{\gamma} - \hat{\gamma}_{(a)})}{k\sigma^2} + \frac{2(\hat{\beta} - \hat{\beta}_{(a)})^T X^T Z (\hat{\gamma} - \hat{\gamma}_{(a)})}{k\sigma^2} \quad (2.7)$$

The first term may be interpreted as the impact of deletion on the fixed effects, the second term as the impact on the subject specific random parameters and the third term as

a measure of the covariation between the change in the mean profile and a change in the position of the subject specific profiles relative to the mean profile. They remark without supporting analysis that the third element can be neglected. This approach is also based on the assumption that V is unchanged following deletion.

The RVC defined in equation 1.17 is one of the few diagnostics that can be easily applied to the multiple variance component LMM case. Longford (1993) observes that the RVC is applicable for the LMM as a means of assessing the impact of deletion on each of the variance parameters $\hat{\sigma}_0^2, \dots, \hat{\sigma}_r^2$. It is used for this purpose in chapter 3

The CVR and AP statistic are significantly more difficult to compute for the LMM with multiple variance components. The issue is similar to that encountered with external studentisation. The CVR simplifies for the single component of variance case as outlined in equation 1.19. There is no easily computable accurate estimate of $V_{(a)}$ unless it is re-estimated. If the estimate is to be accurate each of the variance components needs to be updated.

2.3.3 Leverage

For the LMM both the covariates and the random effects may influence the leverage measure. Demidenko and Stukel (2005) propose a methodology for considering leverage in the LMM. They propose a leverage measure comprising two components; one measures high leverage data in the X space and the other in the Z space. They consider leverage as the partial derivative of the predicted value with respect to the corresponding dependent variable. For the LMM $\hat{Y} = X\hat{\beta} + Z\hat{\gamma}$ is the predicted outcome conditional on the estimated random effects $\hat{\gamma}$ with $H = \frac{d\hat{Y}}{dY} = X(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1} + Z\hat{D}Z^TQ = H_1 + H_2$. High values of h_{1ii} and h_{2ii} the diagonal elements of H_1 and H_2 respectively indicate

high leverage observations in either the X or Z spaces. For any subset a_j with $o > 1$ and $j = 1, \dots, k$ the leverage matrix for subset $a = H_a = X_a (X_a^T V_{aa}^{-1} X_a)^{-1} X_a^T V_{aa}^{-1} + Z_a D_{aa} Z_a' Q_{aa}$. A larger than average value for $tr(H_{a1})$ indicates an influential subset in the X space and for $tr(H_{a2})$ one in the Z space. An average value for $tr(H_{a1})$ is not obvious as leverage is affected by subset size so it is difficult to know when a subset is larger than average if the subsets are not of equal size. However, the approach does provide a leverage measure for both the covariates and random effects.

Haslett and Haslett (2005) suggest an alternative approach that provides a measure of leverage for the random effects. They suggest transforming the data to the OLS situations as outlined in section 1.3.5. They define K as the $(n + q) \times n$ non-square square root of V with the form $K = (\sigma_1^2 Z_1, \dots, \sigma_r^2 Z_r, \sigma_0^2 I)$. The H matrix $KX(X^T V^{-1} X)^{-1} X^T K^T$ provides a leverage measure for the subsets corresponding to the random effects and for each observation. The obvious difficulty with this approach is that it cannot be used to measure the leverage of any arbitrary subset of the data. It will only measure the leverage of the subsets matching the q random effects.

2.3.4 Local Influence and the Influence Graph

Cook's local influence approach has been applied to the LMM. Pan and Fang (2002) and Verbeke and Molenberghs (2000) provide an extensive discussion on the use of local influence for the LMM. Beckman *et al* (1987) focus on observation deletion while Lesaffre and Verbeke (1998) focus on longitudinal data models where often the influence of subjects rather than individual observations is of interest. As outlined in section 1.3.5 this approach measures the sensitivity by the maximum curvature of the log-likelihood function. This technique may be used to assess the impact of perturbing the data, the parameters or individual covariates. The method relies on a well-behaved likelihood. It also requires a distribution assumption. Demidenko (2004) in advocating the more

general infinitesimal influence approach identifies this as a drawback. Infinitesimal influence also measures influence on the parameter of interest directly, unlike local influence, where the measure of influence is the log-likelihood displacement. The dependence on ML is not entirely satisfactory. Indeed Lesaffre *et al.* (1999, section 5.1) recommended that their diagnostics be used only to 'flag' apparently influential cases for subsequent full REML refit.

2.3.5 Infinitesimal influence.

Demidenko (2004) develops Pregibon's (1981) infinitesimal influence and considers its application to the LMM. The desirable attributes of infinitesimal influence are outlined in section 1.3.7. Principal amongst these is its versatility in that it can be applied to any particular statistic of interest. However, for the LMM the approach becomes more complex if one is to take account of the variance components in considering the influence of other statistics. In such situations a more computationally frugal approach would be desirable. Demidenko's (2004) use of this approach for the LMM is unique in the literature. Indeed for the OLS regression model the approach has not been widely adopted. It is clear from their use in most statistical packages that cheaply computable deletion diagnostics are preferred by software developers for the OLS case. This is also likely to be the case for the LMM.

2.4 Influence diagnostics for the covariance structure

By comparison with fixed effect diagnostics a lot less attention has been given to covariance structure diagnostics in the literature. Various contributions in the LMM literature have addressed the issues of influence and outlier detection for such models; however, many focus on the fixed parameters with V being regarded as a nuisance parameter. As outlined earlier such papers include Hodges (1998), Bannerjee (1998), Bannerjee and Frees (1997) and Tan *et al* (2001).

Christensen *et al.* (1992) proposed a one-step approximation of the change in the variance components when an observation or subset of observations is deleted. The equations for finding the REML estimates were outlined in equation 2.5. Starting values for the σ_j^2 s are selected and the estimates derived when the iterations converge. The one-step approximations are estimated from

$$\sum_{j=0}^r \sigma_{j(a)}^2 \text{tr}[Z_{j(a)} Z_{j(a)}^T Q_{(a)} Z_{i(a)} Z_{i(a)}^T Q_{(a)} = Y_b^T Q_{(a)} Z_{i(a)} Z_{i(a)}^T Q_{(a)} Y_b] \quad (2.8)$$

with $i = 0, \dots, r$ and $Q_{(a)}$ based on $V(Y)$. This approach gives estimates of the change in the variance components for both case and subset deletion and is a measure of influence. However, the method must be applied for each observation separately as the estimates cannot be estimated as a result of a single matrix operation. Oman (1995) is critical of the approach because it is computationally expensive.

Bruce and Martin (1989) have explored the application of deletion diagnostics for time series ARIMA models. Their approach differs to the previous methods in that all parameters are re-estimated using ML when observations are deleted. The measures considered mirror the diagnostic measures *DFBETA* and *RVC* outlined earlier. These diagnostics measure the change in the parameter vector and the change in the estimated variance and are defined as

$$DC(A) = n(\hat{\beta} - \hat{\beta}_{(a)})(X^T V^{-1} X)(\hat{\beta} - \hat{\beta}_{(a)}) \quad (2.9)$$

and

$$DV(A) = \frac{N}{2} \left(\frac{\sigma^2}{\sigma_{(a)}^2} - 1 \right)^2 \quad (2.10)$$

with ML estimates $\hat{\beta}_{(a)}$ and $\hat{\sigma}_{(a)}^2$ estimated from the reduced model. The authors compare both diagnostics using various data and conclude that $DC(A)$ was susceptible to smearing but that $DV(A)$ was more robust. They conclude that an outlier can appear significant in $DV(A)$ for two reasons. It inflates the variance by distorting the parameter estimates of the ARIMA coefficients or it inflates the variance because of a large associated residual. Thus they derive an expression for each of the two components and suggest plotting the two against each other.

This paper is unusual in that the deletion diagnostics are based on a full re-fit of the reduced model. The consequences of misspecifying the covariance structure on the mean parameters include incorrect inferences about the regression parameters β and inefficient estimates of these parameters. However, the impact of not estimating the variance and correlation parameters when observations are deleted on the fixed effect diagnostic measures has not been explored in the literature. This issue is considered in detail in chapter 5.

2.5 Conclusions

Model criticism is an essential element of the model building process. There is an extensive literature on diagnostic measures for the LM. These diagnostics help the data analyst to test model assumptions and identify outlying and influential observations. The attraction of Cook's (1977) methods is that the effects on $\hat{\beta}$ of deleting each observation can be cheaply computed as a by-product of the fitting procedure. Thus for the LM with independent errors the principal statistical software packages routinely compute these

measures with the model fit. However, for more complex error structures as for the LMM deletion diagnostics are still not routinely used despite advances in computing.

For such models two elements of influence arise. These are in the estimation of V and the subsequent estimation of the mean parameters β . The principal challenge in the first stage is to derive measures of influence that can be cheaply computed as a by-product of the model fitting process as is the case for the mean parameters.

For the second element of influence the general approach has been to treat V as fixed and assume it is unchanged when observations are deleted. There has been little work in the literature to explore the effect of this approach on the accuracy of the diagnostics.

The alternative approach is to re-fit the model upon deletion and compute the diagnostics based on this re-fit. The obvious difficulty with this approach is the computational cost involved. Despite advances in computing power it is unlikely that software developers would apply this approach. Thus the challenge is to develop an intermediate approach that allows us to measure the impact on the variance structure parameters of deletion and then, given this information, to compute the mean structure diagnostics. It is also clear from the literature that these methods must allow for both observation and subset deletion. The development of such techniques and methodologies is the focus of the remaining chapters of this thesis.

Chapter 3 focuses on diagnostics for the variance components in the LMM. Chapter 4 is concerned with estimating the impact of deletion on AR and MA model parameters as these models are often used to model autocorrelation in data. The methods proposed in these chapters are based on the “delete = replace” results outlined in appendix 1. In chapter 5, an intermediate approach using a Taylor expansion is proposed to estimate the

impact of updating the mean structure diagnostics, given estimates of the change in the variance structure following deletion.

Chapter 3

Variance Component Diagnostics for LMMs

3.0 Introduction

In this chapter, methods are developed to identify individual (or groups of) observations that are influential on the estimation of the variance components in LMMs. The issue of influence on the variance components was considered in section 2.4. The challenge is to develop methodologies that are insightful, are a by-product of the fitting process and are computationally efficient. The ‘delete = replace’ approach is exploited to develop variance component diagnostics based on ANOVA and REML estimation techniques. The proposed methods allow both observation and subset deletion and are a by-product of the fitting process. The methodological approaches outlined in this chapter were first outlined in Haslett and Dillane (2004).

The development of computationally frugal deletion diagnostics for estimates of variance components is the primary contribution in this chapter. The effect of the deletion of individual observations, of “subjects” and of arbitrary subsets is illustrated. The proposed methods are applied to a number of datasets. Their relative merits and deficiencies are considered based on their application to these data. The methods proposed in this chapter are exploited in chapter 5 to develop deletion diagnostics for the fixed effects, which incorporate changes in the variance components following deletion.

3.1 Subset deletion for components of variance in the LMM.

The LMM defined in equation 1.5 may be expressed as

$$Y = X\beta + \varepsilon$$

with $V \equiv \text{Var}(\varepsilon) = \sum_{j=0}^r \sigma_j^2 Z_j Z_j^T$. ANOVA and REML approaches for estimating the vector of variance components $\tau = (\sigma_0^2, \sigma_1^2, \dots, \sigma_r^2)^T$ were outlined in sections 2.2.1 and 2.2.2 respectively. In this chapter, the focus is on the estimates $\hat{\tau}_{(a_i)}$ following deletion of subsets Y_{a_i} ($i = 1, \dots, k$). The objective of the approaches outlined in this chapter is to develop procedures that produce the entire set of estimates $\hat{\tau}_{(P)} = \{\hat{\tau}_{(a_1)}, \hat{\tau}_{(a_2)}, \dots, \hat{\tau}_{(a_k)}\}$ (an $(r+1) \times k$ matrix) arising by deleting each of the Y_{a_i} in turn. The proposed methods are based on ANOVA methods (Henderson's Method III), Christensen, Pearson and Johnson's (CPJ's) (1992) one-step approximation and an alternative approximation also based on the REML equations described in Haslett and Dillane (2004).

3.1.1 Henderson's Method III based approach

In section 2.2.1 Henderson's Method III for estimating the variance components was outlined. In this section the method is used to estimate the change in the variance components following deletion. The 'delete = replace' approach is used to update the various quantities. To obtain an estimate of $\sigma_{r(a)}^2$ using this ANOVA approach, which we denote by $\tilde{\sigma}_{r(a)}^2$, we need updating formulae based on the model fit for the quadratic forms and the traces associated with the reduced model. Using the results outlined in appendix 1 these are easily derived. For example, using Henderson's Method III described in section 2.2.1 and results outlined in appendix 1

$$\tilde{\sigma}_{o(a)}^2 = (Y^T Q_r Y - \tilde{e}_{(a)}^T Q_{raa} \tilde{e}_{(a)}) / \text{tr}[Q_r W_a] \quad (3.1)$$

and

$$\begin{aligned} \check{\sigma}_{r(a)}^2 = & (Y^T Q_{r-1} Y - \tilde{e}_{(a)}^T Q_{r-1_{aa}} \tilde{e}_{(a)} - Y^T Q_r Y + \tilde{e}_{(a)}^T Q_{r_{aa}} \tilde{e}_{(a)} \\ & - \check{\sigma}_{0(a)}^2 (tr[Q_{r-1} W_a] - tr[Q_r W_a])) / tr[Q_{r-1} Z_r Z_r^T W_a] - tr[Q_r Z_r Z_r^T W_a] \end{aligned} \quad (3.2)$$

Using the conditional residual results outlined in appendix 1

$\check{\sigma}_{(a_i)} = (\check{\sigma}_{0(a_i)}^2, \check{\sigma}_{1(a_i)}^2, \dots, \check{\sigma}_{r(a_i)}^2)$ may then be estimated. Repeating the process for Y_{a_i} ($i = 1, \dots, k$) produces estimates of $\check{\sigma}_{(p)}$. As outlined in section 2.2.2 many researchers are still more comfortable with ANOVA techniques than optimisation methods like REML. This approach offers these individuals an ANOVA based approach for identifying data influential on estimates of the variance components. However, the approach is not as computationally economical as those based on REML estimation outlined below.

3.1.2 Methods based on REML Estimation

REML Estimation

The REML equations (equations 2.5) for estimating the $\hat{\sigma}_j^2$ are based on equating the sum of squares $s_j = Y^T Q Z_j Z_j^T Q Y = Y^T R_j Y$ with its expected value. The REML equations can then be expressed as

$$E_{\tau}[s_j(\tau)] = s_j(\tau). \quad (3.3)$$

The LHS of equation 3.3 may be expressed as

$$E_{\tau}[s_j(\tau)] = tr(R_j(\tau) Var_{\tau}(Y)) = tr\left(R_j(\tau) \sum_{i=1}^r \sigma_i^2 Z_i Z_i^T\right) = \sum_{i=1}^r t_{ij}(\tau) \sigma_i^2$$

$$j = 0, 1, 2, \dots, r$$

where $t_{ij} = \text{tr}(Z_i Z_i^T Q Z_j Z_j^T Q)$. This leads to the REML estimates being the iterative solutions to the system of non-linear equations $(T(\tau))\tau = s(\tau)$, where T is $((r+1) \times (r+1))$ with elements t_{ij} and the vector of length $(r+1)$ containing the entire set of terms s_j is defined as $s(\hat{\tau})$. At convergence the REML estimate $\hat{\tau}$ satisfies

$$(T(\hat{\tau}))\hat{\tau} = s(\hat{\tau}). \quad (3.4)$$

For notational simplicity in the following we drop the argument $\hat{\tau}$ from T and from other matrices such as Q and P and from vectors s and $s_{(a)}$. Equation 3.4 may then be expressed as $\hat{T}\hat{\tau} = \hat{s}$ where $\hat{T}$ and $\hat{s}$ indicate evaluation at $\tau = \hat{\tau}$.

Subset Deletion

Following deletion of subset Y_a computations based on Y_b lead to $s_{j(a)}$ being sums of squares for the reduced data set. The REML estimator $\hat{\tau}_{(a)}$ now satisfies

$E_{\hat{\tau}_{(a)}}(s_{j(a)}(\hat{\tau}_{(a)})) = s_{j(a)}(\hat{\tau}_{(a)})$. This leads to the matrix equation

$$(T_{(a)}(\hat{\tau}_{(a)}))\hat{\tau}_{(a)} = s_{(a)}(\hat{\tau}_{(a)}) \quad (3.5)$$

Christensen, Pearson and Johnson (1992) use the approximation $s_{j(a)}(\hat{\tau}_{(a)}) \approx s_{j(a)}(\hat{\tau})$ to develop their one-step approximation $\hat{\sigma}_{(a)}$ to the vector $\hat{\tau}_{(a)}$.

An alternative approach also based on the approximation above and denoted by $\tilde{\sigma}_{(a)}$ is outlined below (Haslett and Dillane, 2004). A desirable feature of $\tilde{\sigma}_{(a)}$ is that it is simple to construct $\tilde{\sigma}_{(P)}$; no such simple construction for $\hat{\sigma}_{(P)}$ or $\tilde{\sigma}_{(P)}$ exists. However, the ‘delete = replace’ approach leads to some simplification of the approach proposed by CPJ.

Application to Christensen Pearson and Johnson’s (CPJ’s) ones step approximation

The CPJ approach is described as a ‘one-step approximation’ with

$E_{\hat{\tau}_{(a)}}(s_{j(a)}(\hat{\tau}_{(a)})) = E_{\hat{\sigma}_{(a)}}(s_{j(a)})$ where $\hat{\sigma}_{(a)}$ satisfies $E_{\hat{\sigma}_{(a)}}(s_{j(a)}(\hat{\tau})) = s_{j(a)}(\hat{\tau}) \cdot \hat{\sigma}_{(a)}$ is

$E_{\hat{\tau}_{(a)}}(s_{j(a)}(\hat{\tau}_{(a)})) = E_{\hat{\sigma}_{(a)}}(s_{j(a)})$ where $\hat{\sigma}_{(a)}$ satisfies $E_{\hat{\sigma}_{(a)}}(s_{j(a)}(\hat{\tau})) = s_{j(a)}(\hat{\tau}) \cdot \hat{\sigma}_{(a)}$ is

estimated from

$$(\hat{T}_{(a)})\hat{\sigma}_{(a)} = \hat{s}_{(a)} \quad (3.6)$$

whence

$$\hat{\sigma}_{(a)} = (\hat{T}_{(a)})^{-1} \hat{s}_{(a)} \quad (3.7)$$

CPJ provide numerical evidence that $\hat{\sigma}_{(a)}$ is a satisfactory approximation to $\hat{\tau}_{(a)}$. The

“delete = replace” approach outlined in appendix 1 leads to greater computational efficiency with

$$\begin{aligned} s_{j\bar{a}} &= Y^T QZ_j Z_j^T QY = Y^T W_a QZ_j Z_j^T QW_a Y \\ E_{\hat{\sigma}_{(a)}}(s_{j\bar{a}}) &= E_{\hat{\sigma}_{(a)}}(Y^T W_a QZ_j Z_j^T QW_a Y) = \text{tr}(W_a^T QZ_j Z_j^T QW_a \text{Var}_{\hat{\sigma}_{(a)}}(Y)) \\ &= \text{tr}(W_a^T QZ_j Z_j^T QW_a \sum_j Z_i Z_i^T \hat{\sigma}_{j(a)}) = \sum_i t_{ij(a)} \hat{\sigma}_{j(a)} \\ &= \hat{T}_{(a)} \hat{\sigma}_{(a)} \end{aligned} \quad (3.8)$$

The elements of the matrix $T_{(a)}$ are $t_{ij(a)} = \text{tr}(W_a^T QZ_j Z_j^T QW_a Z_i Z_i^T)$ and thus $T_{(a)}$ can be computed more easily than CPJ suggest. In particular, it is not necessary to work with any reduced dimension matrices. However, it is necessary to construct and invert $T_{(a)}$ separately for each subset Y_a .

Alternative approximation to CPJ (Haslett and Dillane, 2004)

The alternative approximation $\tilde{\sigma}_{(a)}$ proposed in Haslett and Dillane (2004) follows from the trace result (A1.2) outlined in appendix 1

$$E(s_{j\bar{a}}) = \text{tr}(QZ_j Z_j^T \text{Var}(Y)) - (QZ_j Z_j^T \bar{D}_a)$$

The proposed approximation is

$$E_{\tilde{\sigma}_{(a)}}(s_{j\bar{a}}) \approx \text{tr}(QZ_j Z_j^T Q\text{Var}_{\tilde{\sigma}_{(a)}}(Y)) - \text{tr}(QZ_j Z_j^T Q\bar{D}_a).$$

$$= \sum_i \text{tr}(QZ_j Z_j^T QZ_i Z_i^T \tilde{\sigma}_{i(a)}) - u_{ja}(\hat{\tau}) = \sum_i \hat{T}_{ij} \tilde{\sigma}_{i(a)} - u_{ja}(\hat{\tau}) \quad (3.9)$$

where $\text{tr}(QZ_j Z_j^T Q\bar{D}_a) = u_{ja}$. This leads to

$$\hat{T} \tilde{\sigma}_{(a)} = \hat{s}_{(a)} + u_{(a)} \equiv \tilde{s}_{(a)} \quad (3.10)$$

where $u_{(a)}$ has elements u_{ja} . The matrix T is available from the full fit and is thus already available. The estimate $\tilde{\sigma}_{(a)}$ may be contrasted with $\hat{\tau}$ via

$$\hat{T}(\hat{\tau} - \tilde{\sigma}_{(a)}) = (\hat{s} - \hat{s}_{(a)}) - \hat{u}_a \quad (3.11)$$

This formulation is particularly attractive as it then follows that

$$\tilde{\sigma}_{(a)} = \hat{T}^{-1} \tilde{s}_{(a)} \quad (3.12)$$

which permits computation as

$$\tilde{\sigma}_{(p)} = \hat{T}^{-1} \tilde{s}_{(p)} \quad (3.13)$$

where $\tilde{\sigma}_{(p)}$ denotes a matrix each of whose columns corresponds to one subset (a_i) . It is only necessary to compute $\tilde{s}_{(a)}$ for each subset Y_a under investigation. The advantage of this approach is that following REML convergence the only additional computations are those involved in constructing $\tilde{s}_{(p)}$.

3.2 Evaluation of Proposals

The various approaches are considered by their application to four datasets and the results are presented below. In sections 3.2.1 and 3.2.2 the accuracy of the approaches are contrasted. The dataset considered in section 3.2.1 was used by CPJ to evaluate their proposals. It is also considered in Haslett and Dillane (2004). However, they didn't consider the ANOVA based approach. The example in section 3.2.2 includes covariates. The application of the methods to the longitudinal dataset in section 3.2.3 demonstrates that the approximation may be applied to models including an autocorrelation parameter.

The objective of the analysis in section 3.2.4 is to explore the insight offered by the “delete = replace” approach in such analyses.

These analyses indicate that the approximation $\tilde{\sigma}_{(a)}$ is almost as good an approximation to $\hat{\tau}_{(a)}$ as is $\hat{\sigma}_{(a)}$. The approximation $\check{\sigma}_{(a)}$ is good if following deletion the data retains approximate balance. However, the more unbalanced the data is following deletion the poorer the approximation. In section 3.3 the conditions and situations where $\tilde{\sigma}_{(a)}$ differs from $\hat{\sigma}_{(a)}$ are explored.

3.2.1 Evaluation of Proposals-Nicotine Data

Wagner and Thaggard (1979) conducted an experiment to determine the amount of nicotine in samples extracted from Cambridge filter pads. Determinations are by gas liquid chromatography. Ten samples were analysed in each of 14 laboratories. The data are provided in appendix 3. The following mixed model is fitted.

$$Y_{ij} = \mu + \alpha_i + M_j + \varepsilon_{ij}$$

$i = 1, \dots, 10; j = 1, \dots, 14$ where Y_{ij} is the amount of nicotine in mg, α_i is a fixed effect for the i^{th} sample, M_j is the random effect for the j^{th} laboratory, and ε_{ij} is the error.

There are 138 observations as two data items are missing; the design is thus unbalanced.

In Fig. 3.1 comparative results are presented for computations on the changes, in the estimates of both variance components. These are induced as a consequence (in Figs. 3.1(a)-3.1(f)) of deleting individual observations, and (in Figs. 3.1(g)-3.1(l)) of deleting sets of observations associated with the individual laboratories. These comparisons are between a full REML refit on the one hand and on the other the approximations (i) the one step approximation $\hat{\sigma}_{(a)}$ using expressions for $\hat{T}_{(a)}$ and $\hat{s}_{(a)}$ (ii) the approximation $\tilde{\sigma}_{(a)}$ and (iii) the approximation $\check{\sigma}_{(a)}$ based on Henderson’s Method III approach. These are presented in terms of the Relative Variance Change and the line $Y = X$ is plotted.

Fig. 3.1 (a)-(f) Observation Deletion Plots for Nicotine study

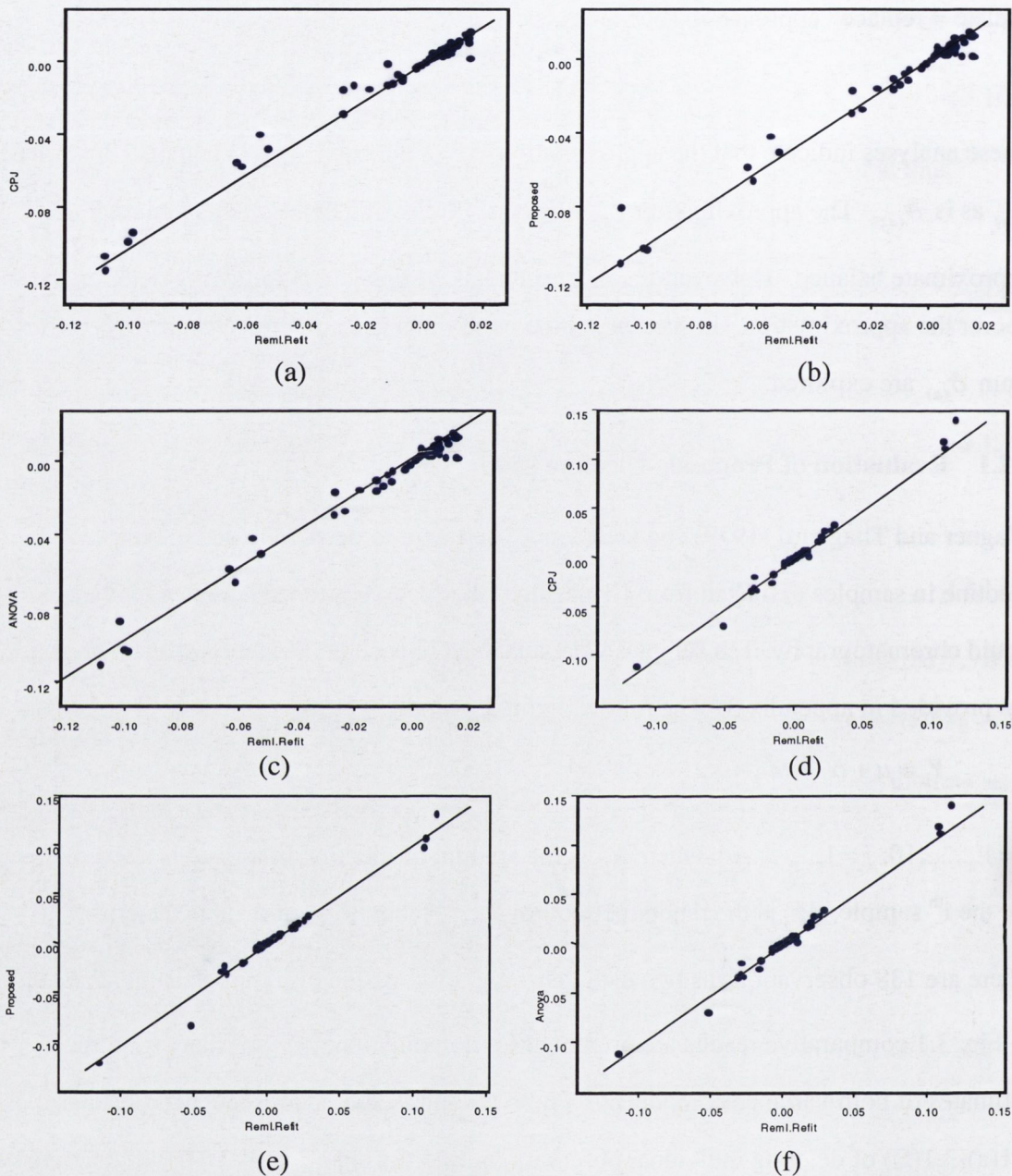


Fig. 3.1 Case deletion plots for the nicotine study: (a), (b) and (c) contrast in terms of relative variance change the approximations $\hat{\sigma}_{0(i)}$ (CPJ), $\tilde{\sigma}_{0(i)}$ (Proposed) and $\check{\sigma}_{0(i)}$ (ANOVA based on Henderson's Method III) with the full REML refit $\hat{\tau}_{0(i)}$ arising from the deletion in turn of each of the 138 observations; similarly (d), (e) and (f) contrast $\hat{\sigma}_{1(i)}$, $\tilde{\sigma}_{1(i)}$ and $\check{\sigma}_{1(i)}$ with $\hat{\tau}_{1(i)}$. Lines show equality of different measures

Fig. 3.1 (g)-(l) Subset Deletion Plots for Nicotine study

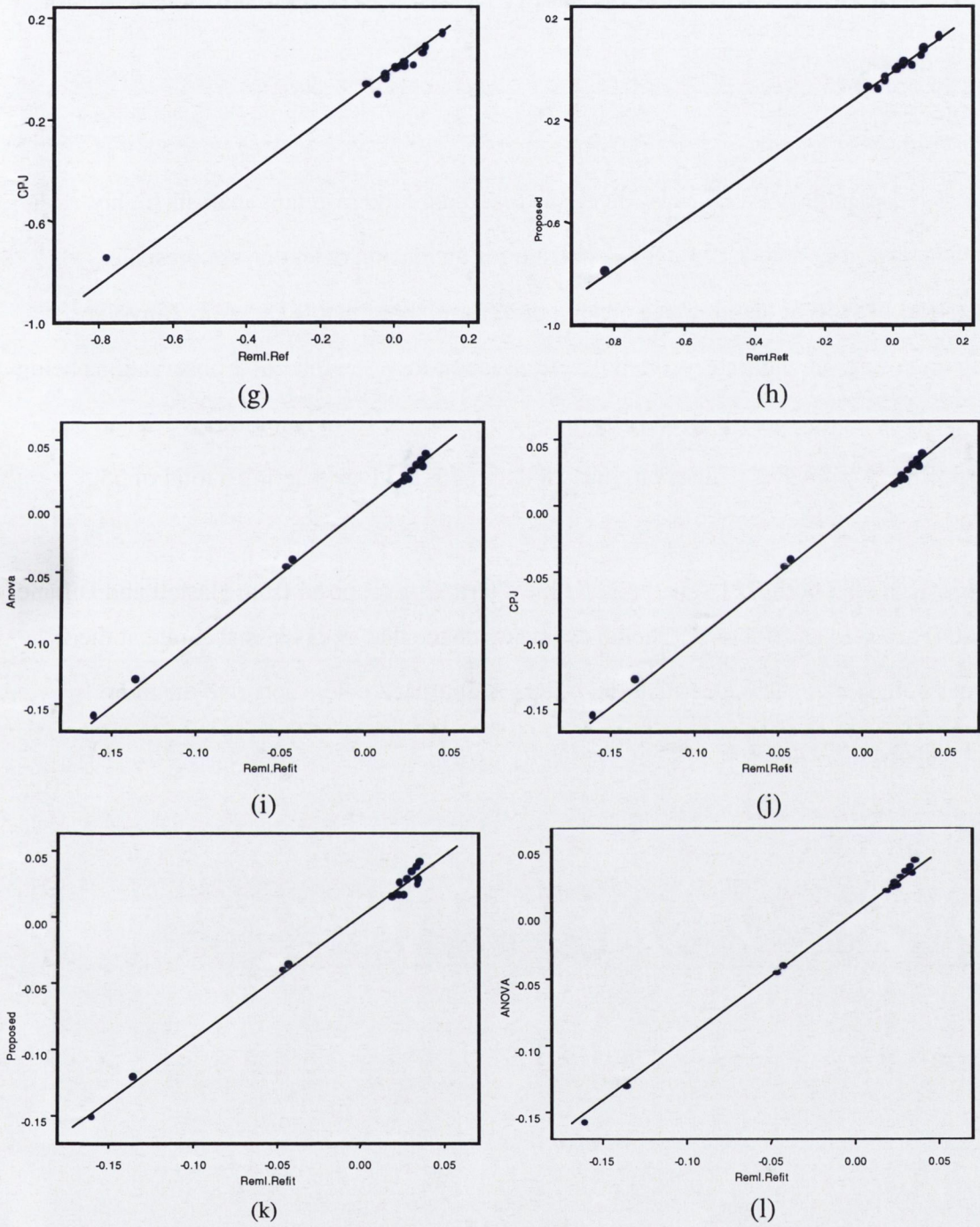


Fig 3.1 Deletion plots for the nicotine study: (g), (h) and (i) contrast in terms of relative variance change the approximations $\hat{\sigma}_{0(a)}$ (CPJ), $\tilde{\sigma}_{0(a)}$ (Proposed) and $\check{\sigma}_{o(a)}$ (ANOVA based on Henderson's Method III) for the full REML refit $\hat{\tau}_{0(a)}$ arising from the deletion of all observations that are associated with each of the 14 laboratories; similarly (j), (k) and (l) contrast $\hat{\sigma}_{1(a)}$, $\tilde{\sigma}_{1(a)}$ and $\check{\sigma}_{1(a)}$ with $\hat{\tau}_{1(a)}$

It is clear from Fig. 3.1 that $\hat{\sigma}_{(a)}$, $\tilde{\sigma}_{(a)}$ and $\check{\sigma}_{(a)}$ are very good approximations to $\hat{\tau}_{(a)}$ for these deletions.

A more demanding test involves deletions that render the resultant analysis highly unbalanced. In Table 3.1, the RVC statistic for increasing deletions are presented. Deletion 1 involves deleting one observation from laboratory 1; Deletion 2 involves deleting one from laboratory 1 and two from laboratory 2, a total of 3 observations being deleted; Deletion 3 involves deleting one, two and three from laboratories 1, 2 and 3 respectively, a total of 6 observations. Deletion 10 involves deleting a total of 55.

From Table 3.1 both CPJ's method and the alternative proposal from Haslett and Dillane (2004) are very satisfactory. There is some evidence that $\tilde{\sigma}_{(a)}$ is less accurate at the extremes than $\hat{\sigma}_{(a)}$. It is clear that the ANOVA approach is less accurate the more unbalanced the data becomes.

that the reported crop hectares are positively correlated within given counties but uncorrelated from different counties.

Table 3.2 Survey and Satellite Data for Corn and Soybeans in 12 Iowa Counties

County	Sample	Corn	Soybeans	Corn	Soybeans	Corn	Soybeans
	(a)	(b)	(b)	(c)	(c)	(d)	(d)
Cerro Gordo	1	165.76	8.09	374	55	295.29	189.70
Hamilton	1	96.32	106.03	209	218	300.40	196.65
Worth	1	76.08	103.6	253	250	289.60	205.28
Humboldt	2	185.35	6.47	432	96	290.74	220.22
		116.43	63.82	367	178		
Franklin	3	162.08	43.50	361	137	318.21	188.06
		152.04	71.43	288	206		
		161.75	42.49	369	165		
Pocahontas	3	92.88	105.26	206	218	257.17	247.13
		149.94	76.49	316	221		
		64.75	174.34	145	338		
Winnebago	3	127.07	95.67	355	128	291.77	185.37
		133.55	76.57	295	147		
		77.70	93.48	223	204		
Wright	3	206.39	37.84	459	77	301.26	221.36
		108.33	131.12	290	217		
		118.17	124.44	307	258		
Webster	4	99.96	144.15	252	303	262.17	247.09
		140.43	103.60	293	221		
		98.95	88.59	206	222		
		131.04	115.58	302	274		
Hancock	5	114.12	99.15	313	190	314.28	198.66
		100.60	124.56	246	270		
		127.88	110.88	353	172		
		116.90	109.14	271	228		
		87.41	143.66	237	297		
Kossuth	5	93.48	91.05	221	167	298.65	204.61
		121.00	132.33	369	191		
		109.91	143.14	343	249		
		122.66	104.13	342	182		
		104.21	118.57	294	179		
Hardin	6	88.59	102.59	220	262	325.99	177.05
		165.35	69.28	355	160		
		104.00	99.15	261	221		
		88.63	143.66	187	345		
		153.70	94.49	350	190		

Batesse *et al* (1988) fit the following model

$$y_{ij} = \beta_0 + \beta_1x_{1ij} + \beta_2x_{2ij} + u_{ij}$$

where i is the subscript for county ($i = 1, 2, \dots, 12$); j is the subscript for a segment within a given county ($j = 1, 2, \dots, n_i$ where n_i is the number of sample segments in the i th county); y_{ij} is the number of hectares of corn (or soybeans) in the j th segment of the i th county; x_{1ij} and x_{2ij} are the number of pixels classified as corn and soybeans respectively in the j th segment of the i th county; and β_0, β_1 and β_2 are unknown parameters. The random error u_{ij} associated with the reported crop area y_{ij} is expressed as $u_{ij} = v_i + e_{ij}$, where v_i is the i th county effect and e_{ij} is the random effect associated with the j th sample segment within the i th county. The random errors v_i and e_{ij} are assumed to be iid Normal random variables with mean zero and variances σ_v^2 and σ_e^2 respectively.

The focus of the Banesse *et al* (1988) analysis is the prediction of the mean crop hectares per segment. The estimated parameters for corn are $\hat{\beta} = (48, .327, -.135)$ with $\hat{\sigma}_e^2 = 150$ and $\hat{\sigma}_v^2 = 138$. The estimated parameters for the soybeans are $\hat{\beta} = (-18, +.028, +.494)$ with $\hat{\sigma}_e^2 = 189$ and $\hat{\sigma}_v^2 = 252$.

The analysis involves comparing $\hat{\sigma}_{(a)}$, $\tilde{\sigma}_{(a)}$ and the complete REML refit $\hat{\tau}_{(a)}$. The various measures are contrasted in Fig. 3.2 below. Figs. 3.2a(a)-(f) and 3.2a(g)-(l) are observation deletion plots for the corn and soybean data respectively. It is clear from the graphs that $\tilde{\sigma}_{(a)}$ is a very good approximation for $\hat{\tau}_{(a)}$. The greatest deviation between both approximate methods arises for the value of segment 1 of Humboldt county though this is still very small. Figs. 3.2b(a)-(f) and 3.2(g)-(l) are the corresponding plots for each county. The greatest deviation at county level arises for Kossuth. Kossuth county is most influential on the county variance component. However again this deviation is small in magnitude. The conditions and situations where deviations arise between $\tilde{\sigma}_{(a)}$ and $\hat{\sigma}_{(a)}$ are explored in section 3.3.

Fig. 3.2a (a)-(f) Observation Deletion Plots for Corn data.

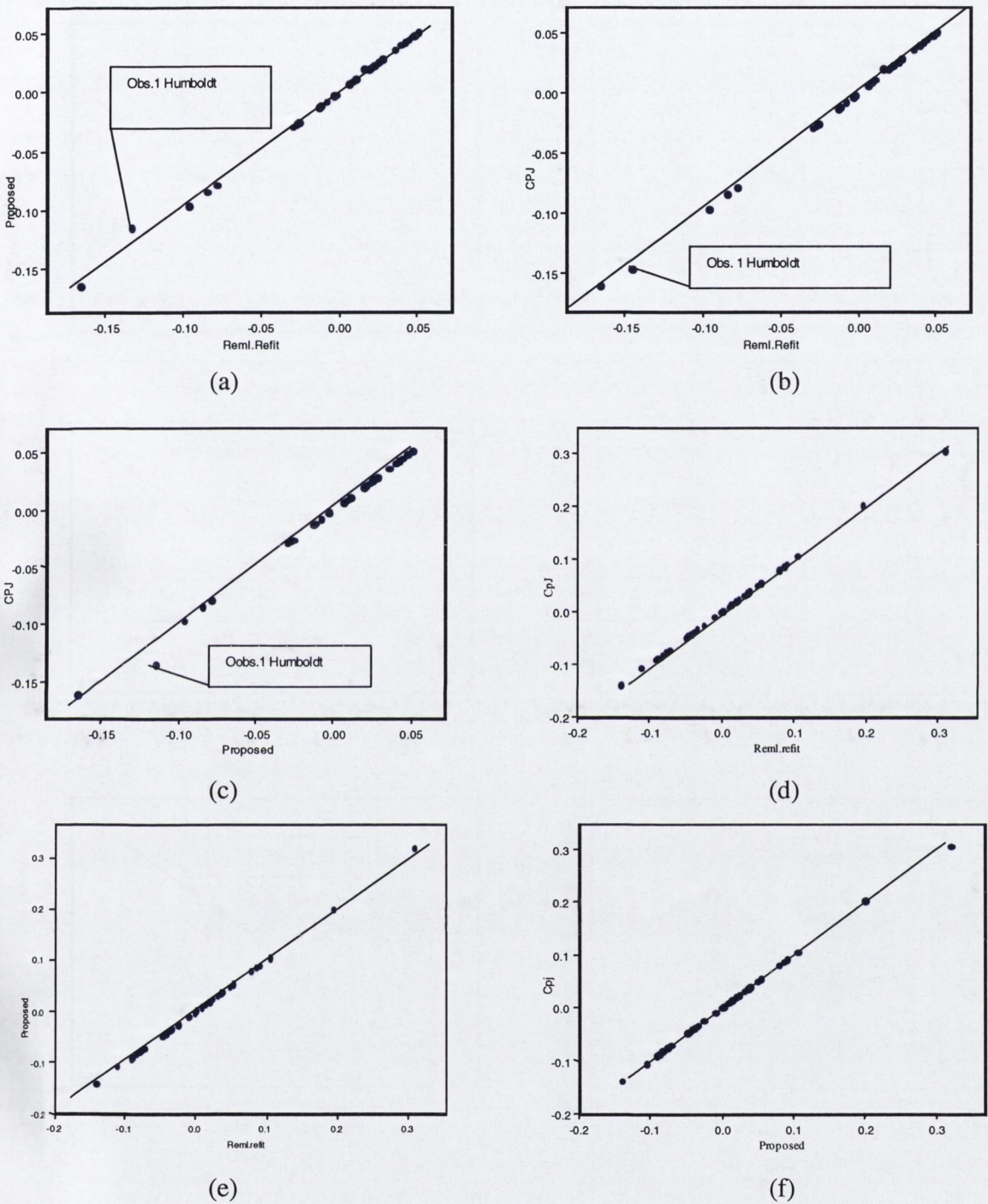
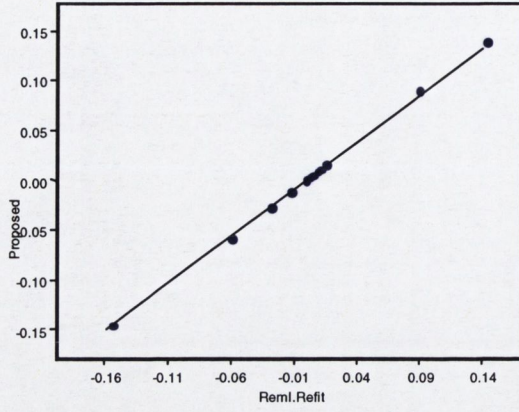
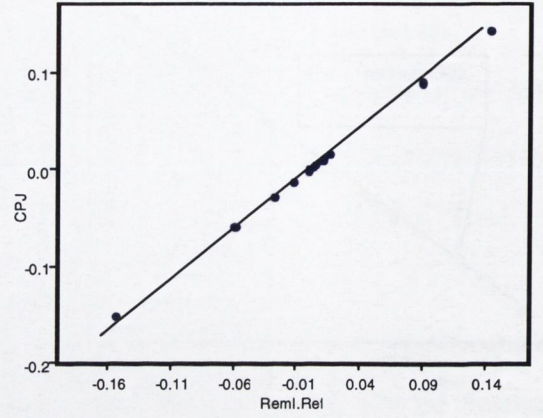


Fig. 3.2(a) Case deletion plots for the corn data. (a) (b) and (c) contrast in terms of relative variance change the approximations $\hat{\sigma}_{e(i)}$ (CPJ), $\tilde{\sigma}_{e(i)}$ (Proposed) for the full REML refit $\hat{\tau}_{e(i)}$ arising from the deletion in turn of each of the 36 observations; similarly (d) (e) and (f) contrast $\hat{\sigma}_{v(i)}$, $\tilde{\sigma}_{v(i)}$ with $\hat{\tau}_{v(i)}$ for each observation.

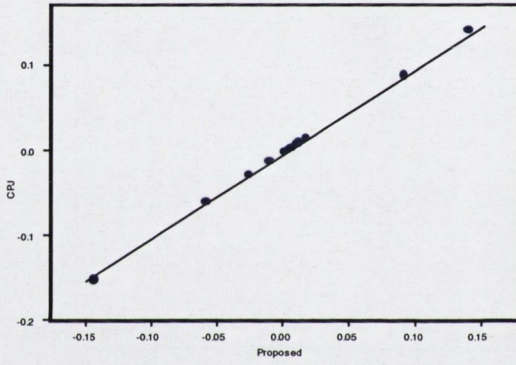
Fig. 3.2a (g)-(l) Subset Deletion (Counties) Plots for corn data



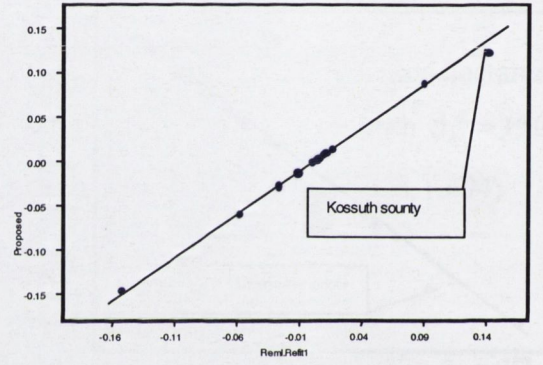
(g)



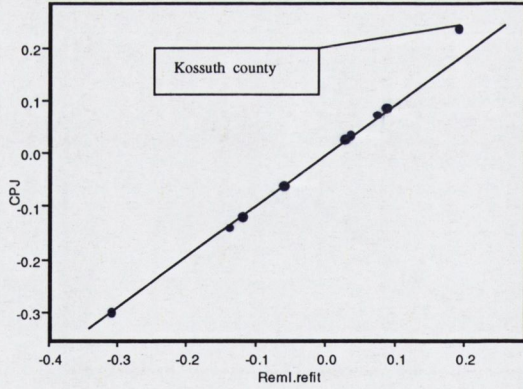
(h)



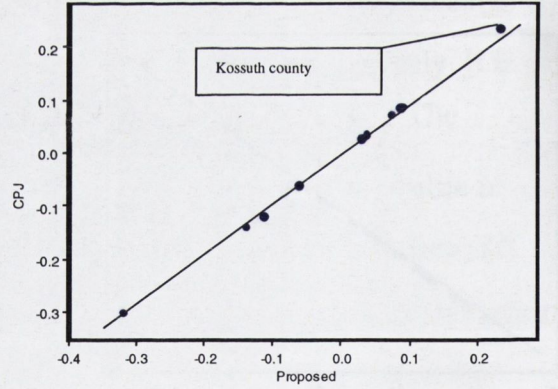
(i)



(j)



(k)



(l)

Fig. 3.2(a) Deletion plots for the corn data. (g) (h) and (i) contrast in terms of relative variance change the approximations $\hat{\sigma}_{e(a)}(\text{CPJ})$, $\tilde{\sigma}_{e(a)}(\text{Proposed})$ for the full REML refit $\hat{\tau}_{e(a)}$ arising from the deletion in turn the entire set of all observations associated with each of the 12 counties; similarly (j) (k) and (l) contrast $\hat{\sigma}_{v(a)}$, $\tilde{\sigma}_{v(a)}$ with $\hat{\tau}_{v(a)}$ for each county.

Fig.3.2b (a)-(f) Observation Deletion Plots for Soybean Data

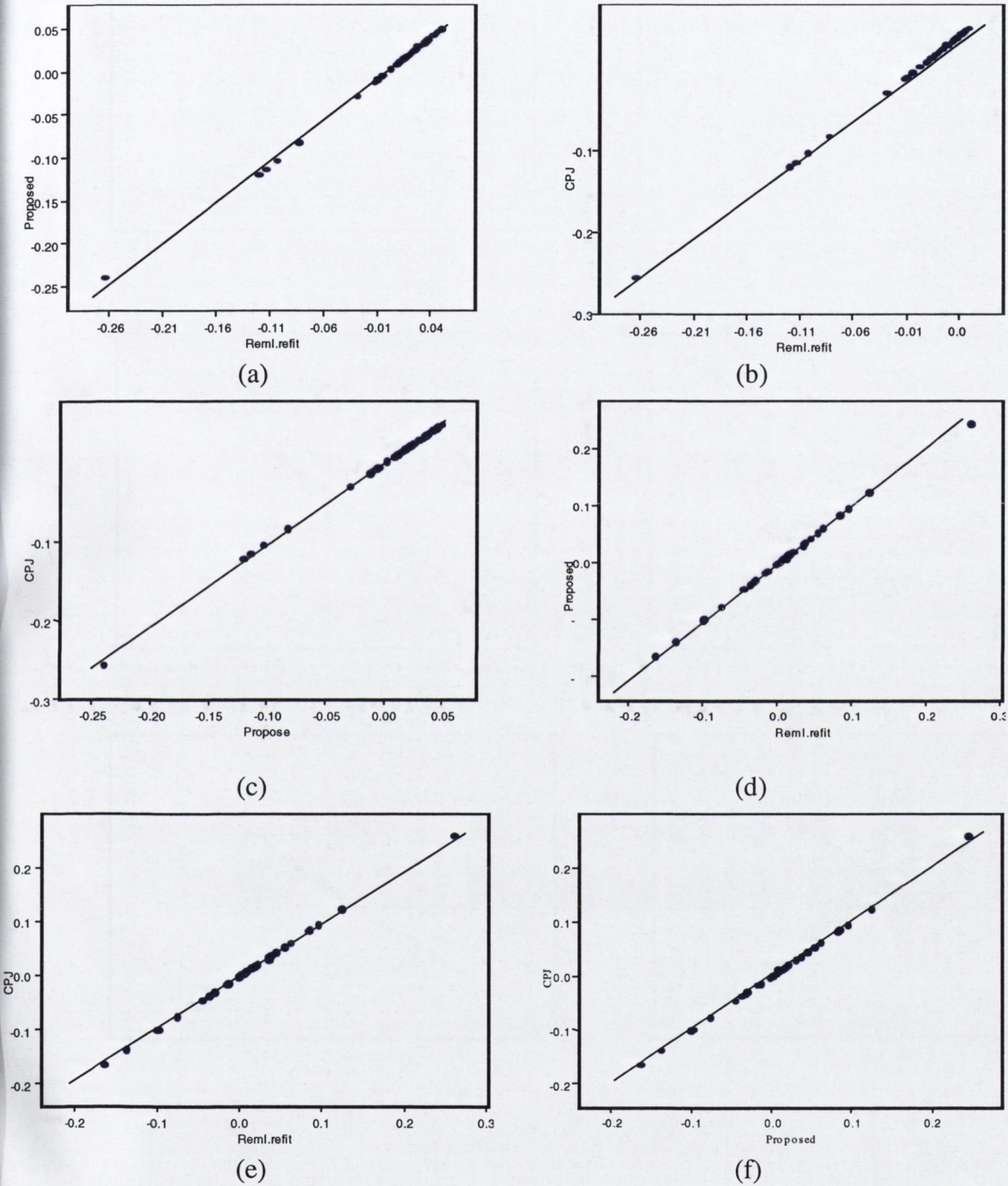


Fig. 3.2(b) Observation deletion plots for the soybeans data. (a) (b) and (c) contrast in terms of relative variance change the approximations $\hat{\sigma}_{e(i)}$ (CPJ), $\tilde{\sigma}_{e(i)}$ (Proposed) for the full REML refit $\hat{\tau}_{e(i)}$ arising from the deletion in turn of each of the 36 observations; similarly (e) (d) and (f) contrast $\hat{\sigma}_{v(i)}$, $\tilde{\sigma}_{v(i)}$ with $\hat{\tau}_{v(i)}$ for each observation.

Fig. 3.2b (g)-(l) Subset Deletion (Counties) Plots for Soybean data

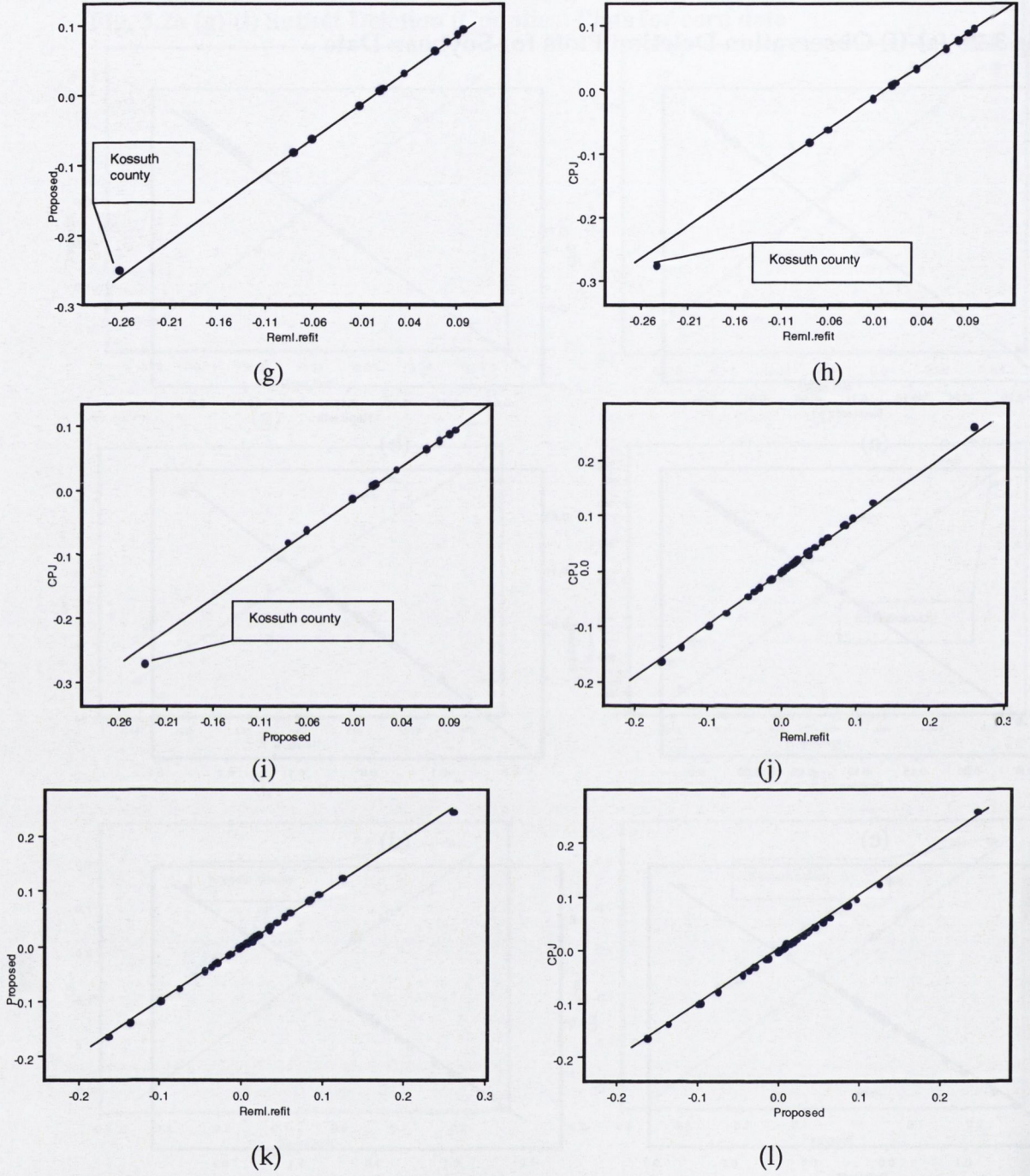
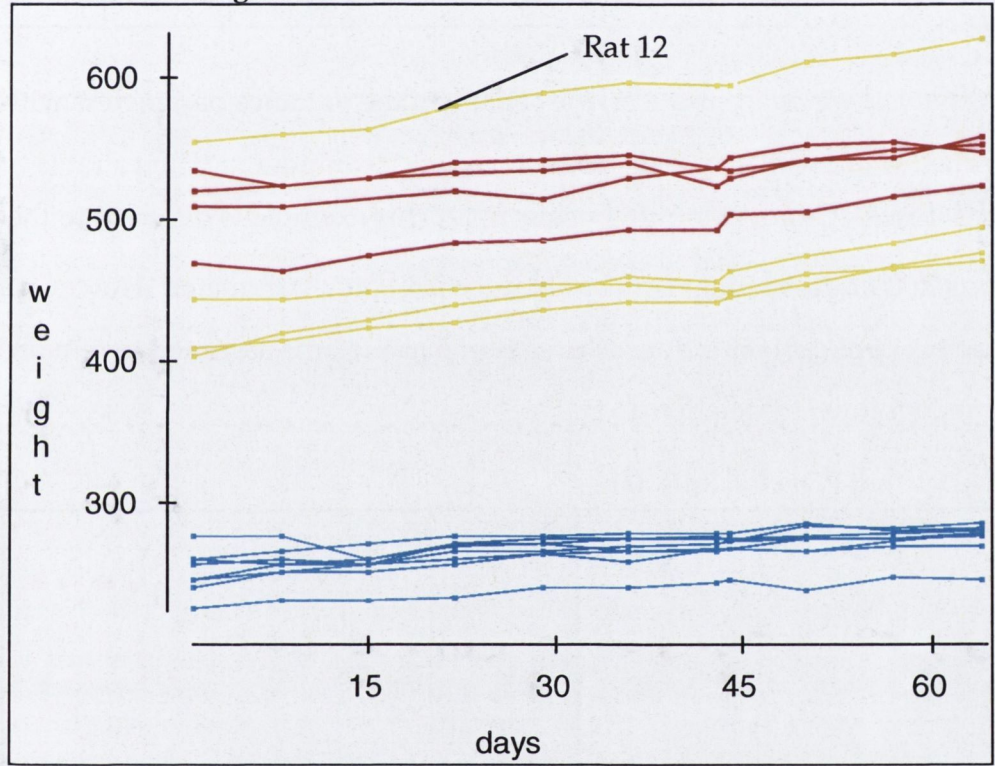


Fig. 3.2(b) Deletion plots for the soybeans data. (g) (h) and (i) contrast in terms of relative variance change the approximations $\hat{\sigma}_{e(a)}$ (CPIJ), $\tilde{\sigma}_{e(a)}$ (Proposed) for the full REML refit $\hat{\tau}_{e(a)}$ arising from the deletion in turn the entire set of all observations associated with each of the 12 counties; similarly (j) (k) and (l) contrast $\hat{\sigma}_{v(a)}$, $\tilde{\sigma}_{v(a)}$ with $\hat{\tau}_{v(a)}$ for each county.

3.2.3 Analysis of rats data (Crowder and Hand, 1990)

These data refer to an experiment in which the weights of three groups of rats were measured over a nine week period on days $t=1,8,15,22,29,36,43,44,50,57$ and 64 . The group sizes were $8,4$ and 4 respectively. The data is plotted in Fig.3.3 below.

Fig. 3.3 Plot of rat weights over time



The model fitted is as in Crowder and Hand (1996) and is given below.

$$Y = \alpha_1 + \beta t$$
$$Y = \alpha_2 + \beta t$$
$$Y = \alpha_3 + \beta t$$

For group 1

For group 2

For group 3

The model was fitted using REML estimation and the V matrix was block diagonal with each block of the form $\sigma_1^2 J + E$ where J is an 11×11 matrix whose elements are all 1, σ_1^2 reflects the covariance within each rat's measurements by its deviation from the

group intercept and E is an 11×11 matrix and reflects the correlation between successive measurements on the same rat; Crowder and Hand (1990) selected a Markovian structure here with the (j, k) element of E being $\sigma^2 \rho^{|j-k|/7}$. The results of the model fit using REML estimation were as follows

Mean Parameters	α_1	α_2	α_3	β
	243.25	465.493	507.693	0.602
Variance Parameters	σ	ρ	σ_1	
	10.24	0.96	69.72	

In this analysis we concentrate on the change in the variance parameters with Fig. 3.4 below showing the RVC for each variance component when a rat is deleted. The RVCs are computed using $\tilde{\sigma}_{(a)}$. This example is different to the previous in that the model includes an autocorrelation parameter. However, the approach is applicable if the autocorrelation parameter is assumed to remain unchanged following deletion.

Fig. 3.4 Deletion Plots for Rats data

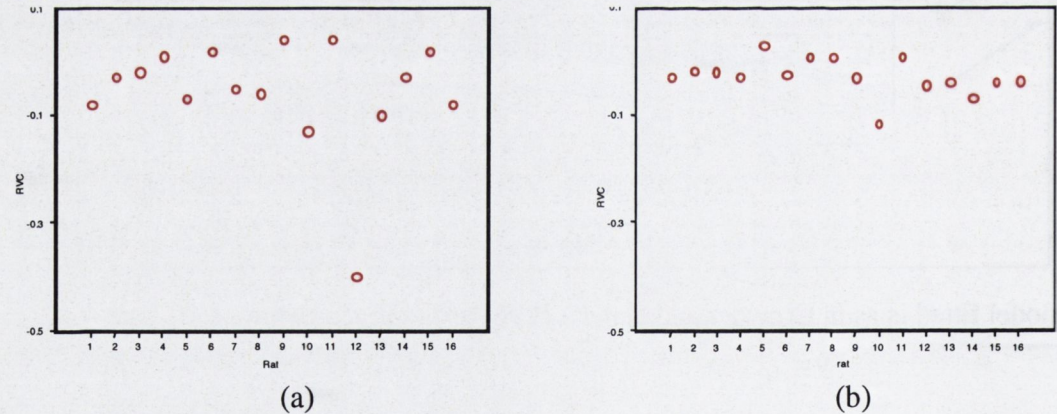


Fig. 3.4 (a) and (b); Plots of RVC for $\hat{\sigma}_1^2$ and $\hat{\sigma}^2$ respectively

From Fig. 3.4 (a) it is clear that rat 12 is highly influential in the estimation of σ_1^2 . This is not surprising as from Fig. 3.3 rat 12 is clearly far away from the group 2 behaviour. From Fig. 3.4 (b) rat 10 appears most influential on σ^2 . Rat 10 while somewhat distant

from the others in group 3 shows a large jump from day 43 to 44 and is influential on both variance components. Such large jumps between successive measurements are unlikely given the very high fitted correlation in the model. This example demonstrates that the Haslett and Dillane (2004) approach is applicable in such situations if it is assumed that the autocorrelation parameter is fixed following deletion.

3.2.4 Analytical Chemistry Data

The final example is used to demonstrate how the ‘delete = replace’ approach may be useful in interpreting a variance component influence analysis. The dataset below in Table 3.3 is a subset of data arising in a co-operative trial in analytical chemistry reported by Analytical Methods Committee (AMC) (1987); this subset is also discussed in Venables and Ripley (1994, Table 6.1). Estimates based on the proposed approximation (Dillane and Haslett, 2004) of equation 3.13 of the implications of case and subset deletion for each of the three variance components are presented in Fig. 3.5

Table 3.3 Analytical Chemistry Data-Amylase concentration (g/kg) from 6 laboratories and 3 batches

Batch	Laboratory					
	1	2	3	4	5	6
1	0.29	0.40	0.40	0.90 ^{19,10}	0.44	0.38
	0.33	0.40	0.35	1.30 ^{20,10}	0.44	0.39
2	0.33	0.43	0.38	0.90	0.45	0.40
	0.32	0.36	0.32	1.10	0.45	0.46
3	0.34	0.42	0.38	0.90	0.42	0.72 ^{35,18}
3	0.31	0.40	0.33	0.90	0.46	0.79 ^{36,18}

The superscripts refer to the labels (Observation, Batch) so for example the value 1.30 corresponds to Lab 4, Batch 10, Observation 20.

Seven specimens were sent to 6 laboratories, each 3 times a month apart, for duplicate analysis. The response is the concentration of (unnamed) analyte in g/kg. The analysis they report was on specimen 1. The purpose of the study was to assess components of variation in co-operative trials. For this purpose laboratories and batches are regarded as random, so a model for the response for laboratory i , batch j and duplicate k is

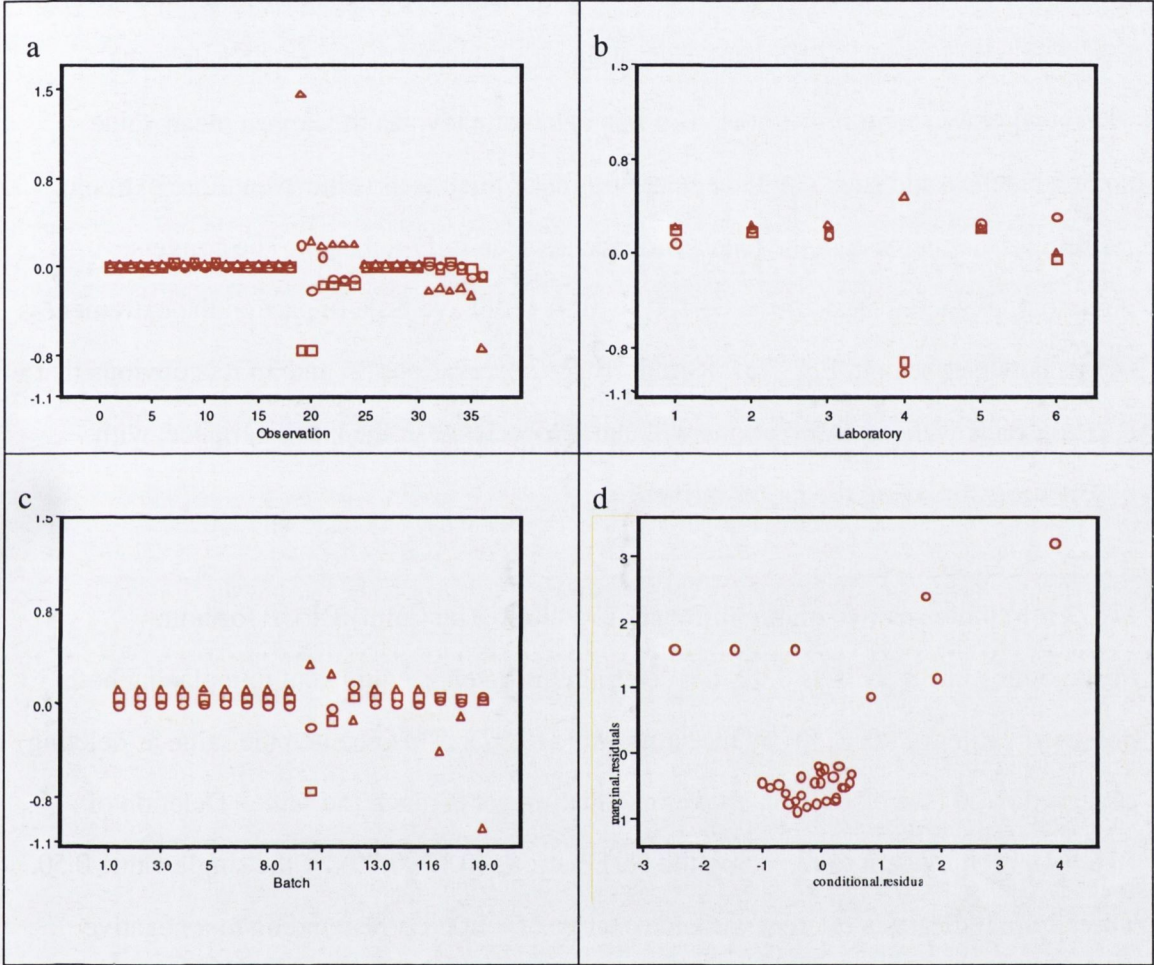
$$y_{ijk} = \mu + \xi_i + \beta_{ij} + \varepsilon_{ijk}; i = 1..6; j = 1, 2, 3; k = 1, 2 \quad (3.14)$$

where ξ , β and ε are independent random variables with zero means and variances σ_L^2 , σ_B^2 and σ_e^2 . The design is balanced and components of variance are estimated (whether by ANOVA or by REML) as $\hat{\sigma}_L^2 = 0.0602$, $\hat{\sigma}_B^2 = 0.0054$ and $\hat{\sigma}_e^2 = 0.0063$.

Figure 3.5(d) presents a scatterplot of standardised conditional and marginal residuals; see Haslett and Hayes (1998). It is clear that Laboratory 4 and Batch 18 are outlying. It is not clear how they influence the estimates of the three components of variance. Fig. 3.5(a) indicates that observations (19, 20) (from Batch 10, Lab 4) and (35, 36) (from Batch 18, Lab 6) are very influential on some or all of the variance components. Observations 19 and 20 both cause a large decrease in $\hat{\sigma}_e^2$ on deletion; but deleting observation 19 causes a very large increase (more than 100%) in $\hat{\sigma}_B^2$. Observation 36 and to a lesser extent observation 35 cause a substantial decrease in $\hat{\sigma}_B^2$ upon deletion.

Fig. 3.5(b) provides information on the influence of batches. It is clear that both Batch 10 and Batch 18 are highly influential. Deleting Batch 10 greatly reduces $\hat{\sigma}_e^2$. Deletion of Batch 18 reduces $\hat{\sigma}_B^2$ even more dramatically, with an RVC of -1.02 , indicating a negative (!) variance estimate; see discussion below. Fig. 3.5 (c) shows that Lab 4 is very influential on both $\hat{\sigma}_L^2$ and $\hat{\sigma}_e^2$.

Fig. 3.5 Deletion Plots for the AMC data.



Figs. 3.5 (a),-(c); Plots of RVC for each observation, batch and laboratory respectively. $\circ$ laboratory variance; Δ batch variance; $\square$ error variance. Fig. 3.5 (d) Scatterplot of marginal and conditional residuals.

The use of the “delete = replace” approach is helpful in understanding the impact of these observations on the various variance component estimates. To consider Fig. 3.5(a), note that the conditional expected values given the others (BLUPs) for these observations are

$\tilde{y}_{(19)} = 1.11$, $\tilde{y}_{(20)} = 0.91$, $\tilde{y}_{(35)} = 0.46$ and $\tilde{y}_{(36)} = 0.47$. Deleting case 19 (value 0.9) is (to a good approximation; as shown in section 3.1) the same as replacing its value by 1.11. This value (1.11) has been computed from the rest, but is naturally much influenced by its neighbour's high value of 1.3. Such a replacement will reduce the error variance, which is primarily concerned, in this example, with the inter-batch variation. This is what we see in Fig. 3.5(a). But this batch already has the largest mean value among batches, and such a replacement will make this mean value even more extreme, resulting in an increase in the batch variance, as seen in Fig. 3.5(a). The converse argument, replacing case 20 (value 1.3) by 0.91 will have little impact on the extremeness of this batch, as seen in Fig. 3.5a. Similarly for observations 35 and 36 it is obvious that deleting each of these observations will cause a decrease in the batch variance, with observation 36 having the larger influence.

Fig. 3.5(b) indicates two highly influential batches. The (joint) BLUP for both observations in Batch 10 is 0.92; i.e. deleting this batch is equivalent to replacing both observed values (0.90, 1.30) by the same value (0.92). This is much the same as deleting observation 20 (see discussion above) and the impact is much the same. Deletion of Batch 18 is equivalent to replacing the observations (0.72, 0.79) by the single value 0.50. The Figure indicates a relative variance change of -1.02 corresponding to a negative estimate $\hat{\sigma}_B^2$! This is entirely consistent with the REML methodology, which returns a 0, after a full refit, as it does also after a replacement of the pair of observations by the BLUP for that pair given the remaining data, both equal to 0.50. The application of ANOVA methods to the data, using 0.50 for both values, returns a negative $\hat{\sigma}_B^2$. This is a well documented shortcoming of the methodology. The value of σ_B^2 which maximises the likelihood (for the reduced data set, Batch 18 having been removed) is on the edge of the feasible region for the proposed model. Searle *et al* (1992; p 130) discuss a number of options facing the analyst in such circumstances. One is the REML strategy of

returning $\sigma_B^2=0$ and it is not clear how to interpret such a result, except to say that the model is inappropriate.

Fig. 3.5(c) indicates that deleting Lab 4 (which is approximately equivalent to replacing all six data values by the 0.41, the joint BLUP for all observations in this lab) reduces both $\hat{\sigma}_e^2$ and $\hat{\sigma}_L^2$. The latter is reduced since the value 0.41 is much closer to the other laboratory means. (For this design the 0.41 is easy to interpret, being the mean of the other five laboratories). The former is reduced as the variability within Lab 4 is greater than for any other laboratory. The fact that $\hat{\sigma}_B^2$ is not much impacted by this deletion/replacement indicates that the extra variability within Lab 4 is not attributable to inter-batch variation within Lab 4, as can be seen by inspection.

In summary, Fig. 3.5 has provided us with much insight into the sources of the evidence for the reported values $\hat{\sigma}_L^2 = 0.0602$, $\hat{\sigma}_B^2 = 0.0054$ and $\hat{\sigma}_e^2 = 0.0063$. It seems that laboratory 4 overall contributes too much for comfort to $\hat{\sigma}_e^2$ and $\hat{\sigma}_L^2$; similarly Batch 18 dominates $\hat{\sigma}_B^2$. This broadly concurs with the conclusions of Venables and Ripley (1997, p299) although they observe “a simple residual analysis will not expose this”. We have offered such an analysis.

The Analytical Methods Committee (1987) provides further details on their residual methodology and it is useful to contrast our approach with that recommended there, which involves computation of ‘residuals’ (being for the batches $\hat{\gamma}_{B,k}$, $k = 1, 2, 3$; for the laboratories $\hat{\gamma}_{L,h}$, $h = 1, \dots, 6$ and for the observation errors $\hat{\epsilon}_m$, $m = 1, \dots, 36$). For a balanced design, such as here, these are easy to compute by the subtraction of the appropriate means, and are closely related to the conditional residuals. For example, the ‘residual’ for Lab 1 is computed as $\hat{\gamma}_{L1} = \bar{y}_{L1} - (\bar{y}_{L1} + \bar{y}_{L2} + \dots + \bar{y}_{L6})/6$ and similarly $\hat{\gamma}_{Lh}$ $h = 2, \dots, 6$. (Observe that the BLUP for Lab 1 is $(\bar{y}_{L2} + \dots + \bar{y}_{L6})/5$; see paragraph above).

A large laboratory 'residual' suggests that $\hat{\sigma}_L^2$ will be much influenced by the corresponding laboratory. The laboratory 'residuals' are in fact (-0.188, -0.106, -0.1481, 0.492, -0.064, 0.015) and it is clear that Laboratory 4 has a large 'residual' and thus a large inflationary influence on $\hat{\sigma}_L^2$. No doubt this procedure will often work, although it is clear that with small samples (here $n = 6$ laboratories) the position is more complicated.

3.3 Comparison of $\hat{\sigma}_{(a)}$ with $\tilde{\sigma}_{(a)}$

It is clear from our analyses that in many situations both $\hat{\sigma}_{(a)}$ and $\tilde{\sigma}_{(a)}$ provide very similar results. The objective of this section is to explore the situations and conditions

where differences are likely to occur between the approximations. From section 3.1

$$E(s_{j(a)}) = E(Y_b^T A_{j(a)} Y_b^T) = \text{tr}(A_{j(a)} V_{(a)}) \text{ with } A_j = QZ_j Z_j^T Q$$

The CPJ approximation $\hat{\sigma}_{(a)}$ satisfies the equation

$$s_{j(a)}(\hat{\tau}) = \text{tr}(A_{j(a)}(\hat{\tau}) V_{(a)}(\hat{\sigma}_{(a)}))$$

Haslett (1999) shows that $E(s_{j\bar{a}}) = E(s_{j(a)})$ and from appendix 1

$$E(S_{j(a)}) = E(S_{j\bar{a}}) = E(Y_a^T A_j Y_a) = \text{tr}(A_j V - 2A_j V Q \bar{D}_a + A_j \bar{D}_a) \quad (3.15)$$

The Haslett and Dillane (2004) approximation $\tilde{\sigma}_{(a)}$ satisfies the equation

$$S_{j(a)}(\hat{\tau}) = \text{tr}(A_j(\hat{\tau}) V(\tilde{\sigma}_{(a)}) - A_j(\hat{\tau}) \bar{D}_a(\hat{\tau})) \quad (3.16)$$

It is clear from equation 3.15 that $\hat{\sigma}_{(a)} = \tilde{\sigma}_{(a)}$ if $VQ\bar{D}_a = \bar{D}_a$. The covariance matrix of the leave- a -out deletion prediction residual is $\bar{D}_a$ while $\text{Cov}(\hat{e}_a, \tilde{e}_{(a)}) = VQ\bar{D}_a$ as $\tilde{e}_{(p)} = D_{(p)} V^{-1} \hat{e}$. Hence greater differences will arise between $\hat{\sigma}_{(a)}$ and $\tilde{\sigma}_{(a)}$ the larger the difference between $\text{Cov}(\tilde{e}_{(a)}, \tilde{e}_{(a)})$ and $\text{Cov}(\hat{e}_a, \tilde{e}_{(a)})$. This will occur when observations or subsets of observations are outlying or unusual in the X and V spaces. Thus high

leverage subsets may be expected to lead to the biggest discrepancies between $\hat{\sigma}_{(a)}$ and $\tilde{\sigma}_{(a)}$.

To investigate further this issue a highly unbalanced dataset was created by modifying the rats dataset in Hand and Crowder (1996, pg 171) and considered in section 3.2.3. The modified dataset consists of 45 observations on 9 rats. These are rats 1,2,3,4,10,11,12,13 and 14. The data comprises of one observation on rat 1 at time point 1, two observations at time points 1 and 2 on rat 2 with nine observations on rat 14 over the first 9 time points. There are three groups with rats 1,2,3 and 4 in group 1, rats 10,11 and 12 in group 2 and rats 13 and 14 in group 3. The covariance matrix is block diagonal of the form $\eta J + E$ where J is an 11 x 11 matrix whose elements are all 1, η reflects the covariance within each rat's measurements by its deviation from the group intercept and E is a diagonal matrix of the form $\sigma_0^2 I$ This is the compound symmetry case which is very common in some applications.

The first phase of the analysis involved singleton deletion to consider the conditions when differences arise between $Cov(\tilde{e}_{(a)}, \tilde{e}_{(a)})$ and $Cov(\hat{e}_a, \tilde{e}_{(a)})$.

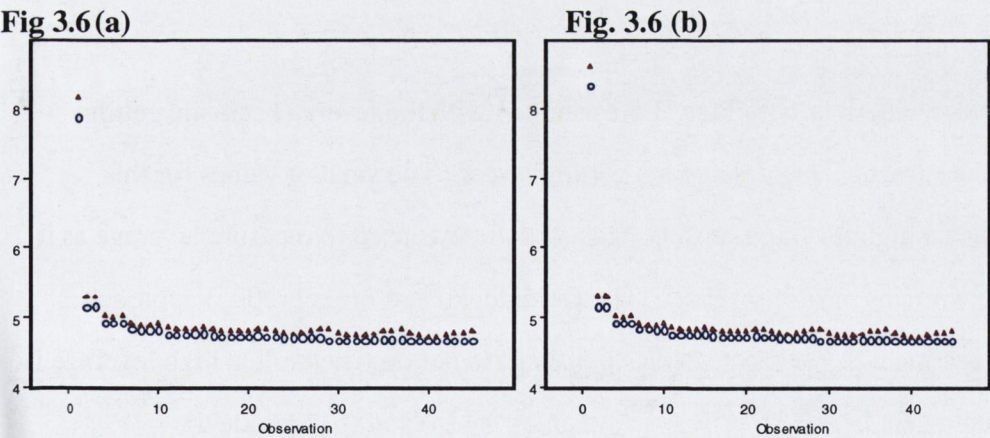
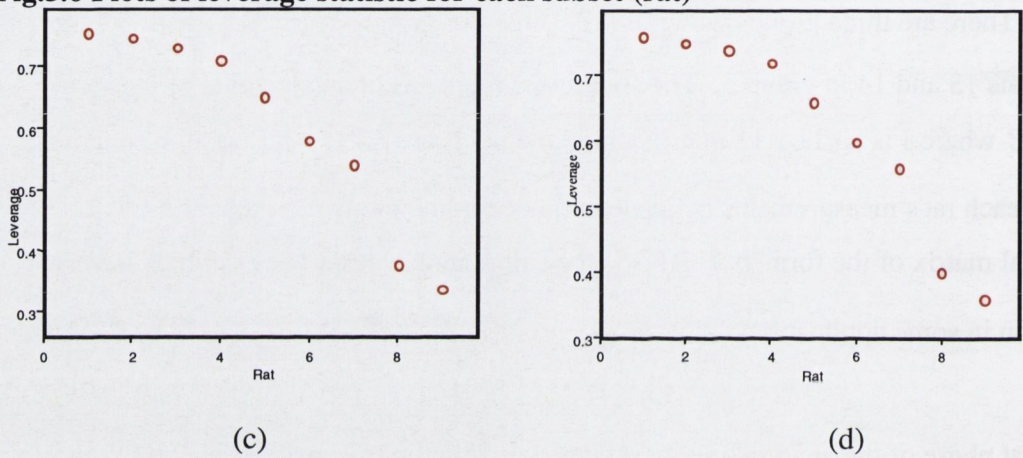


Fig. 3.6a and b; Plots of $\text{Log}(Cov(\tilde{e}_{(a)}, \tilde{e}_{(a)}))$ and $\text{Log}(Cov(\hat{e}_a, \tilde{e}_{(a)}))$ for each observation.

In Fig. 3.6a $\eta / \sigma_0^2 = 20$ while for Fig. 3.6b $\eta / \sigma_0^2 = 40$.

The largest difference arises for the observation on rat 1. This is an extreme observation in the X sense given it is the only observation on this rat and it is the first observation in the dataset. There is also a pattern of greater differences occurring at both the earlier and later time points for each rat. This is due to the correlation between each observation within each rat. The increase in the ratio of the variance components between Figs.3.6(a) and (b) produces an almost identical pattern. The next element of this analysis involved considering subset deletion with the subset size being larger than a singleton.

Fig.3.6 Plots of leverage statistic for each subset (rat)



Figs. 3.6c and d; Plots of $\frac{|X_{(a)}^T V_{bb}^{-1} X_{(a)}|}{|X^T V^{-1} X|}$ for each subset. In Fig. 3.6(c) $\eta / \sigma_0^2 = 20$ while for Fig. 3.6(d)

$\eta / \sigma_0^2 = 40$.

There is a clear pattern in both Figs. 3.6(c) and (d) with the leverage statistic getting smaller as the subsets being deleted are getting larger. The smaller values for this statistic indicate high leverage subsets. This statistic was used to measure leverage as it can be used with any arbitrary subset. The Demidenko and Stukel (2005) approach outlined in section 2.3.3 was not used as our focus is not on whether the high leverage is due to extreme data in the X or Z space. Tables 3.4 (a) and (b) provide the corresponding values for the RVCs of $\tilde{\sigma}_{(a)}$ and $\hat{\sigma}_{(a)}$ for each deletion.

It is evident again that both approximations are very satisfactory. But there is evidence that $\hat{\sigma}_{(a)}$ is more accurate than $\tilde{\sigma}_{(a)}$ for high leverage subsets. In this example the high leverage values are due primarily to the size of the groups being deleted. To explore the issue further a balanced dataset was also considered.

Table 3.4 RVC using full REML refit, the CPJ approx and the proposed $\tilde{\sigma}_{(a)}$ for (a) σ_0^2 and (b) η under increasingly large and unbalanced deletions

(a)

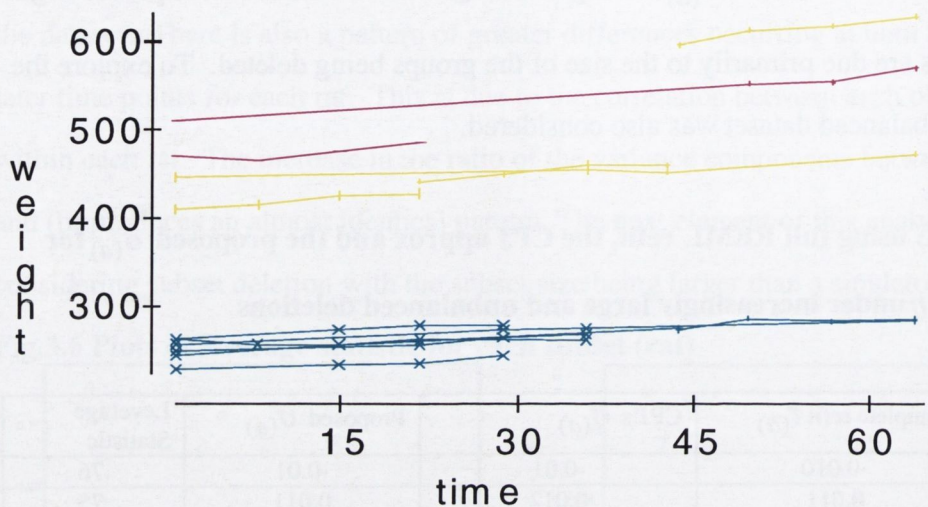
Deletion	Complete refit $\hat{\tau}_{(a)}$	CPJ's $\hat{\sigma}_{(a)}$	Proposed $\tilde{\sigma}_{(a)}$	Leverage Statistic
1	-0.010	-0.01	-0.01	.76
2	0.011	0.012	0.011	.75
3	0.013	0.013	0.012	.73
4	0.025	0.022	0.021	.71
5	0.041	0.038	0.036	.68
6	0.046	0.044	0.042	.58
7	0.042	0.039	0.036	.54
8	0.038	0.034	0.042	.38
9	0.045	0.039	0.048	.34

(b)

Deletion	Complete refit $\hat{\tau}_{(a)}$	CPJ's $\hat{\sigma}_{(a)}$	Proposed $\tilde{\sigma}_{(a)}$	Leverage Statistic
1	-0.008	-0.008	-0.008	.76
2	0.015	0.016	0.017	.75
3	0.042	0.040	0.039	.73
4	0.060	0.064	0.062	.71
5	0.062	0.060	0.058	.68
6	0.078	0.076	0.072	.58
7	0.093	0.100	0.104	.54
8	0.240	0.235	0.227	.38
9	0.450	0.441	0.426	.34

An alternative dataset was formed by randomly selecting four observations on rats 1-7 in group 1 on each of the four rats in group 2 and on rats 13 and 14 in group 3. The data are plotted in Fig. 3.7.

Fig. 3.7 Plot of modified rats dataset.



The first element of the analysis involved considering singleton deletion to identify the observations which showed large differences between $Cov(\tilde{e}_{(a)}, \tilde{e}_{(a)})$ and $Cov(\hat{e}_a, \tilde{e}_{(a)})$. Figs. 3.8 a and b are plots of $\text{Log}(Cov(\tilde{e}_{(a)}, \tilde{e}_{(a)}))$ and $\text{Log}(Cov(\hat{e}_a, \tilde{e}_{(a)}))$ for each observation.

Fig. 3.8 (a)

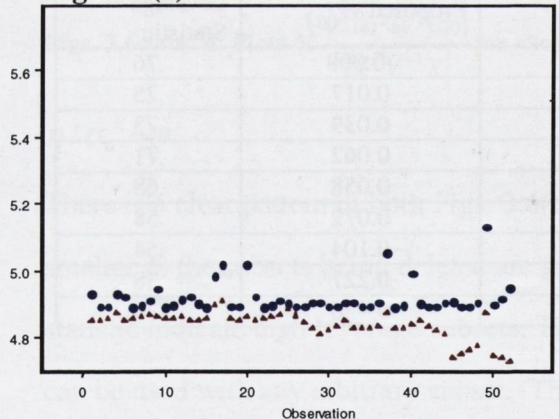


Fig. 3.8(b)

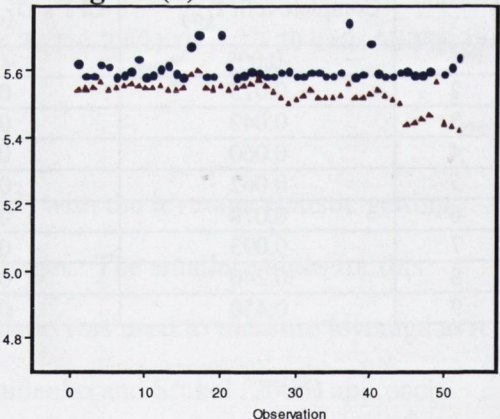


Fig. 3.8a and b; Plots of $\text{Log}(Cov(\tilde{e}_{(a)}, \tilde{e}_{(a)}))$ denoted by $\bullet$ and $\text{Log}(Cov(\hat{e}_a, \tilde{e}_{(a)}))$ denoted by $\blacktriangle$ for each observation. In Fig. 3.8(a) $\eta / \sigma_0^2 = 20$ while for Fig. 3.8(b) $\eta / \sigma_0^2 = 40$.

A similar pattern to Figs. 3.6 (a) and (b) is observed here for the difference in the ratio between the variance components. The plots show very similar patterns. The theory

would suggest that greater differences between $\hat{\sigma}_{(a)}$ and $\tilde{\sigma}_{(a)}$ will occur the larger the difference between $Cov(\tilde{e}_{(a)}, \tilde{e}_{(a)})$ and $Cov(\hat{e}_a, \tilde{e}_{(a)})$. Figs.3.8 (c) and (d) show the corresponding plots of the differences between these two quantities to Figs 3.8 (a) and (b).

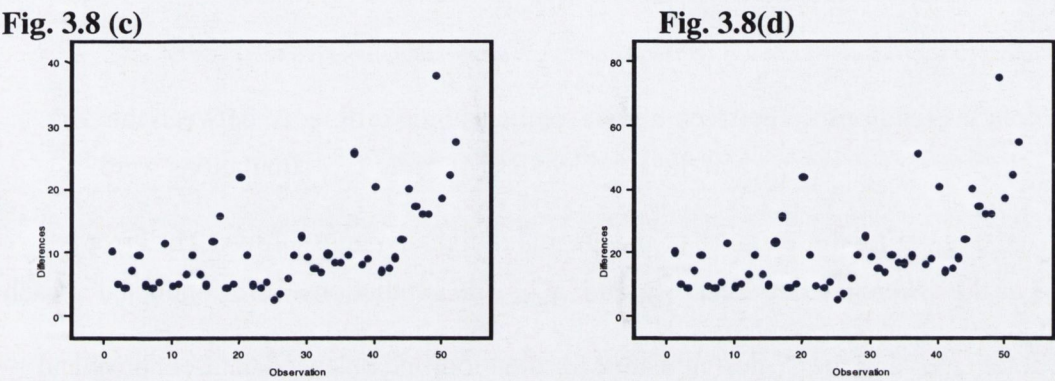


Fig. 3.8(c) and (d) Plots of the difference between $Cov(\tilde{e}_{(a)}, \tilde{e}_{(a)})$ and $Cov(\hat{e}_a, \tilde{e}_{(a)})$ for each observation. In Fig. 3.8(c) $\eta / \sigma_0^2 = 20$ while for Fig. 3.8(d) $\eta / \sigma_0^2 = 40$.

The three observations with the largest differences are observations 49,52 and 37. These are observations 1 and 4 on rat 13 and observations 1 on rat 10. These observations are extreme in the X sense. For example observations 49 and 37 are at time point 1 while the other observations on these rats are at the final three time points.

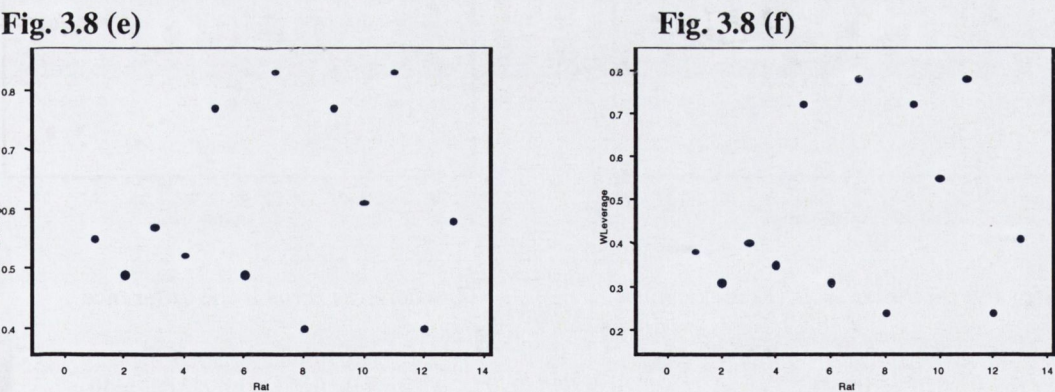


Fig.3.8 (e)-(f) Plot of $\frac{|X_{(a)}^T V_{bb}^{-1} X_{(a)}|}{|X^T V^{-1} X|}$ for each rat. In 3.8(e) $\eta / \sigma_0^2 = 20$ while for 3.8(f) $\eta / \sigma_0^2 = 40$.

These points are the highest leverage points based on the q_{ii} values. This next stage of the analysis involves considering subset deletion. The rats with highest leverage are rats 8 and 12. Both rats are extreme in the X sense in that they are the only rats in their group with observations at these time points.

To explore the relationship between high leverage and the difference between the proposed approximation $\hat{\sigma}_{(a)}$ and the complete REML refit $\hat{\tau}_{(a)}$ simulations were conducted comparing the variance of the error and high leverage subsets. The error is defined as the difference between $\hat{\sigma}_{(a)}$ and $\hat{\tau}_{(a)}$. 100 simulations were conducted at each of the deletions using the X design matrix for the modified balanced dataset above and for both ratios of variance components. The plots are given below. The $Var(\text{error})$ is multiplied by 10^6

Fig. 3.8(g)

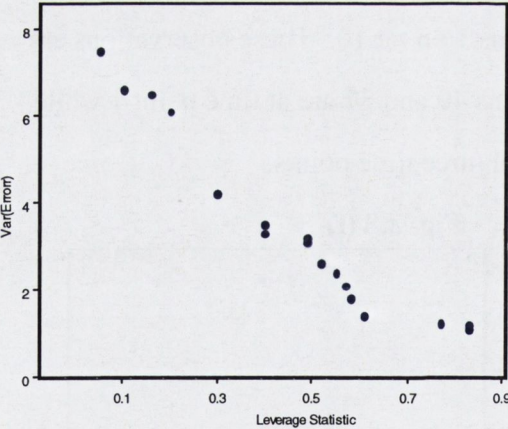


Fig.3.8(h)

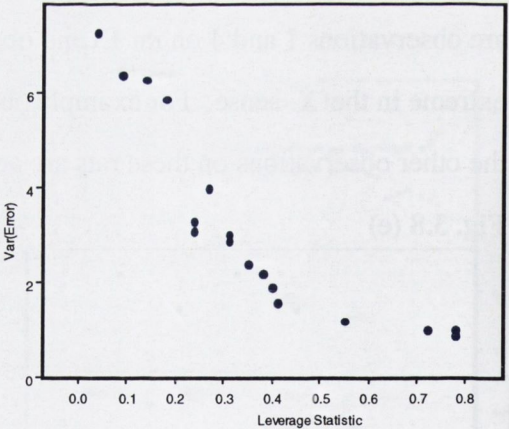


Fig. 3.8(g) and (h) Plot of $Var(\text{Error})$ against $\frac{|X_{(a)}^T V_{bb}^{-1} X_{(a)}|}{|X^T V^{-1} X|}$ where the error is the difference

between the REML estimate $\hat{\tau}_{(a)}$ and $\tilde{\sigma}_{(a)}$. In 3.8(g) $\eta / \sigma_0^2 = 20$ while for 3.8(h) $\eta / \sigma_0^2 = 40$

An analogous analysis was conducted to consider the difference between $\tilde{\sigma}_{(a)}$ and $\hat{\sigma}_{(a)}$. The corresponding graphs are shown below.

Fig. 3.8(i)

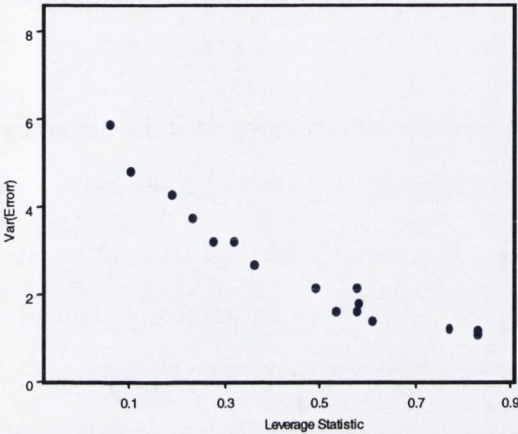


Fig.3.8(j)

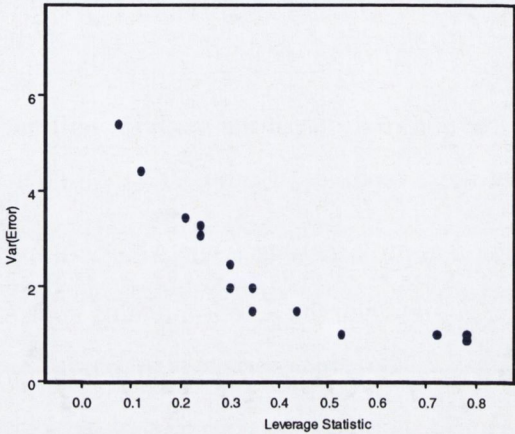


Fig. 3.8(g) and (h) Plot of $Var (Error)$ against $\frac{|X_{(a)}^T V_{bb}^{-1} X_{(a)}|}{|X^T V^{-1} X|}$ where the error is the difference between the REML estimate $\hat{\tau}_{(a)}$ and $\tilde{\sigma}_{(a)}$. In 3.8(g) $\eta / \sigma_0^2 = 20$ while for 3.8(h) $\eta / \sigma_0^2 = 40$

There is a clear pattern in both Figs. 3.7(g) and (h). Small values of the leverage statistic indicate high deletions. Thus there is evidence that the variance of the error is greater for leverage high leverage subsets. A similar pattern is present in Figs. 3.7 (i) and (j). It is clear that from this analysis that both approximations provide very good approximations for $\hat{\tau}_{(a)}$ even for high leverage subsets.

3.4 Summary and Discussion

In this chapter, the focus has been on identifying observations and groups of observations which are influential on the estimates of the variance components in the LMM. The “delete = replace” identity (Haslett, 1999) is developed further to enable a more computationally efficient approach for CPI’s one-step approximation. An alternative approach is developed which is even more computationally frugal. It may also be used

with ANOVA estimation techniques to provide estimates of the change in the variance components following deletion. Thus it is easy to generate excellent approximations for the influence of subsets of the data as measured by deletion on the estimates of variance components in the LMM as by products of the REML and ANOVA fitting procedures.

The comparisons of the methods outlined in this chapter provide evidence of the accuracy of these methods. Both CPJ's method and the proposed approximation $\tilde{\sigma}_{(a)}$ are very good approximations. The ANOVA approach using Henderson's Method III is very accurate when the deletion retains approximate balance. However, the more unbalanced the dataset becomes following deletion the more they deviate from the REML estimates. The inadequacies of ANOVA methods for estimating variance components are well documented in the literature. They may produce negative estimates and are non-optimal for unbalanced data. Thus the effectiveness of ANOVA methods as diagnostic tools is limited.

Of the three approaches the proposed approximation of equation 3.13 is the most computationally efficient. This method is also applicable where there is autocorrelation in the model if the correlation parameter is assumed to remain unchanged following deletion; this is typically modeled as $Var(\gamma_0) = \sigma_0^2 R(\Theta)$ with R block diagonal. Such models are frequently used in the analysis of longitudinal data. In chapter 4 consideration is given to assessing the influence of data on autocorrelation model parameters.

The approaches outlined in this chapter deal exclusively with variance structure diagnostics. They cannot be used directly to assess the impact on the fixed parameters of observations that are highly influential on the variance parameters. The effect of the deletion on the fixed parameters is estimated as $\hat{\beta}_{(a)}(\hat{\tau})$ and so a more desirable approach

would be $\hat{\beta}_{(a)}(\tilde{\sigma}_{(a)})$ or $\hat{\beta}_{(a)}(\hat{\sigma}_{(a)})$. The development of 1-step diagnostic measures for the mean parameters based on $\tilde{\sigma}_{(a)}$ or $\hat{\sigma}_{(a)}$ is considered in Chapter 5.

Chapter 4

Autocorrelation parameter diagnostics for the LMM

4.0 Introduction

The LMM is an important tool for modeling longitudinal data. In this context, the LMM will often incorporate an autocorrelation parameter for its error structure. The methods proposed in chapter 3 apply in situations where the covariance matrix may be represented by a linear combination of fixed matrices. Some models for autocorrelated data may be expressed in such a form. However, other models including the AR(p) model cannot be expressed in this form and other methods must be used. The focus in this chapter is on developing computationally efficient methods for estimating the change in the autocorrelation parameters following deletion of individual (or groups of) observations. The methods proposed are used in chapter 5 for the development of deletion diagnostics for the fixed effects that incorporate changes in V following deletion.

The methods proposed exploit the “delete = replace” identity, are a by-product of the fitting process and are computationally efficient. The method proposed for MA(q) models are based on the approaches in chapter 3. For the AR(p) family of models, method of moments estimation is used. The usefulness and accuracy of the techniques are considered by their application to data.

4.1 Linear mixed effects models for autocorrelated data

The LMM was defined in equation 1.5 where V was expressed as $A + \sum_{i=1}^r \sigma_i^2 Z_i Z_i^T$. The variance component diagnostics proposed in chapter 3 are feasible because it is possible to represent the covariance matrix by a linear combination of fixed matrices. For autocorrelated data, A will be of the form $\sigma_0^2 W(\rho)$. In time series or longitudinal data applications the γ_j parameters are often null with A modeled by an ARIMA (p,d,q) process. The principal models used in such applications are AR type models with MA models also used. The focus of the work in this chapter is to evaluate $\hat{\rho}_{(a)}$ for these models.

For situations where the covariance matrix may be expressed in the form

$V = \sum_{j=0}^r \sigma_j^2 Z_j Z_j^T$ the REML based approaches of chapter 3 are applicable. LMMs

involving moving average processes may be expressed in this form. The MA(1) model has a one-lag correlation, meaning that $cor(\varepsilon_t, \varepsilon_{t+1}) = \rho$ but $cor(\varepsilon_t, \varepsilon_{t+l}) = 0$ for $l > 1$.

The covariance matrix V for an LMM is expressed for such a model as

$\sigma_0^2 W(\rho) + \sum_{i=1}^r \sigma_i^2 Z_i Z_i^T$. This may be expressed as

$$V = \sigma_0^2 (I + \rho T_1) + \sum_{j=1}^r \sigma_j^2 Z_j Z_j^T \quad (4.1)$$

where T_1 is an elementary Toeplitz matrix of the first order which has 1 at the two parallel subdiagonals and zero elsewhere. The MA(2) model accommodates autocorrelation of the second order with a covariance matrix of the form

$$V = \sigma_0^2(I + \rho_1 T_1 + \rho_2 T_2) + \sum_{j=1}^r \sigma_j^2 Z_j Z_j^T \quad (4.2)$$

where T_2 is a second order Toeplitz matrix. For example the 5 x 5 matrices T_1 and T_2 are

$$T_1 = \begin{pmatrix} 01000 \\ 10100 \\ 01010 \\ 00101 \\ 00010 \end{pmatrix} \text{ and } T_2 = \begin{pmatrix} 00100 \\ 00010 \\ 10001 \\ 01000 \\ 00100 \end{pmatrix} \quad (4.3)$$

It is possible to estimate $\rho_{(a)}$ using the methods in chapter 3 when V is expressed in this form. The CPJ (equation 3.8) and Haslett and Dillane (2004) (equation 3.13) approaches are now applicable with for example $\tilde{\sigma}_{(\alpha_i)} = (\tilde{\rho}_{(a_i)}, \tilde{\sigma}_{0(a_i)}^2, \tilde{\sigma}_{1(a_i)}^2, \dots, \tilde{\sigma}_{r(a_i)}^2)$ and $\tilde{\sigma}_{(p)}$ is easily computable.

Autoregressive models.

Another popular statistical model of a time-series is autoregression. The simplest is the first order autoregressive AR(1) model, which is widely used for modelling autocorrelation in longitudinal data. Unlike the moving average model described above it is not possible to express the LMM incorporating an AR model for the autocorrelation in an analogous form to that defined in equations 4.1 or 4.2. Thus the one-step REML approaches for LMMs outlined in chapter 3 are not applicable for AR models and an alternative approach must be used.

The sample ACF defined in equation 4.4 is the key to method of moments (MM) estimation for AR(p) models. Cryer (1986) compares the estimation of model parameters by MM on a number of simulated ARIMA series. This analysis showed that this approach is very accurate for estimating autoregressive model parameters especially in

the AR(1) case. For the general AR(p) model let ϕ denote the vector of parameters. The focus of interest in this chapter is on estimating $\hat{\phi} - \hat{\phi}_{(a)}$, that is the impact of deleting individual (or groups of) observations on the model parameter estimates. The expression for $r_{l(p)}$ derived below is used for this purpose.

Consider any sequence of data $y_1, y_2, \dots, y_n$ whether residuals, original data, or some transformation of the original data such as differenced data. The autocorrelation function is the basic tool underlying time series analysis and using method of moments estimation provides estimates for the AR(p) parameters. The sample ACF for a stationary process, which is defined as

$$r_l = \frac{\sum_{t=1}^n (y_t - \bar{y})(y_{t+l} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}, \quad l = 1, 2, \dots \quad (4.4)$$

may be expressed as

$$r_l = \frac{\hat{e}^T D \hat{e}}{\hat{e}^T \hat{e}} \quad (4.5)$$

where $\hat{e}$ is the vector of residuals and D is a square matrix whose structure depends on the lag l being considered. For $l = 1$ D is of the form

$$D = \begin{pmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & 1 & \dots & . \\ . & . & . & 1 \\ 0 & 0 & . & 0 \end{pmatrix}.$$

Given that $\hat{e} = GQY$ (4.5) may be expressed as

$$r_l = \frac{Y^T Q A Q Y}{Y^T Q G^T G Q Y} \quad (4.6)$$

where $A = G^T D G$ and G will depend on the residuals being used. For the OLS residuals $G = I$. For the LMM defined in equation 1.5 $G = A$. The impact of deleting a subset a on r_i is

$$r_{l(a)} = \frac{Y_{\tilde{a}}^T Q A Q Y_{\tilde{a}}}{Y_{\tilde{a}}^T Q M Q Y_{\tilde{a}}} \quad (4.7)$$

with $M = G^T G$. From appendix 1 the numerator and denominator in equation 4.7 may be written as $Y_{\tilde{a}}^T R Y_{\tilde{a}}$ and thus computed as $Y^T R Y - 2Y^T R \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix} + \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix}^T R \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix}$. Thus we can estimate $r_{l(a)}$ as a by-product of the fitted model. The principal motivation for using this approach is that it allows easy computation of $r_{l(p)} = (r_{l(a_1)}, r_{l(a_2)}, \dots, r_{l(a_k)})$, a $(1 \times k)$ matrix using the complete set of conditional residuals $\tilde{e}_{(p)} = \{\tilde{e}_{(a_1)}^T, \tilde{e}_{(a_2)}^T, \dots, \tilde{e}_{(a_k)}^T\}$.

The numerator and denominator in equation 4.7 may be computed as $Y^T Q A Q Y - \tilde{e}_{(a)}^T Q_{aa} A_{aa} Q_{aa} \tilde{e}_{(a)}$ and $Y^T Q M Q Y - \tilde{e}_{(a)}^T Q_{aa} M_{aa} Q_{aa} \tilde{e}_{(a)}$ respectively. This allows $r_{l(p)}$ to be computed *en bloc* using the results in Appendix 1. Let A_p and M_p be block diagonal with blocks A_{a_j} and M_{a_j} respectively with $j = 1, \dots, k$. Let S_p denote the $K \times 1$ matrix with elements $Y Q A Q Y$. Let N_p denote the $K \times 1$ matrix with elements $\tilde{e}_{(a_j)}^T Q_{aa_j} A_{aa_j} Q_{aa_j} \tilde{e}_{(a_j)}$. Using the approaches outlined in section 1.3.4 N_p may be computed *en bloc* as $\tilde{e}_p^T D_p A_p \square D_p \tilde{e}_p$. The $K \times 1$ matrix containing the numerator terms is simply $S_p - N_p$. The denominator may be computed in a similar fashion and thus $r_{l(p)}$ can be computed *en bloc*.

For an AR(1) model $\hat{\phi}_1 = r_1 = \hat{\rho}$ and from above $\hat{\rho} - \hat{\rho}_{(p)}$ is easily estimated. For the AR(2) model $\hat{\phi}_1 = \frac{r_1(1-r_2)}{1-r_1^2}$ and $\hat{\phi}_2 = \frac{r_2-r_1^2}{1-r_1^2}$. The impact of deleting a subset of size a on

the parameters is thus $\hat{\phi}_{1(a)} = \frac{r_{1(a)}(1-r_{2(a)})}{1-r_{1(a)}^2}$ and $\hat{\phi}_{2(a)} = \frac{r_{2(a)}-r_{1(a)}^2}{1-r_{1(a)}^2}$. The method of

moments approach is very accurate for a stationary AR process (Cryer, 1986). However, in this context the focus is on the change in the parameter estimates arising from deletion rather than the actual parameter estimates. A number of methods were considered for measuring this change including a 1-step ML or REML approach. The method of moments approach outlined above was the only methodology, which produced $\hat{\rho} - \hat{\rho}_{(p)}$ in a computationally efficient manner.

4.2 Evaluation of Proposals

The approaches are considered by their application to data. In section 4.2.1 the Dillane and Haslett (2004) approach using equation 4.1 is applied to a longitudinal dataset where the autocorrelation is modeled by an MA(1) process. The accuracy of the method is contrasted with re-fitted REML estimates. In section 4.2.2 the accuracy of the method of moments based approach is contrasted with the REML estimates for a linear regression model with an AR(1) model for the autocorrelation in the data. In section 4.2.3 the proposed methods are applied to the longitudinal dataset described in section 3.2.3. Finally in section 4.3 consideration is given to the use of these methods in computing time series diagnostics proposed by Bruce and Martin (1989). The primary motivation for developing these approaches is their contribution to the methods proposed in chapter 5 for fixed effect deletion diagnostics.

The usefulness of the methods in chapter 5 depends on the changes in the estimates of the variance and correlation parameters following deletion being easily computable. The proposed methods provide easily computable updates of the changes in the autocorrelation parameters.

4.2.1 longitudinal data with MA(1) errors.

The data shown in Figure 4.1 refer to an experiment conducted to consider the effect of three preservatives on the growth rate of cultures of Pseudomonas fungi, a food spoiler. The response noted is the percentage change in cell mass at hourly intervals over nine hours. The group sizes were 4,4 and 4 respectively. The model fitted is given below.

$Y=\alpha_1 + \delta_1t$

For group 1

$Y=\alpha_2 + \delta_1t$

For group 2

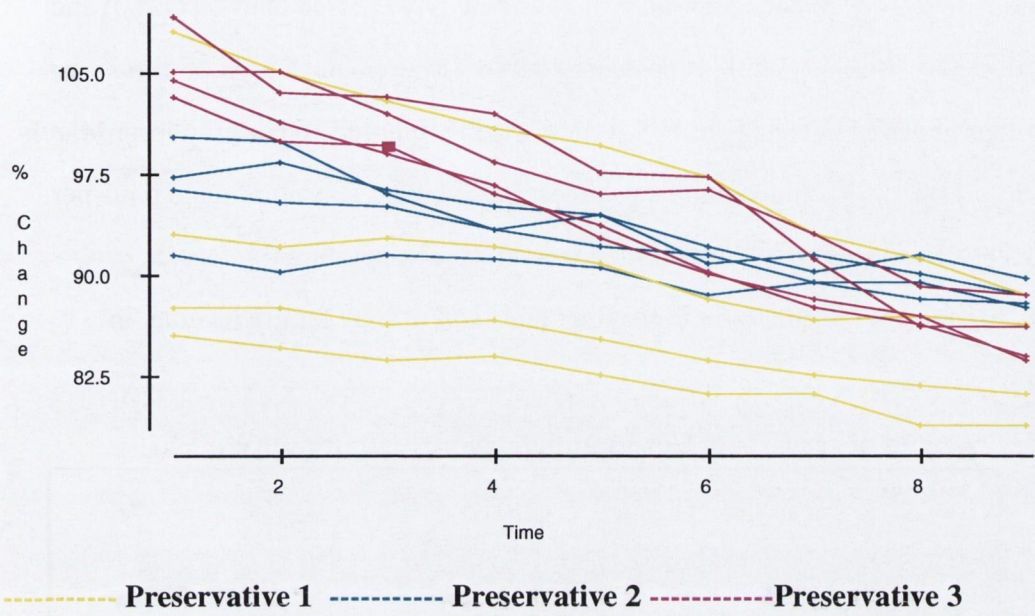
$Y=\alpha_3 + \delta_1t$

For group 3

The model was fitted using REML and the V matrix was block diagonal of the form $\sigma_1^2J + E$ where J is a 9 x 9 matrix whose elements are all 1, σ_1^1 reflects the covariance within each culture's measurements by its deviation from the group intercept and E is 9 x 9 and reflects the correlation between successive measurements on the same experimental unit with an MA(1) structure fitted. The results of the model fit using REML estimation were as follows

Mean Parameters	α_1	α_2	α_3	δ_1
	106.2	96.8	93.4	-1.18
Variance Parameters	σ	ρ	σ_1	
	8.22	0.78	14.99	

Fig. 4.1 Longitudinal plot of % change in *Pseudomonas* fungi.



Our focus in this analysis is on the relative change in the autocorrelation parameter ρ .

Fig.4.2(a)

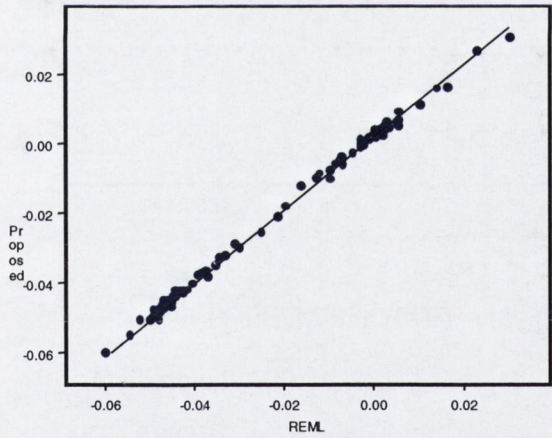


Fig.4.2(b)

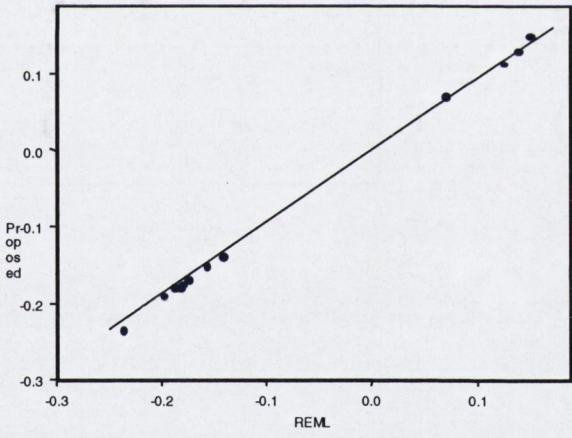


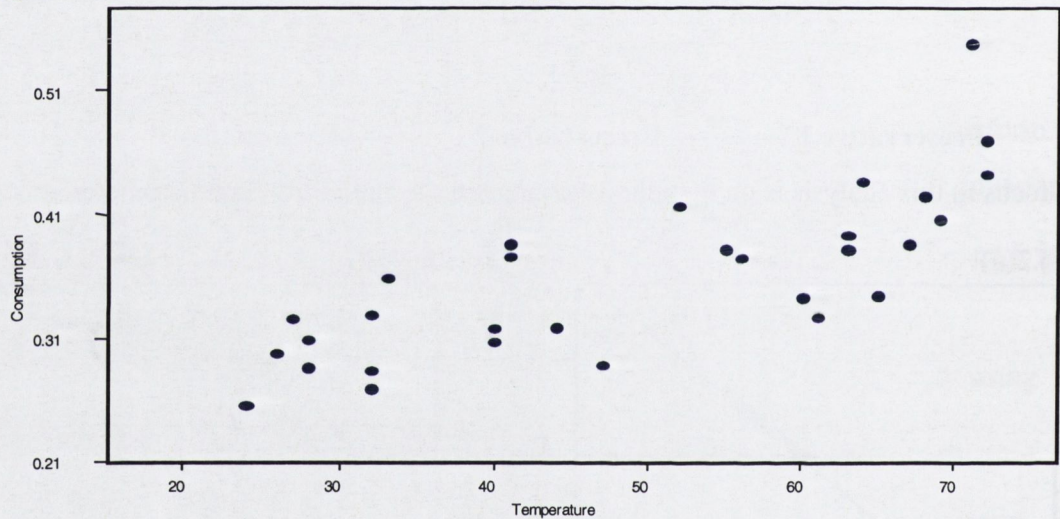
Fig. 4.2 Plot of $\frac{\hat{\rho}_{(a)}}{\hat{\rho}} - 1$ using REML and proposed method for each observation in (a) and each preservative in (b)

The Haslett and Dillane (2004) proposed method is used to estimate the change in the autocorrelation parameter for the MA(1) model. It is clear from Figs.4 2 (a) and (b) that the proposed method provides a good estimate of the changes in the autocorrelation parameter for these data.

4.2.2 Ice cream data

The dataset on ice-cream consumption first considered by Hildreth and Lu (1960) and later analysed by Martin (1992) is considered below. These data consist of thirty observations gathered in Michigan U.S.A. at 4-weekly intervals, reportedly from March 18th to July 11th 1953. The variable of interest is ice-cream consumption in pints per person. The other three variables recorded were mean temperature (in degrees Fahrenheit), price of ice cream (in dollars per pint) and weekly family income (in dollars).

Fig. 4.3 Plot of Ice Cream Consumption against temperature.



Based on analyses by Hildreth & Lo (1960), Kadiyala (1970) and Martin (1992) a regression model with just mean temperature (in degrees Fahrenheit) as the predictor using AR(1) errors appears to be adequate. The REML fit estimates ρ as 0.885. The mean parameter estimates are $\hat{\beta}_0 = .228$ and $\hat{\beta}_1 = .003$. The data are represented graphically in Fig. 4.3. At this stage our interest is in estimating the impact on the AR model parameter of observation and subset deletion. Figs. 4.3.1 (a) and (b) show the leave one-out and leave two out relative change in the autocorrelation parameter defined as $\frac{\hat{\rho}_{(a)}}{\hat{\rho}} - 1$. From the analysis it is clear that observations 1 and 30 are most influential in

estimating the autocorrelation parameter ρ . Figs. 4.3.1 (c) and (d) compare the relative change in the autocorrelation parameter computed using a complete REML re-fit and using the proposed method. The proposed method provides a very good approximation for the complete REML re-fit.

This analysis identifies data influential on the AR(1) parameter. In chapter 5 our interest focuses on the mean structure diagnostics. However, consideration of data influential on the correlation structure may be an important first step in any diagnostic analysis.

Fig. 4.3.1 Deletion plots for ice cream data

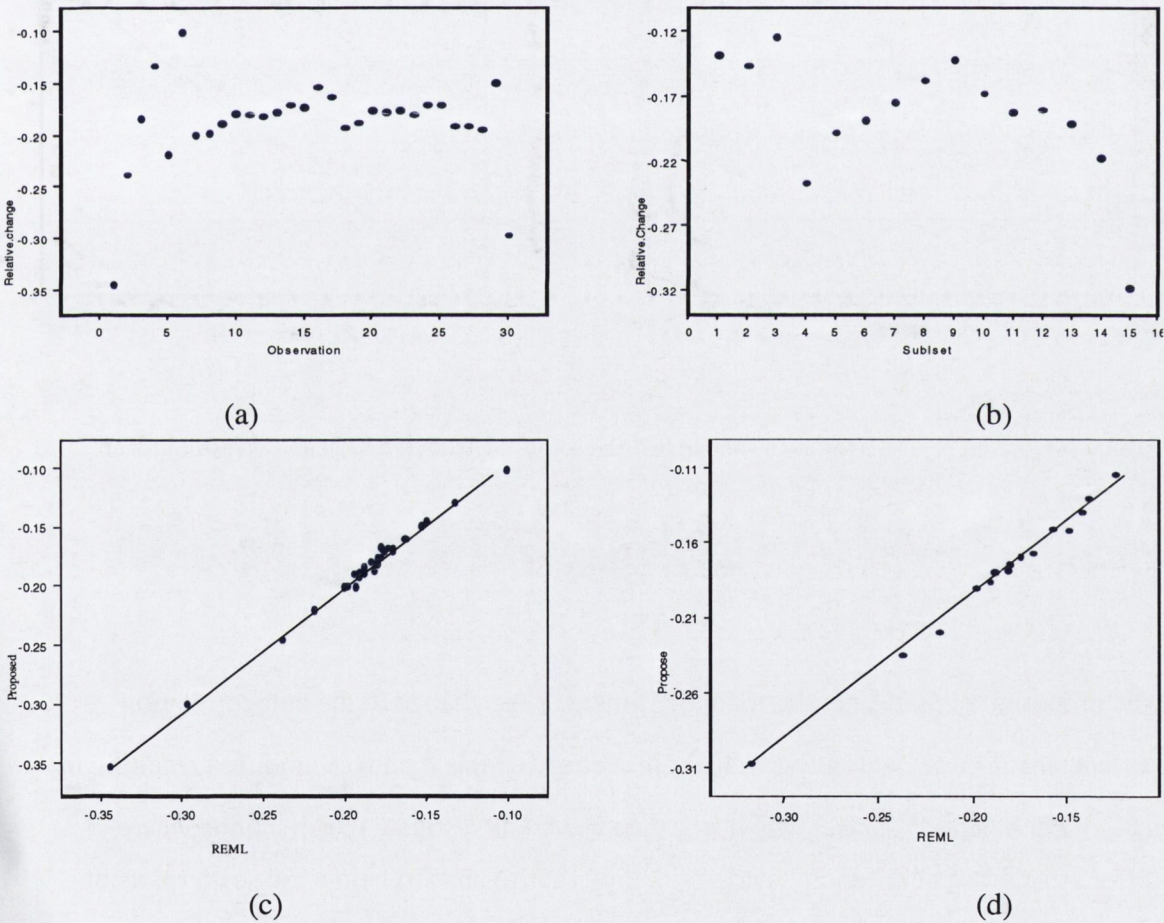


Fig. 4.3.1 (a) and (b) show the relative change in the autocorrelation parameter based on leave-one-out and leave-two out deletions respectively: (c) and (d) compare $\frac{\hat{\rho}_{(a)}}{\hat{\rho}} - 1$ computed using a full REML fit and using the proposed method.

4.2.3 Rats Data (Crowder and Hand, 1990)

The final example in this chapter was considered in section 3.2.3. The effect of deleting each rat on the variance components was plotted in Figs. 3.4 (a) and (b). Fig. 4.4 below is a similar plot except the focus is on the autocorrelation parameter ρ . It is clear from the plot that no individual rat has a very large influence on the correlation parameter. Intuitively it is unlikely that changes in the correlation parameter in such models solely would have large impacts on the fixed effect diagnostics. This is considered in chapter 5.

Fig.4.4 Deletion plots for autocorrelation parameter for rats data

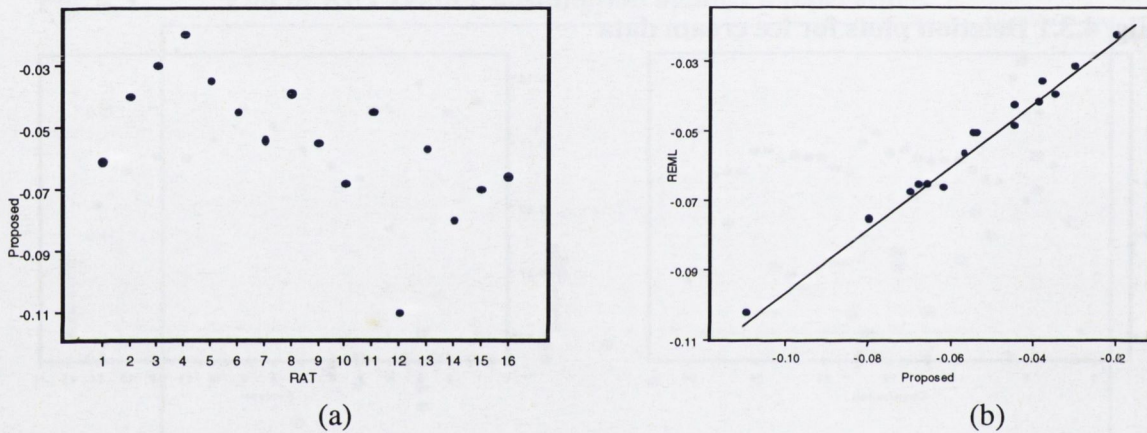


Fig.4.4 (a) Plot of $\frac{\hat{\rho}_{(a)}}{\hat{\rho}} - 1$ for each rat: (b) Scatterplot of the REML and proposed estimates of

$$\frac{\hat{\rho}_{(a)}}{\hat{\rho}} - 1.$$

The proposed method provides a good estimate of the change in the autocorrelation parameter for these data in Fig. 4.4(b). In section 3.3 this data was modified resulting in a highly unbalanced dataset which was then used to investigate the situations where differences arose between $\tilde{\sigma}_{(a)}$ and $\hat{\sigma}_{(a)}$. Figs. 4.5 (a) and (b) below are scatterplots of $\frac{\hat{\rho}_{(a)}}{\hat{\rho}} - 1$ estimated using a complete REML refit following deletion and using the

proposed method for both observation and subject deletion respectively for the modified dataset

Fig. 4.5 Deletion Plots for modified rats dataset

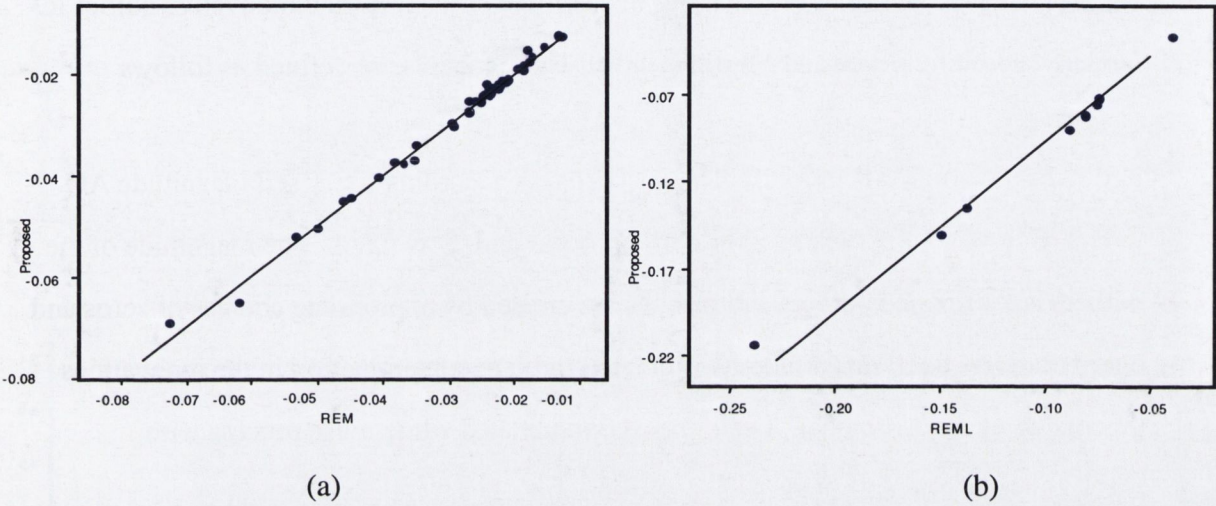


Fig. 4.5 Deletion plots for modified rats dataset (a) and (b) compare $\frac{\hat{\rho}_{(a)}}{\hat{\rho}} - 1$ computed using a full REML fit and using the proposed method for observation and subject deletion respectively.

While there is some scatter around the line $Y = X$ the approximation provides a good approximation to the full REML refit for these data

4.3 A further application of these proposals

Bruce and Martin (1989) have explored the application of deletion diagnostics for time series ARIMA models. Their analysis involved considering deletion of both individual and arbitrary subsets of observations. Their proposed diagnostics $DC(A)$ and $DV(A)$ measure the change in the parameter vector and the change in the estimated variance and were defined in equations 2.9 and 2.10 respectively. Their approach differs to other methods in that all parameters are re-estimated using ML or REML when observations are deleted.

In this section the proposed methods of this chapter are used to compute the $DC(A)$ and $DV(A)$ diagnostics and these are compared with those computed using REML. In the examples below the focus is on influential data caused by outliers. The examples involve simulated AR(p) and MA(q) data using the additive outlier AO and innovative outlier IO models as used by Bruce and Martin (1989). These models are defined as follows

Let x_t be a Normal AR process. Then y_t behaves according to a fixed magnitude AO model if $y_t = x_t + \xi_t z_t$ where z_t is a 0–1 process and ξ_t is fixed. The magnitude of the outliers is ξ_t ; isolated outliers and patches are created by appropriate choices of zeros and ones for z_t . A fixed magnitude IO is created through contamination in the innovations process ε_t of the AR model. Let ε_t be a contaminated white noise process with

$$\varepsilon_t = \hat{\varepsilon}_t + \xi_t z_t$$

where $\hat{\varepsilon}_t$ are independent Normal random variables with zero mean and variance σ^2 with ξ_t and z_t as above.

4.3.1 a Autoregressive model

This example involves an AR(2) series of 50 points from Normal white noise with an IO type outlier at point 7 of +3. The data are plotted in Fig. 4.6(a) with the leave-one-out DC diagnostic computed using REML and MM in Fig 4.6 (b.) A scatterplot of both is shown in Fig. 4.5 (c) with the corresponding plots for DV in Figs. 4.6 (d) and (e).

Example 1 AR(2) $\phi_1 = .5, \phi_2 = .4, \sigma^2 = 1$, innovations outliers model with one isolated outlier (obs. 7).

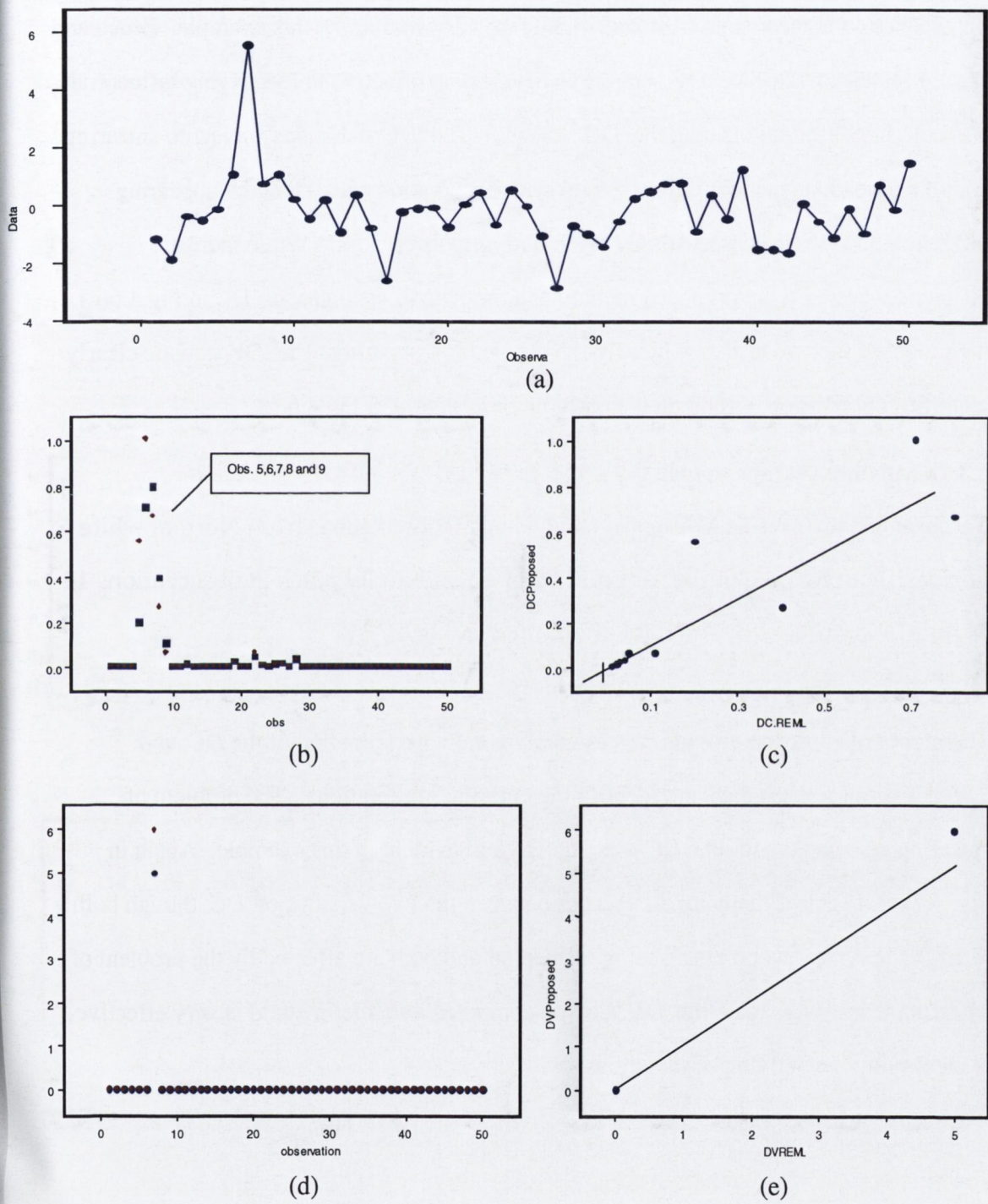


Fig.4.6 Analysis on simulated data. (a) Data (b) Plot of *DC* for each observation * denotes computed using method of moments approach,* denotes computed using REML (c) Scatterplot of *DC* computed using REML and using proposed method with line $Y = X$ * denotes computed using method of moments approach,* denotes computed using REML (d) Plot of *DV* for each observation . (e) Scatterplot of *DV* based on REML and proposed method with line $Y = X$.

The example above provides some evidence of the usefulness of the “delete = replace” approach for identifying outlying observations in time series data. The *DC* and *DV* diagnostics computed using MM and REML are very similar for this example. Bruce and Martin (1989) identify the *DV* statistic as being more effective in identifying influential and outlying observations than the *DC* statistic. The *DC* statistic is subject to smearing and this problem is evident in Fig. 4.6 (b) with observations 5,6 7,8 and 9 appearing influential. However, observation 7 is the outlying observation. While there are differences between the proposed method and the REML estimates of *DC* in Fig.4.6(c) the proposed method is also subject to the problem of smearing. The *DV* statistic clearly identifies observation 7 as being the outlying observation in Fig.4.6 (d).

4.3.1b Moving Average model

This example involves an MA model comprising 50 observations from Normal white noise with an IO type outlier at point 35 of +3 and an IO patch at observations 14 15 and 16 of order +1. The data are plotted In Fig.4.7 (a).

The proposed approach provides a reasonably good approximation of the *DC* and *DV* statistics for these data and identify the outlying observations. The problem of smearing associated with the *DC* statistic is again evident in this example. Again in Fig. 4.7 (c) it is clear there are differences between the two versions of *DC* though both identify the same data points as being influential and both are affected by the problem of smearing. However, again the *DV* statistic computed by either method is very effective at identifying the outlying observations.

Example 2 MA(2) $\theta_1 = .5, \theta_2 = .4, \sigma^2 = 1$, innovations outliers model with one isolated outlier (obs. 7).

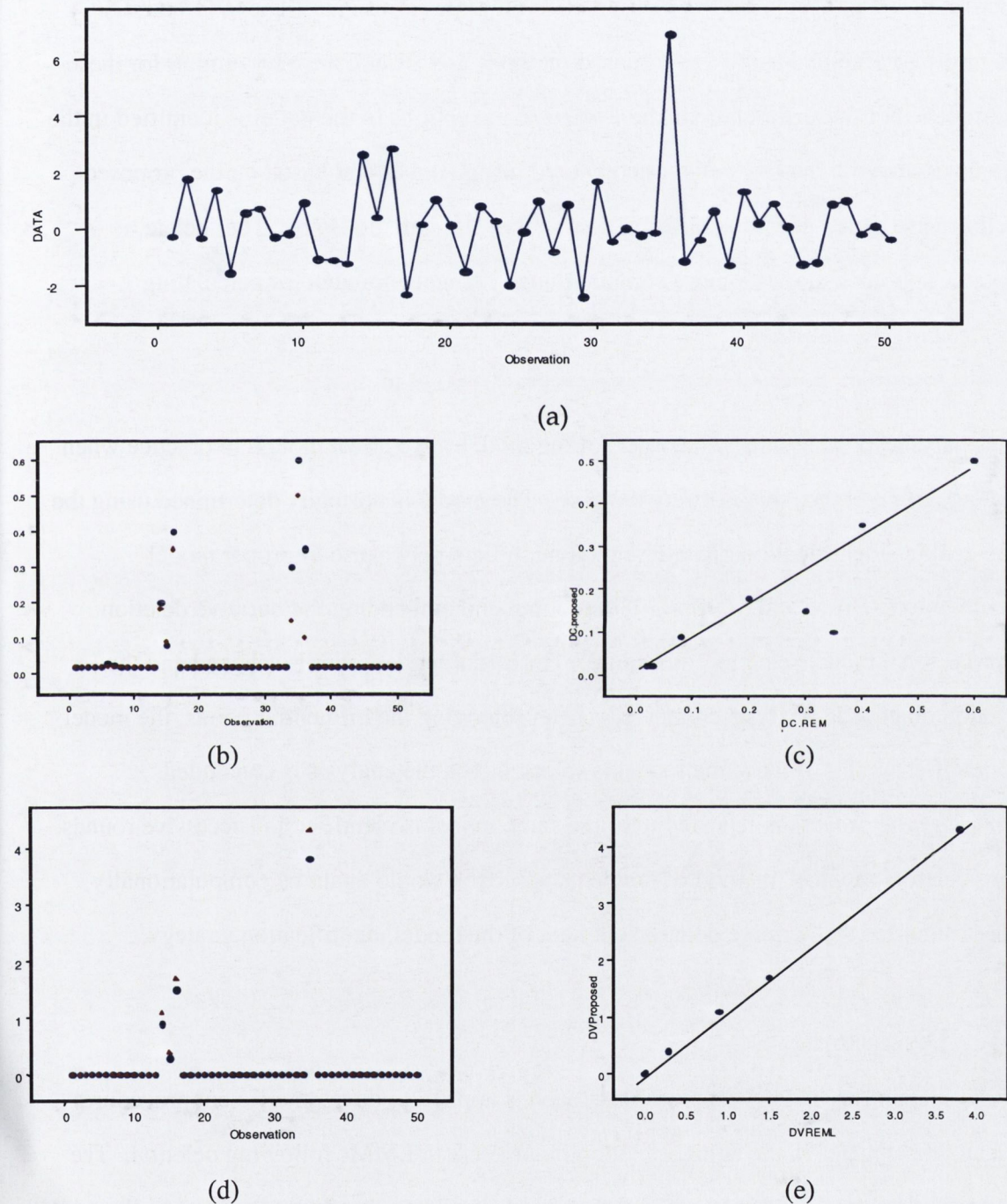


Fig.4.7 Analysis on simulated data ; (a) Plot of data; (b) Plot of leave-1-out DC , * denotes computed using proposed method * denotes computed using REML; (c) Scatterplot of DC computed using REML and one-step REML with line $Y = X$. (d) Plot of leave-one-out DV * denotes computed using proposed method * denotes computed using REML. (e) Scatterplot of DV based on REML and proposed method with line $Y = X$.

The examples above provide some evidence of the usefulness of the “delete = replace” approach for identifying outlying observations in time series data. The *DC* and *DV* diagnostics computed using the proposed methods and REML are very similar for these examples. For these simulations there were no exceptions to the patterns identified in the examples above. The *DC* statistics computed using REML and based on the proposed methods are both subject to smearing though the *DV* statistic is not. The ‘delete = replace’ approach allows a more computationally frugal approach for calculating the *DC* and *DV* statistics.

The analysis above assumes the order for the model was known though in practice when dealing with real data this is rarely the case. The model is normally determined using the Box-Jenkins identification procedure though outliers may cause improper model specification. Bruce and Martin (1989) recommend embedding the iterative deletion approach in a model selection procedure. The initial model order is selected and the deletion diagnostic analysis conducted. After removing the influential points, the model is identified again. If the same model is selected then the analysis is concluded. Otherwise the process is repeated until the same model is identified in successive rounds. The “delete = replace” method of moments approach would again be computationally more frugal for the iterative deletion element of the model identification strategy.

4.4 Discussion

In this chapter the ‘delete = replace’ approach is applied to develop efficient procedures to estimate changes in the autocorrelation parameters in LMMs following deletion. The primary motivation for developing such methods is to help develop computationally efficient fixed effects diagnostics that incorporate changes in the covariance parameters following deletion.

In considering fixed effects diagnostics for such models it is usual to assume that the covariance matrix V remains unchanged upon deletion and the various diagnostic measures outlined in section 1.3 are computed on this basis. An alternative and perhaps more prudent approach would be to first identify observations influential on the covariance parameters by using the approaches outlined in chapter 3 and in this chapter. This would identify observations or subsets of data where it may be wise to estimate the mean structure diagnostics based on $V(Y_b)$ rather than the usual $V(Y)$. In chapter 5 a method is proposed that uses estimates of $V(Y_b)$ based on the proposed methods for estimating the fixed effects diagnostic measures.

Chapter 5

Mean structure Diagnostics

5.0 Introduction

In this chapter the focus returns to mean structure deletion diagnostics. In chapters three and four, deletion diagnostics based on the “delete = replace” approach were proposed for the variance and autocorrelation parameters in LMMs. As observed in section 2.3.2 the mean structure diagnostics proposed in the literature are generally computed on the basis that the variance and correlation parameters are unchanged following deletion. The question of whether ignoring the changes in these parameters following deletion gives rise to misleading inferences from the mean structure diagnostics has not been explored in the literature. However, it is acknowledged (Christensen *et al*, 1992) that this is likely, particularly in the case of subset deletion. This question is addressed in section 5.1 through an analysis on the county crop areas dataset described in section 3.4.2 and the rats dataset in section 3.4.3.

One approach that could be adopted is to use the methods outlined in chapters 3 and 4 to identify observations or subsets that are influential on the covariance structure parameters and then conduct a full re-fit, using REML or ML to compute the mean structure deletion diagnostics for these subsets. It would undoubtedly be more desirable to develop a 1-step procedure for computing the mean structure diagnostics that incorporates the changes in the variance parameters arising from deleting observations and subsets. In section 5.2 such a procedure based on Taylor series approximations is proposed. In section 5.4 we subject this to some rather extreme deletions to explore the generality of the approximation.

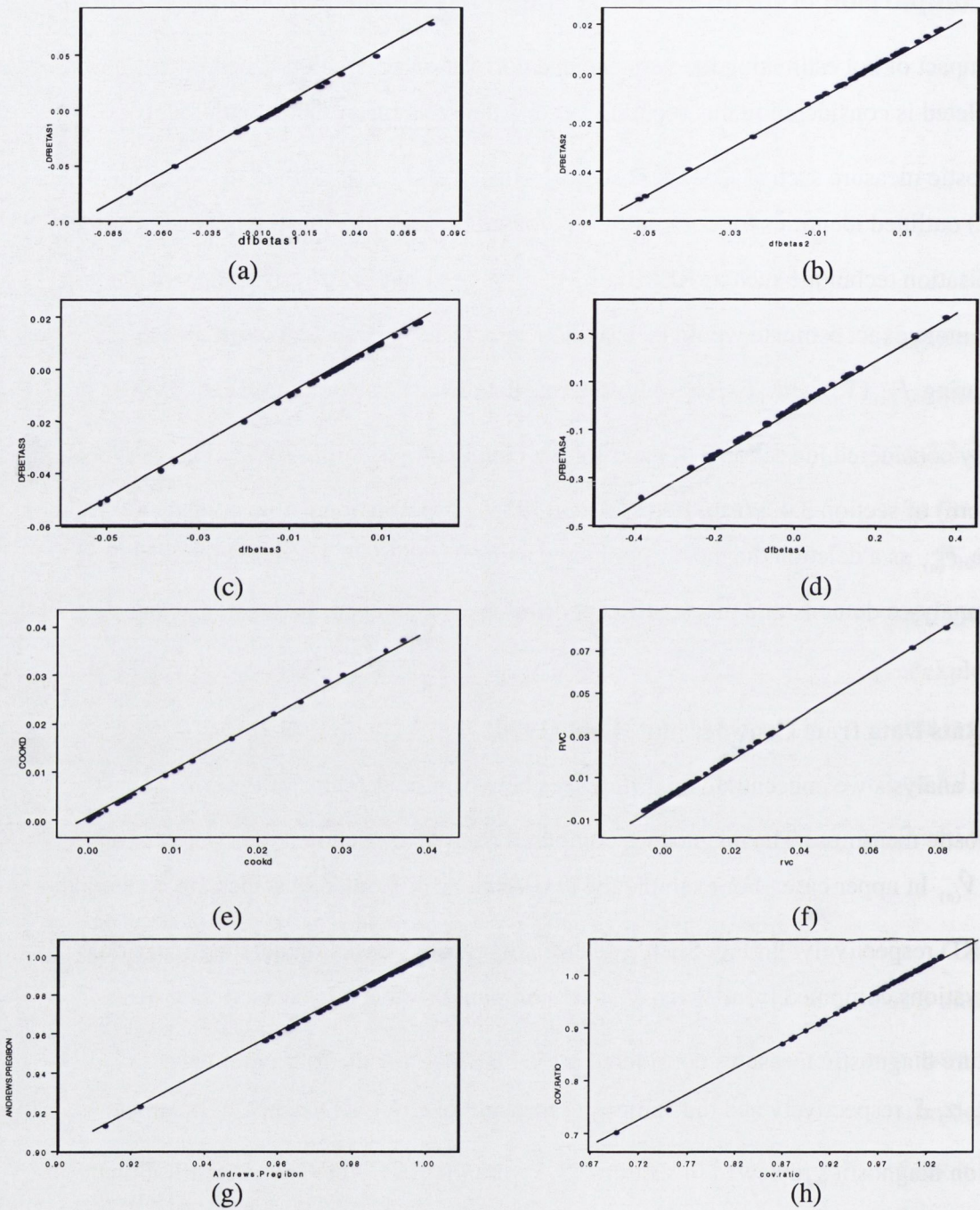
5.1 Comparison of mean structure diagnostics based on $\hat{V}$ and $\hat{V}_{(a)}$.

The impact of not estimating the variance and correlation parameters when observations are deleted is considered in this section. Let $F_{(a)}$ denote a mean structure deletion diagnostic measure such as Cook's Distance. Haslett and Hayes (1998) and Martin (1992) outlined identities for computing $F_{(a)}$ based on $\hat{V}$ with $\hat{V}$ estimated using some optimisation technique such as REML or ML. This estimate of $F_{(a)}$ is an approximation as the more exact estimate would be based on $\hat{V}_{(a)}$. The analysis below involves comparing $F_{(a)}(\hat{V})$ with $F_{(a)}(\hat{V}_{(a)})$ for the rats data from Crowder and Hand (1990) already considered in section 3.4.3 and for the county crop satellite survey data (soybean and corn) of section 3.4.2 from Batesse *et al*(1988). For notational convenience we define $F_{(a)}$ as a deletion diagnostic computed using $\hat{V}$ while $F_{((a))}$ is computed using $\hat{V}_{(a)}$. Both analyses demonstrate that significant differences may occur between $F_{(a)}$ and $F_{((a))}$.

5.1a Rats Data from Crowder and Hand (1990)

In this analysis we concentrate on differences between the versions of the various diagnostic measures. The diagnostics computed from $\hat{V}$ are denoted in lower case and from $\hat{V}_{(a)}$ in upper case. For example the two versions of Cook's Distance are cookd and COOKD respectively. In Fig. 5.1 the deletion diagnostics based on deleting individual observations computed from $\hat{V}$ and $\hat{V}_{(b)}$ are compared for the rats dataset. The mean structure diagnostic measures considered are DFBETAS for the four parameters $\alpha_1, \alpha_2, \alpha_3, \delta$ respectively and the composite measure of Cook's Distance. The other deletion diagnostics reviewed in section 2.3.2 were the RVC, the Covariance Ratio and the Andrews Pregibon statistic and these are also considered below though the focus is on the mean structure diagnostics. For this dataset the observation deletion diagnostic measures computed from $\hat{V}$ are a very good approximation for those based on $\hat{V}_{(a)}$ as is evident in 5.1 below.

Fig. 5.1 (a)-(h) Singleton deletion scatterplots of $F_{(a)}$ versus $F_{((a))}$ for rats data



Figs..5.1(a)-(h) Singleton deletion scatterplots of $F_{(a)}$ versus $F_{((a))}$ for various diagnostic measures.

Line $Y = X$ show equality of measures

The next element of the analysis involves comparing subset deletion diagnostics. For these data we consider the impact of deletion of each subject (a rat in this case) using the various diagnostic measures based on both $\hat{V}$ and $\hat{V}_{(a)}$. Figs. 5.1 i,j,k and l compare the DFBETAS statistic for the four mean parameters. It is clear from the graphs that there are appreciable differences between the two versions of the DFBETAS diagnostics particularly for the case of rat 12. This fact is also clearly evident in the case of Cook's Distance. Both the CVR and the AP statistic show quite dramatic differences between the two versions for the case of rat 12. As observed in section 3.2.3 Rat 12 is very influential in the estimation of σ_1 and thus its deletion has a large impact on the $\hat{V}$ matrix, which in turn has a substantial impact on the various diagnostic measures

Fig. 5.1 (i)-(p) Subset (rat) deletion scatterplots of $F_{(a)}$ versus $F_{((a))}$ for rats data

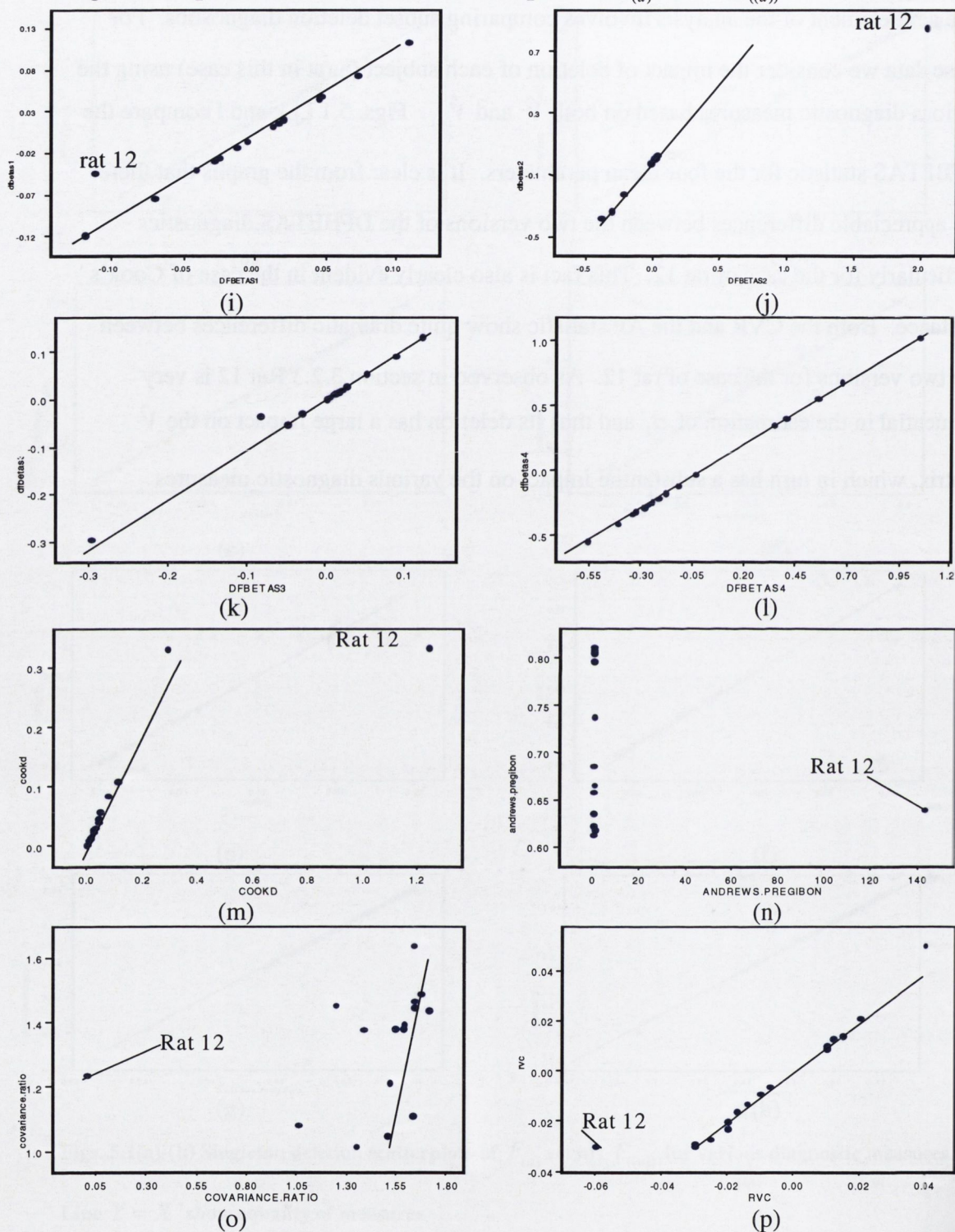


Fig.5.1(i)-(p) Case (rats) deletion scatterplots of $F_{(a)}$ versus $F_{((a))}$ for various diagnostic measures.

Line $Y = X$ show equality of measures

It is clear for such situations that diagnostics based on $\hat{V}$ may be misleading. Fig. 5.2 below illustrates this point.

Fig.5.2 Plots of Cook's Distance for each rat

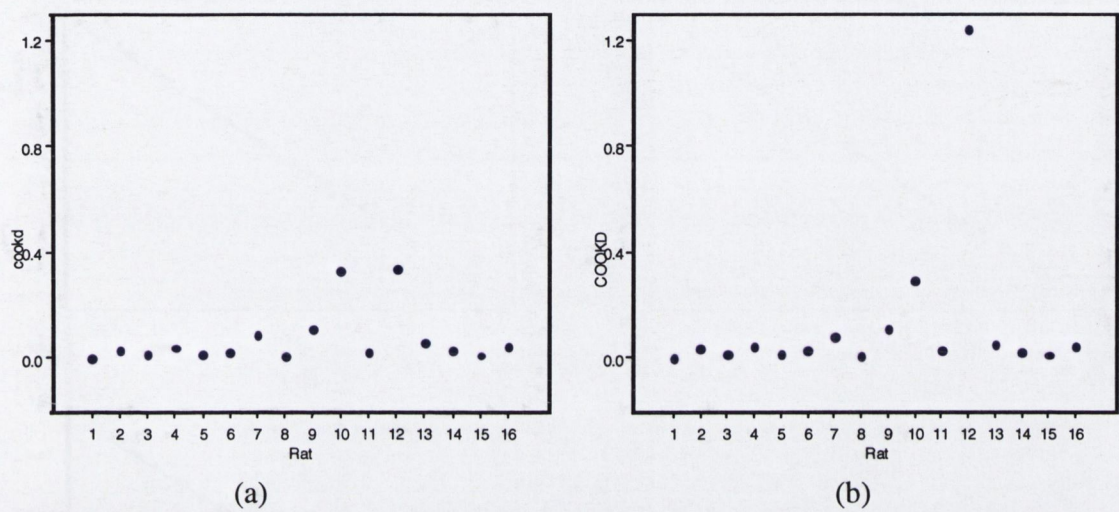


Fig.5.2(a) Plot of Cook's Distance for each rat based on $\hat{V}$ (b) Plot of Cook's Distance for each rat based on $\hat{V}_{(a)}$.

In Fig. 5.2 (a) where Cook's Distance is calculated without re-estimation of the covariance structure subsequent to deletion of the rats, both rats 10 and 12 are equally influential on the regression coefficients. In Fig.5.2 (b) with Cook's Distance estimated from $\hat{V}_{(a)}$ rat 12 is clearly most influential on the mean parameter estimates. The misleading effect of not computing the diagnostic measures using $\hat{V}_{(a)}$ is more serious in the case of the CVR and AP statistic as we saw in Fig. 5.1 and so the approach of computing the various diagnostic measures using $\hat{V}$ is not very satisfactory for these data. One approach would be to estimate $\hat{V}_{(a)}$ using REML or ML as above for each deletion, this is computationally burdensome and an alternative more computationally frugal approach would be desirable.

5.1b County Crop Areas data Batesse, Harter and Fuller. (1988)

For these data we again compare the same diagnostic measures based on $\hat{V}$ and $\hat{V}_{(a)}$. In Figs. 5.3a (a)-(f) the singleton deletion diagnostics for the corn data are compared while the corresponding diagnostics for the soybeans data are presented in Figs. 5.3a (g)-(l)

Fig.5.3a (a)-(f) Singleton deletion plots for corn data

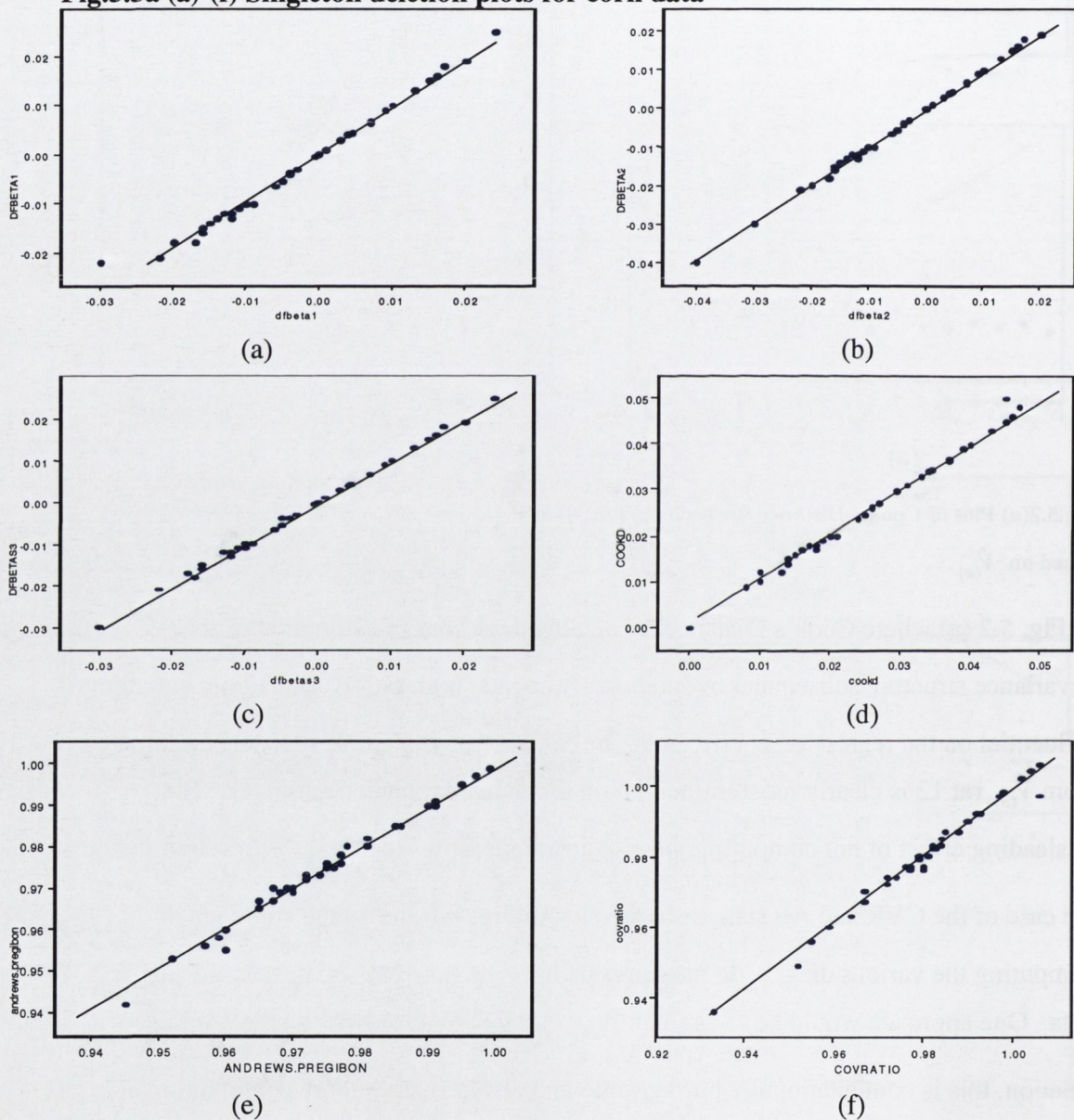


Fig.5.3a (a)-(f) Case deletion scatterplots of $F_{(a)}$ versus $F_{((a))}$ for various diagnostic measures for corn data. Line $Y = X$ shows equality of measures

Fig 5.3a (g)-(l) Singleton deletion plots for Soybeans data

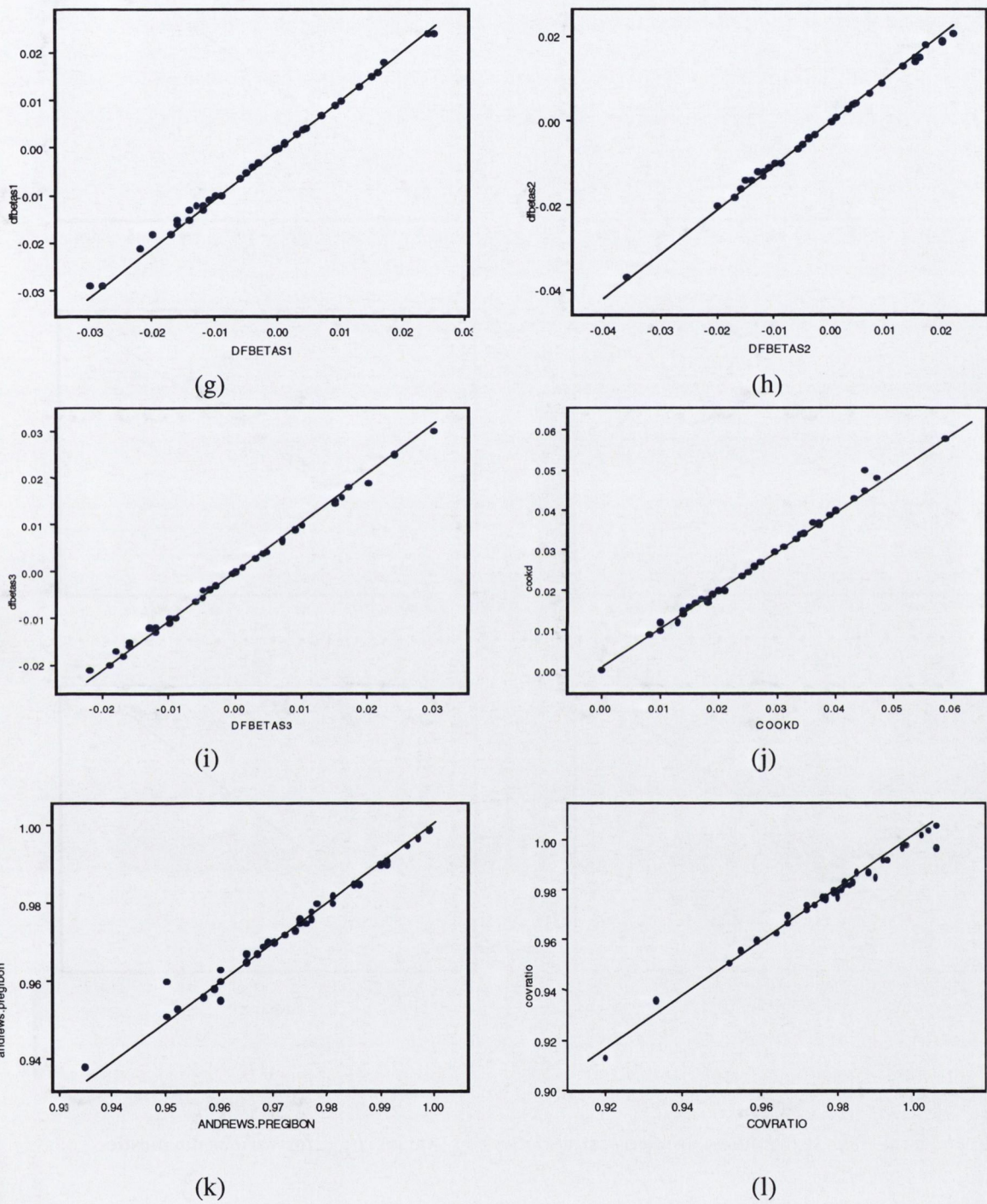


Fig.5.3a (g)-(l) Observation deletion scatterplots of $F_{(a)}$ versus $F_{((a))}$ for various diagnostic measures

for soybeans data. Line $Y = X$ show equality of measures

Fig5.3b(a)-(f) Subset (counties) deletion plots for corn data

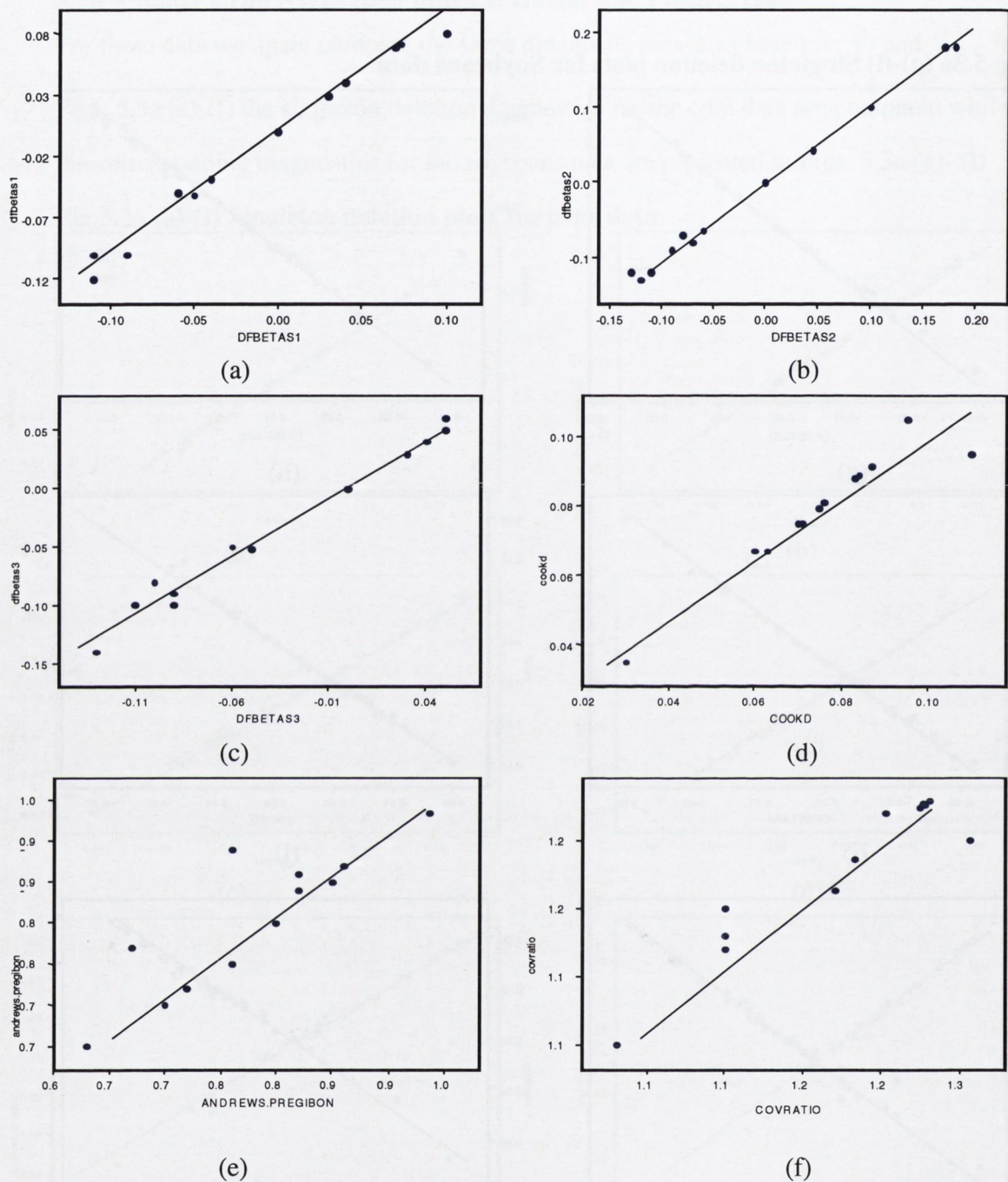


Fig.5.3b (a)-(f) Case (counties) deletion scatterplots of $F_{(a)}$ versus $F_{((a))}$ for various diagnostic measures for corn data. Line $Y = X$ show equality of measures

Fig. 5.3b (g)-(i) Case (County) deletion plots for soybeans data

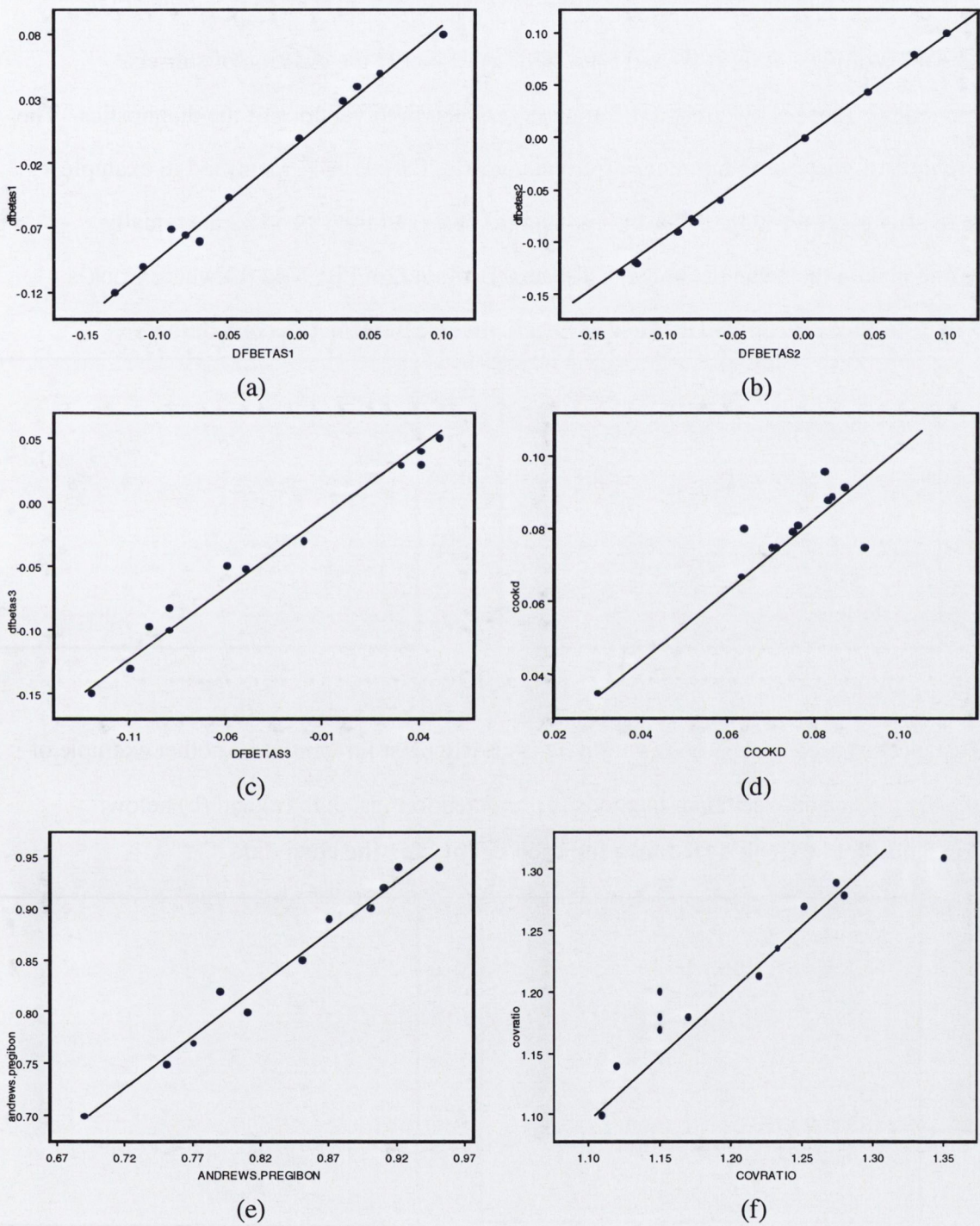
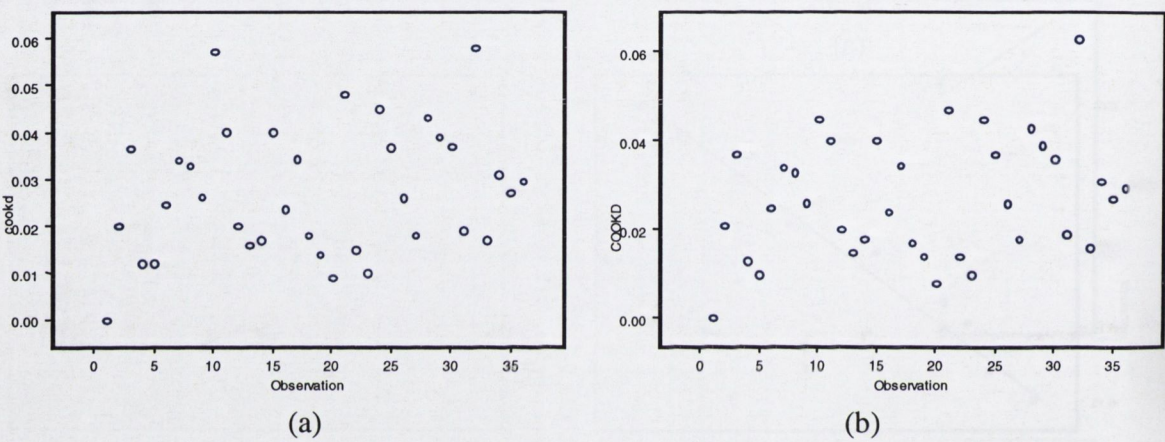


Fig.5.3b (g)-(i) Case (counties) deletion scatterplots of $F_{(a)}$ versus $F_{((a))}$ for various diagnostic measures for soybeans data. Line $Y = X$ show equality of measures

The mean structure diagnostics based on $\hat{V}$ provide a good approximation to those based on $\hat{V}_{(a)}$ for singleton deletion as is evident in Figs. 5.3a (a)-(f) and Figs. 5.3a(g)-(i).

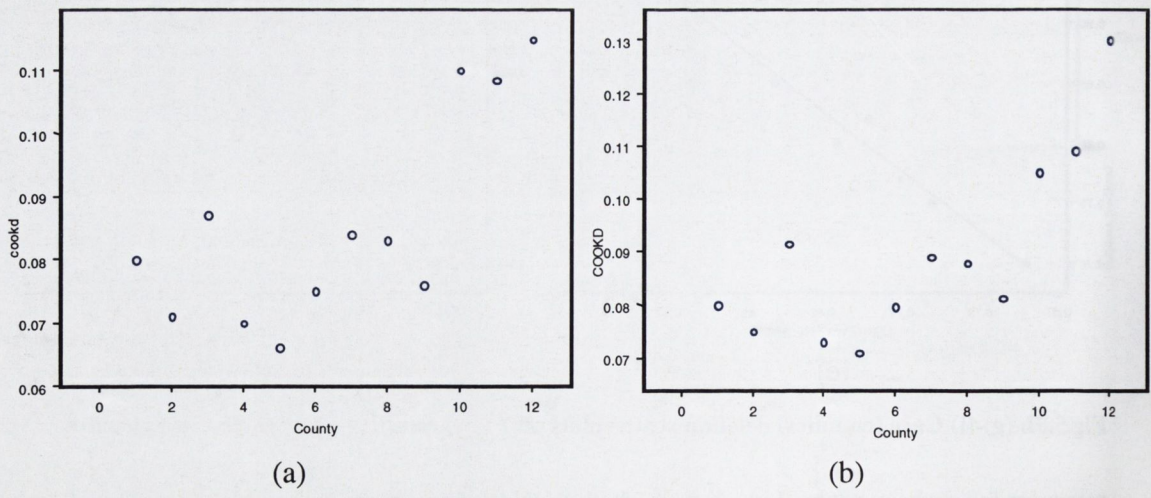
The plots of Figs. 5.3b(a)-(f) and Figs.5.3b (g) -(i) show the deletion of subsets (counties). These show greater differences between both versions of the diagnostics. The problem of misleading inferences from diagnostics based on $\hat{V}$ identified in example 5.1a also arises for these data. In Fig. 5.4a (a) observations 10 and 32 are equally influential on the mean parameter estimates. However, in Fig. 5.4a (b) where Cook's

Fig. 5.4a Plot of Cook's Distance for each observation for the corn data



Distance is based on $\hat{V}_{(a)}$ observation 32 is clearly most influential. Another example of the divergence between both methods is presented in Figs. 5.4b (a) and (b) below.

Fig.5.4b Plot of Cook's Distance for each county for the corn data



In Fig. 5.4b (a) counties 12, 10 and 11 appear to be the most influential on the mean parameters and in that order. However, in Fig. 5.4b (b) with Cook's Distance based on $\hat{V}_{(a)}$ county 12 is clearly most influential with county 11 more influential than county 10.

It is clear from the analysis above that misleading inferences may be drawn from mean structure diagnostics based on $\hat{V}$. In section 5.2 below a method is proposed to estimate the mean structure diagnostics using Taylor series approximations.

5.2 Approximation based on Taylor Series

In this section an intermediate method is proposed which while not based on $\hat{V}_{(a)}$ incorporates estimates of the changes in the variance and correlation parameters. A modified mean structure diagnostic measure $F_{(a)}^{Tayl}$ based on a Taylor expansion is proposed below for the LMM of equation 1.5 with V expressed as $\sigma_0^2 A(\rho) + \sum_{j=1}^r \sigma_j^2 Z_j Z_j^T$

$$F_{(a)}^{Tayl} = F_{(a)} + (\hat{\sigma}_0^2 - \hat{\sigma}_{0(a)}^2) \frac{dF_{(a)}}{d\sigma_0^2} + (\hat{\rho} - \hat{\rho}_{(a)}) \frac{dF_{(a)}}{d\rho} + \dots + (\hat{\sigma}_r^2 - \hat{\sigma}_{r(a)}^2) \frac{dF_{(a)}}{d\sigma_{r(a)}^2} \quad (5.1)$$

We need estimates for $(\hat{\rho} - \hat{\rho}_{(a)}, \hat{\sigma}_0^2 - \hat{\sigma}_{0(a)}^2, \hat{\sigma}_1^2 - \hat{\sigma}_{1(a)}^2, \dots, \hat{\sigma}_r^2 - \hat{\sigma}_{r(a)}^2)$ and for the derivative of $F_{(a)}$ with respect to the variance and autocorrelation parameters to compute $F_{(a)}^{Tayl}$. Methods to estimate the changes in the variance components and the autocorrelation parameters were outlined in chapters 3 and 4. The principal results for computing $\frac{d}{d\sigma_j^2}$ for the principal mean structure diagnostic measures are given below.

Appendix 2 details the derivation of these results.

$$\frac{d\tilde{e}_{(a)}}{d\sigma_j^2} = Q_{aa}^{-1} \left(Q \frac{dV}{d\sigma_j^2} Q \right)_{aa} \tilde{e}_{(a)} - Q_{aa}^{-1} (Q \frac{dV}{d\sigma_j^2} Q Y)_a \quad (5.2)$$

$$\frac{dDFBETA_a}{d\sigma_j^2} = -B \frac{dV}{d\sigma_j^2} Q \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix} + B \begin{pmatrix} \frac{d\tilde{e}_{(a)}}{d\sigma_j^2} \\ 0 \end{pmatrix} \quad (5.3)$$

$$\frac{dCook'sDistance}{d\sigma_j^2} = \frac{p\hat{\sigma}^2 \frac{d}{d\sigma_j^2} (\tilde{e}_{(a)}^T H_{AA} \tilde{e}_{(a)}) - \tilde{e}_{(a)}^T H_{AA} \tilde{e}_{(a)} \frac{d}{d\sigma_j^2} p\hat{\sigma}^2}{(p\hat{\sigma}^2)^2} \quad (5.4)$$

For this approach to be of interest $F_{(P)}^{Tayl}$, the $k \times 1$ matrix with elements $F_{(a_i)}^{Tayl}$ must be easily computable. Using results in chapters 3 and 4 the proposed approach produces an approximate computational procedure for $F_{(P)}^{Tayl}$ which is fast, cheap and results from the fitting process.

$$F_{(P)}^{Tayl} = F_{(P)} + \begin{pmatrix} \hat{\sigma}_0^2 - \hat{\sigma}_{0(a_1)}^2 \\ \hat{\sigma}_0^2 - \hat{\sigma}_{0(a_k)}^2 \end{pmatrix} * \begin{pmatrix} \frac{dF_{(a_1)}}{d\sigma_{(0)}^2} \\ \frac{dF_{(a_k)}}{d\sigma_{(0)}^2} \end{pmatrix} + \begin{pmatrix} \hat{\rho} - \hat{\rho}_{(a_1)} \\ \hat{\rho} - \hat{\rho}_{(a_k)} \end{pmatrix} * \begin{pmatrix} \frac{dF_{(a_1)}}{d\rho} \\ \frac{dF_{(a_k)}}{d\rho} \end{pmatrix} +, +, \begin{pmatrix} \hat{\sigma}_r^2 - \hat{\sigma}_{r(a_1)}^2 \\ \hat{\sigma}_r^2 - \hat{\sigma}_{r(a_k)}^2 \end{pmatrix} * \begin{pmatrix} \frac{dF_{(a_1)}}{d\sigma_{(r)}^2} \\ \frac{dF_{(a_k)}}{d\sigma_{(r)}^2} \end{pmatrix} \quad (5.5)$$

$$= F_{(P)} + M_{0(P)} D_{0(P)} + M_{1(P)} D_{1(P)} +, \dots, M_{r(P)} D_{r(P)}$$

where the elements of $M_{i(P)}$ are available from the columns of $\tilde{\sigma}_{(P)}$ in equation 3.13.

The elements of $D_{i(P)}$ are available through equations 5.6, 5.7 and 5.8 using the conditional residuals and associated results in appendix 1.

$$\frac{d\tilde{e}_{(P)}}{d\sigma_j^2} = D_P \left(-Q \frac{dV}{d\sigma_j^2} Q - \frac{dD_P}{d\sigma_j^2} \tilde{e}_{(P)} \right) \quad (5.6)$$

$$\frac{dDFBETA_p}{d\sigma_j^2} = -B \frac{dV}{d\sigma_j^2} QE + B \frac{dE}{d\sigma_j^2} \quad (5.7)$$

where E is an $N \times k$ matrix with the j th column of E containing $\tilde{e}_{(a_j)}$ and 0.

$$\frac{dCook'sDistance_p}{d\sigma_j^2} = \frac{p\hat{\sigma}^2 \frac{d}{d\sigma_j^2} (\tilde{e}_{(p)}^T GE) - \tilde{e}_{(a)}^T GE \frac{d}{d\sigma_j^2} p\hat{\sigma}^2}{(p\hat{\sigma}^2)^2} \quad (5.8)$$

$$\text{with } \frac{d}{d\sigma_j^2} (e_{(p)}^T GE) = \left(\frac{d\tilde{e}_{(p)}}{d\sigma_j^2} G + e_{(p)}^T \frac{dG}{d\sigma_j^2} \right) E + (e_{(p)}^T G \frac{dE}{d\sigma_j^2})$$

$$\text{and } \frac{d\hat{\sigma}}{d\sigma_j^2} = -Y^T Q \frac{dV}{d\sigma_j^2} QY / N - p$$

5.3 Evaluation of Proposals

The plots below compare the fully refitted diagnostic measures denoted in upper case and the proposed approximations based on the Taylor approximation for each rat in the example from Crowder and Hand (1993). In our analysis in section 5.1 differences between both versions of the mean structure diagnostics arose when cases were deleted. The Taylor Approximation provides a good estimate of the diagnostics based on $\hat{V}_{(a)}$ for these data. It is clear from the graphs below that the use of the Taylor expansion to provide a one-step approach for updating the mean structure diagnostics may give a better approximation for these diagnostics than those currently in the literature based on $\hat{V}$. Figs. 5.2 (a) and (b) and Fig. 5.1 (m) demonstrate the difference between Cook's Distance for rat 12 based on $\hat{V}$ and $\hat{V}_{(a)}$. From Fig. 5.5(e) it is clear that the proposed approach provides a considerably better estimate of Cook's Distance for rat 12 than that based on $\hat{V}$.

Fig.5.5 Scatterplots of $F_{(a)}^{Tayl}$ versus $F_{((a))}$ for rats data

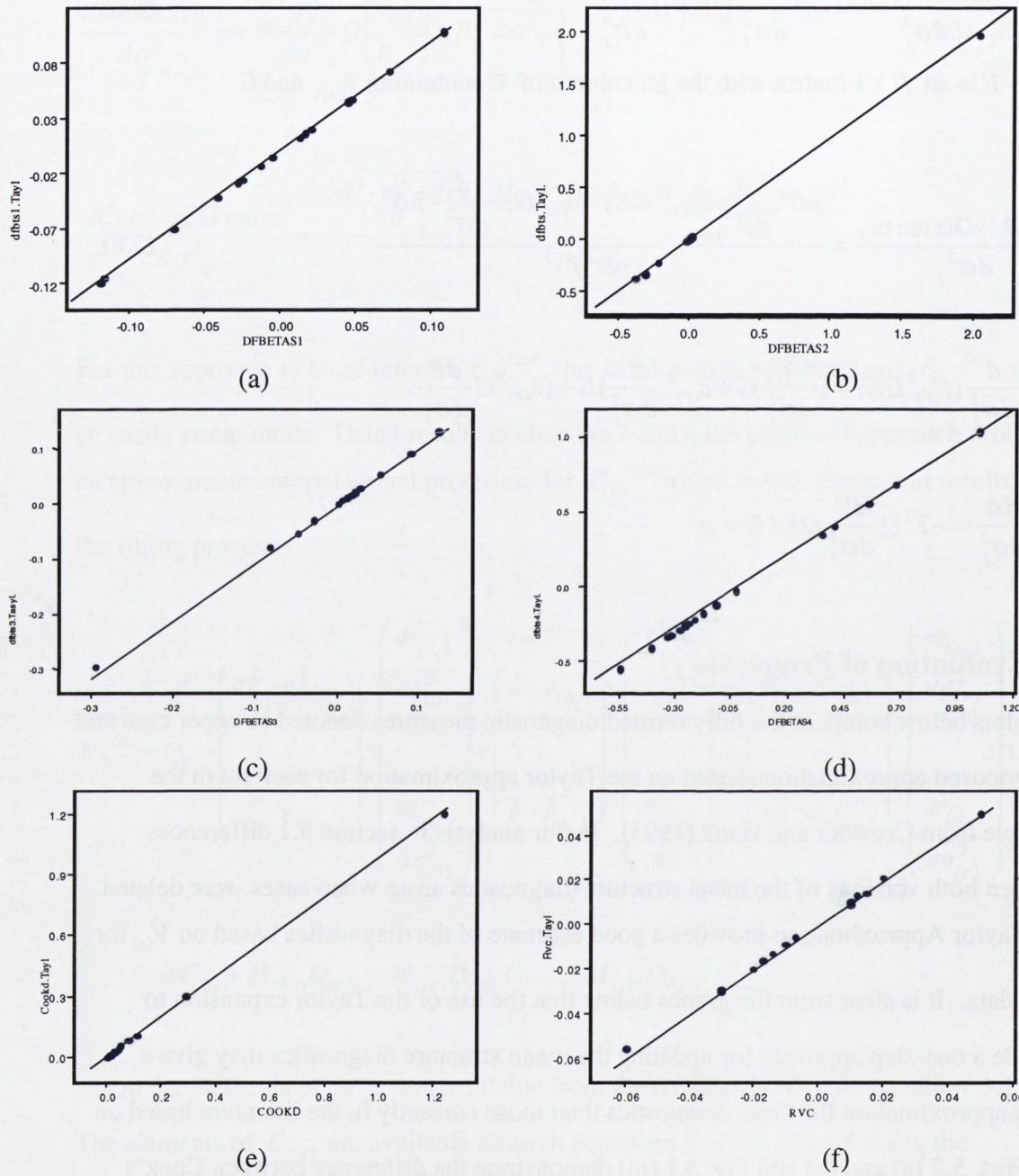


Fig.5.5 (a)-(f) Subset (rats) deletion scatterplots of $F_{(a)}^{Tayl}$ versus $F_{((a))}$ for various diagnostic measures for rats data. Line $Y = X$ shows equality of measures

For the corn and soybeans data above we again use the Taylor Approximation for the subset deletion. The results are presented in the plots below.

Fig. 5.6 (a)-(h) Scatterplots of $F_{(a)}^{Tayl}$ versus $F_{((a))}$ for corn and soybean data

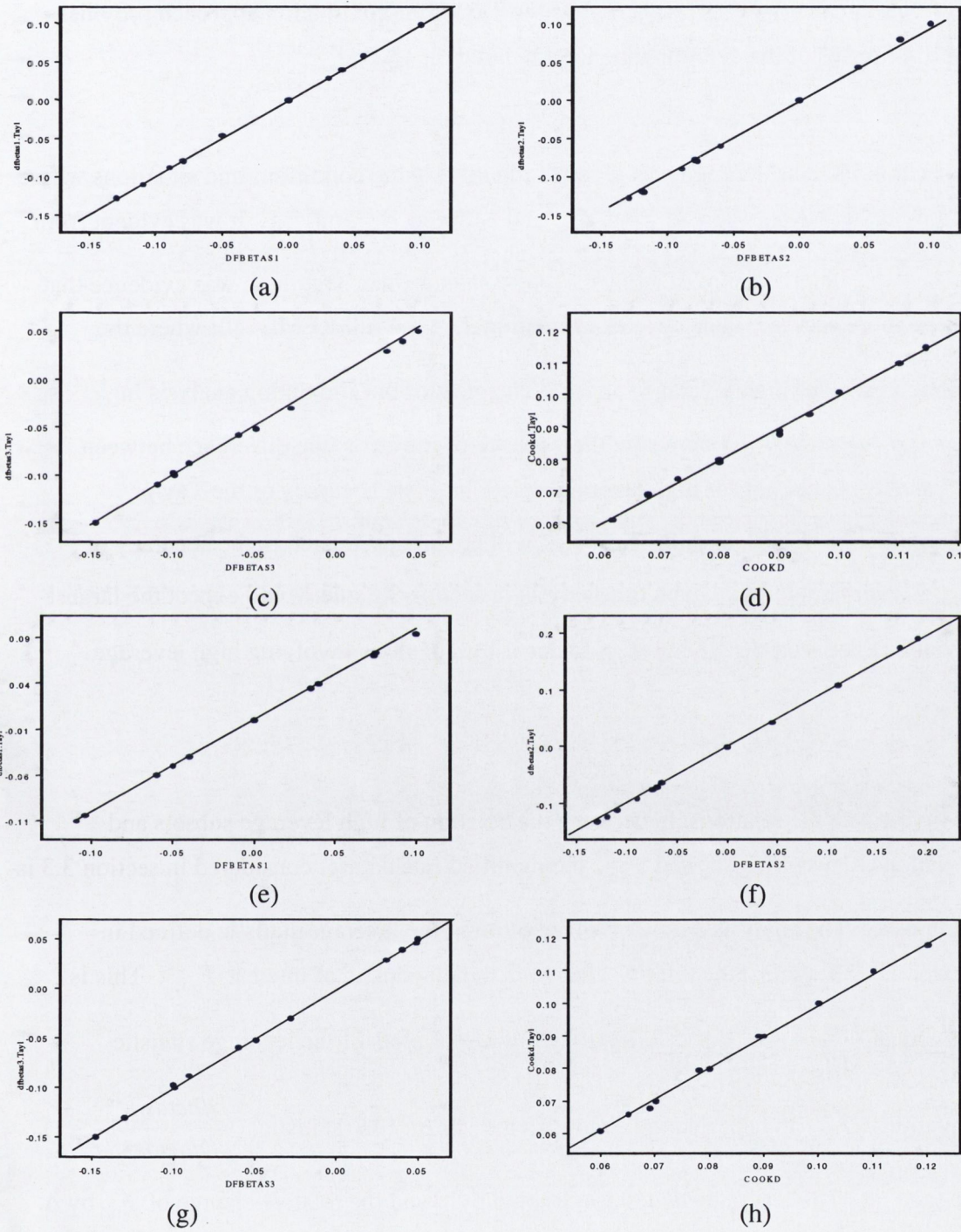


Fig.5.6 (a)-(d) Case (counties) deletion scatterplots of $F_{(a)}^{Tayl}$ versus $F_{((a))}$ for various diagnostic measures for corn data. Fig.5.6 (e)-(h) Corresponding scatterplots for soybeans data Line $Y = X$ show equality of measures

Figs. 5.6(a)-(h) show that for these data the Taylor approximation approach provides good estimates of the deletion diagnostics based on $\hat{V}_{(a)}$.

In section 3.3 consideration was given to identifying the conditions and situations where differences arose between $\hat{\sigma}_{(a)}$ and $\tilde{\sigma}_{(a)}$ and between $\tilde{\sigma}_{(a)}$ and $\hat{\tau}_{(a)}$. It was evident from the analysis that the approximations are very satisfactory. But there was evidence that $\hat{\sigma}_{(a)}$ is more accurate than $\tilde{\sigma}_{(a)}$ for estimating $\hat{\tau}_{(a)}$, the full REML refit where the deletion involves high leverage subsets. The plots of the simulation analysis in Figs. 3.7 (g) and 3.7 (h) show that the variance of the error (the difference between $\tilde{\sigma}_{(a)}$ and $\hat{\tau}_{(a)}$) is larger for high leverage deletions. The accuracy of the Taylor approximation $F_{(a)}^{Tayl}$ in estimating $F_{((a))}$ will be directly related to the accuracy of $\tilde{\sigma}_{(a)}$ in estimating $\hat{\tau}_{(a)}$. Given the analysis in section 3.3 one would expect the largest deviations between $F_{(a)}^{Tayl}$ and $F_{((a))}$ to occur for deletions involving high leverage subsets.

To investigate the relationship between the deletion of high leverage subsets and deviations between $F_{(a)}^{Tayl}$ and $F_{((a))}$ the modified rats dataset considered in section 3.3 is again used. The analysis below involved plotting the leverage statistic defined in equation 1.2.5 against the relative change in the diagnostic of interest $F_{((a))}$. This is defined as $\frac{F_{(a)}^{Tayl}}{F_{((a))}} - 1$. Fig. 5.7 (a)-(d) below are of plots of the leverage statistic

(equation 1.2.5) and the relative change in the $DFBETAS$ statistic $(\frac{Dfbetas_{(a)}^{Tayl}}{Dfbetas_{((a))}} - 1)$

for each rat. The leverage values are denoted by $\circ$ and the relative change of $F_{(a)}$ by Δ .

Fig. 5.7 Plots of leverage versus $\frac{F_{(a)}^{Tayl}}{F_{((a))}} - 1$ for the rats data

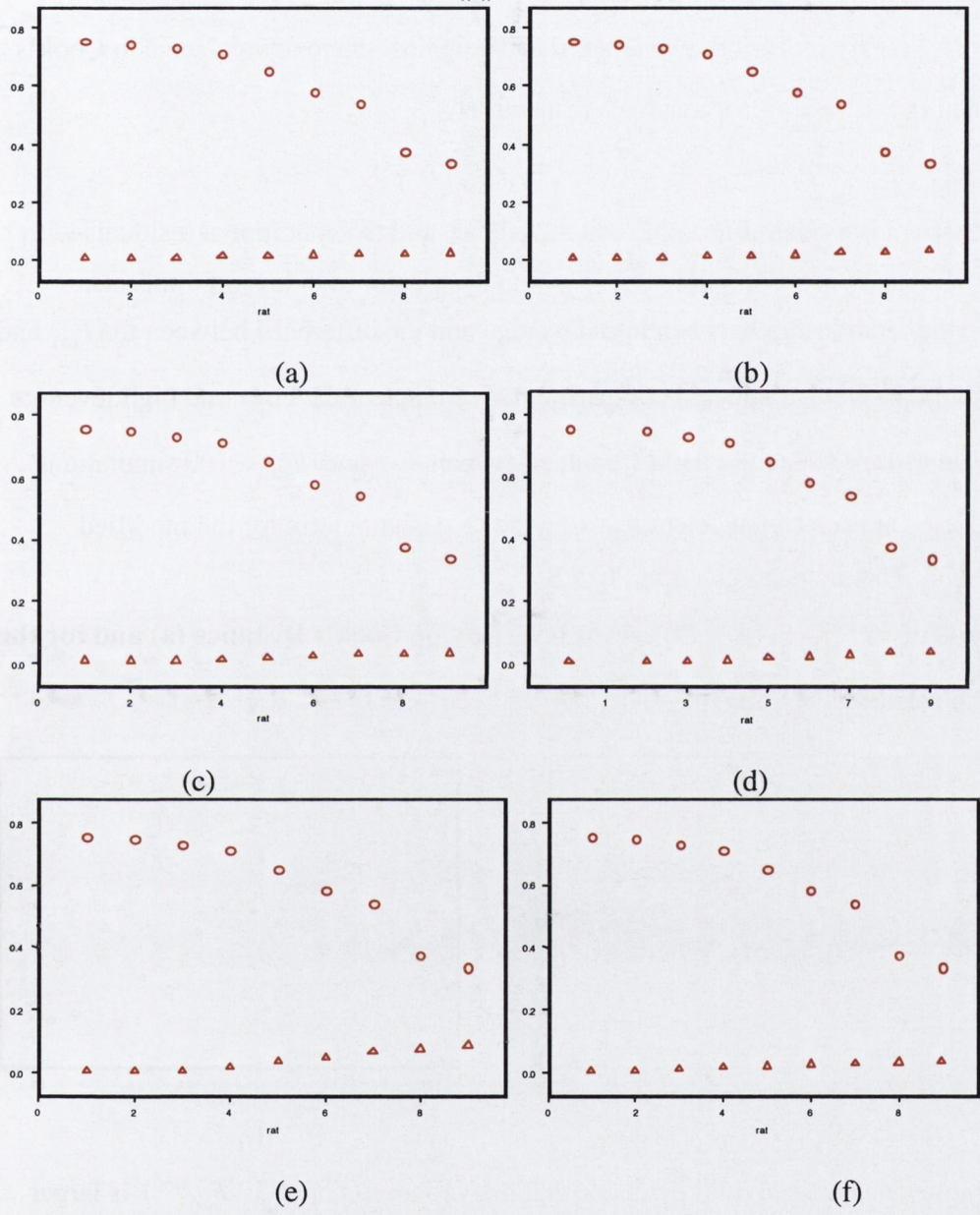


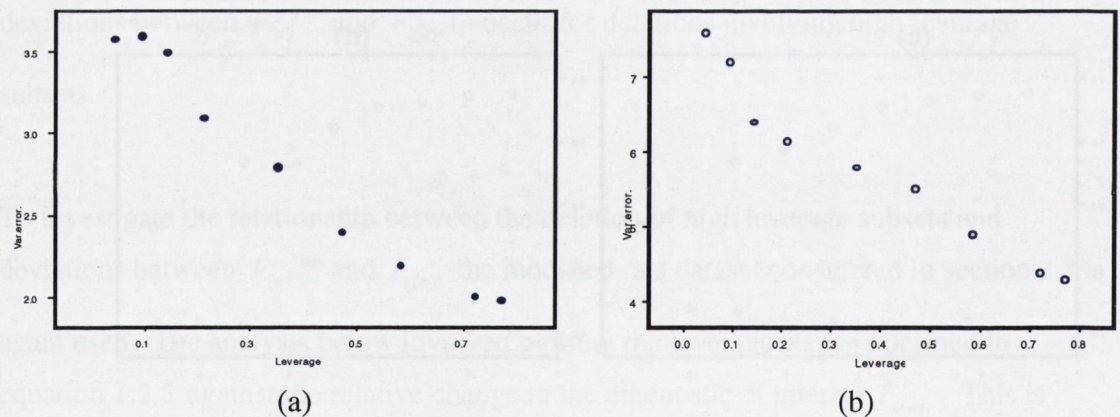
Fig.5.7 (a)-(d) Plots of $\frac{Dfbetas_{(a)}^{Tayl}}{Dfbetas_{((a))}} - 1$ denoted by $\triangle$ for each of the mean parameters $\alpha_1, \alpha_2, \alpha_3$

and β with the leverage statistic denoted by $\bigcirc$. Figs. 5.7 (e)-(f) Plots of leverage statistic denoted by $\triangle$ versus the relative change in Cook's Distance and in the conditional residuals respectively denoted by $\bigcirc$

There is evidence from the plots that deviations between $F_{(a)}^{Tayl}$ and $F_{((a))}$ are larger when the deletions involve high leverage subsets. However, even for such subsets the deviations are relatively small between $DFBETAS_{(a)}^{Tayl}$ and $DFBETAS_{(a)}(\hat{V}_{(a)})$. Figs. 5.7 (e) and 5.7 (f) are plots of the leverage statistic against the relative change in Cook's Distance and in the conditional residuals respectively.

The same pattern is evident for both Cook's Distance and the conditional residuals with the larger deviations occurring between $F_{(a)}^{Tayl}$ and $F_{((a))}$ for high leverage subsets. To explore the relationship between high leverage and the difference between the $F_{(a)}$ and $F_{((a))}$, simulations were conducted comparing the variance of the error and high leverage subsets. The error is defined as the difference between $F_{(a)}$ and $F_{((a))}$. 100 simulations were conducted at each of the deletions using the X design matrix for the modified dataset above. The plots are given in Fig.5.8

Fig.5.8 Plots of $Var(F_{((a))} - F_{(a)}^{Tayl})$ versus leverage for Cook's Distance (a) and for the conditional residuals $\tilde{e}_{(a)}$ in (b)



The simulation analysis provides evidence that the variance of $(F_{((a))} - F_{(a)}^{Tayl})$ is larger when the deletions involve high leverage subsets. It would be expected that differences between $F_{((a))}$ and $F_{(a)}^{Tayl}$ would be influenced by the differences between $\tilde{\sigma}_{(a)}$ and $\hat{\tau}_{(a)}$ from the definition of $F_{(a)}^{Tayl}$ in 5.1. The plots in Fig.5.8 (a) and (b) show a similar

relationship to that in Figs. 3.7(g) and (h). This analysis provides evidence that $F_{(a)}^{Tayl}$ is a very good approximation even for high leverage subsets.

5.4 Discussion

In this chapter the impact of assuming the variance structure is unchanged when observations are deleted on the mean structure diagnostic measures is explored. It is evident from the longitudinal data example in section 5.1 that in certain situations significant differences may occur between the statistics computed on the basis of $\hat{V}$ and $\hat{V}_{(a)}$. For longitudinal data the focus is often on the influence of the experimental unit or subject and it is for such situations that differences occur in section 5.1. One approach would be to completely refit the model for each deletion though this would be computationally burdensome and unlikely to be adopted by software developers.

In section 5.2 an approximation based on a Taylor expansion is outlined which exploits the developments in Chapters 3 and 4. It may be computed as a by-product of the fitting process and is computationally frugal. Based on the analysis in section 5.3, when the proposed approach is feasible it is a good approximation for diagnostic measures based on a full re-fit. The accuracy of the approximation is directly related to the accuracy of the methods proposed in Chapter 3.

Chapter 6

Conclusions and Discussion

6.0 Outline

This chapter provides a review of the thesis. Its principal achievements are outlined in section 6.1. The focus of the thesis is deletion diagnostics for the LMM. In chapter 2, the various diagnostics from the literature were reviewed in the context of the LMM. Deficiencies and inadequacies in these methods were identified. The issues and challenges arising in chapter 2 are the motivation for the proposed methods described in chapters 3,4 and 5. These proposals are reviewed in 6.1 and their merits outlined.

While the proposals address many of the challenges posed by the review in chapter 2 there are still outstanding issues. Section 6.2 considers possible inadequacies in the proposed methods. Shortcomings and disadvantages of the proposed methods are clearly enunciated. These pose new challenges and further opportunities for research are outlined in this section. These focus on addressing the shortcomings in the proposals in this thesis and for possible adaptation of these methods to other models.

6.1 Thesis Summary and Achievements

The LMM is a very important member of the LM family of models. It has a very versatile structure, allowing it to be applied in many situations. A major subset of this class of models is the LM with only one source of random variation; the variance covariance matrix is of the form $\sigma^2 V$ with γ_j null for $j > 0$. In the classical case $Var(\varepsilon) = \sigma^2 I$ and OLS applies. Deletion diagnostics were originally developed for this model. They have since been adapted for the more general single source of variation model described above. However, for the LMM with more than one component of variation various challenges and issues arise.

Influence arises at two stages in these models. These are in the estimation of V by some $\hat{V}$ and at the subsequent estimation of β and γ , given $\hat{V}$. Much of the work on deletion diagnostics for the LMM has focused on the second stage. This work has been based on the assumption that $\hat{V}$ is fixed. Intuitively, for the single component of variance case with variance covariance matrix $V = \sigma^2 R(\rho)$, the assumption is unlikely to be problematic. Indeed our work in chapter 4 would support this view. However, from section 5.1 it is clear that for the multiple variance component case, basing the computation of deletion diagnostics on the assumption of $\hat{V}$ being fixed is unsatisfactory. The clear demonstration of this issue in section 5.1 motivates much of the developments in this thesis

Chapter 3 focuses on assessing influence at the first stage of the fitting process, the estimation of V by some $\hat{V}$. Christensen *et al* (1992) emphasized the importance of conducting an influence analysis at this stage and provided a means of doing this. The principal criticism of their approach was that it was computationally burdensome (Oman, 1995). An alternative, computationally efficient approach based on the “delete = replace”

identity is proposed (Haslett and Dillane, 2004). The identity is also exploited to enhance the computational feasibility of the CPJ method and to offer an ANOVA based approach based on the Henderson's Method III approach.

Our evaluation of these proposals has shown that it is simple to generate as by-products of the fitting procedures approximations for assessing the influence of subsets of data, as measured by deletion, on the estimates of variance components in the LMM. These approximations are useful for singletons or for larger subsets of the data. The key to these developments is the "delete = replace" identity. It is intrinsically linked with the conditional residual, which is central to the easy computation of the proposed diagnostic measures.

In many studies arising in the analysis of longitudinal, time series and spatial data, there can be, in addition to covariance induced by the random effects, significant autocorrelation. For the LMM defined in 1.5 A would have the form $\sigma^2 R(\rho)$. In many applications the autocorrelation is modeled by an AR(1) or MA(1) process. In chapter 4 methods are proposed to provide estimates for $\rho_{(a)}$. The proposed method (Haslett and Dillane, 2004) is adapted to provide estimates of the MA parameters when singletons or larger subsets are deleted. For AR models the proposals are a by-product of the method of moments fitting procedure and are useful for single observations and for subsets of the data. Again the "delete = replace" identity and the conditional residual are key to the feasibility of these approaches.

Influence analysis for the variance components and the autocorrelation parameters is only a preliminary stage to influence analysis for the regression parameters. The various approaches in the literature for estimating deletion diagnostics for the fixed effects are based on $\hat{V}$ as apposed to the more exact $\hat{V}_{(a)}$. The analysis in section 5.1 showed clearly

that significant differences may arise between deletion diagnostics based on $\hat{V}$ and $\hat{V}_{(a)}$.

In chapter 5, a method based on a Taylor series approximation is proposed which allows computation of deletion diagnostics based on $\hat{V}_{(a)}$. The method exploits the proposed methodologies in chapters 3 and 4. The approach is computationally frugal and is a by-product of the modeling fitting process. Our evaluation of the proposal demonstrates that $F_{(a)}^{Tayl}$ provides a good approximation for $F_{((a))}$.

6.2 Deficiencies of thesis and further research opportunities.

The methods proposed in this thesis have a number of shortcomings and leave a number of issues unresolved. These are outlined below.

The conditional residual may be either studentised internally or externally. In the case of external studentisation, the conditional residual is divided by $\sigma_{(a)}^2(Q_{AA})^{-1}$. For the single component of variance case this is easily computable. However, for the LMM with more than one variance component, external studentisation is problematic. In this case each of the variance components should be updated. While the methods in chapter 3 provide a computationally inexpensive method of estimating the impact on the variance components of deleting a subset of data, there is no 1-step procedure for externally studentising the conditional residuals. Externally studentised conditional residuals are thus not computationally feasible.

The Covariance Ratio and Andrews-Pregibon statistics are significantly more difficult to compute for the LMM with more than one variance component. The issue is similar to that encountered with external studentisation. The updating formulae for both measures outlined in Haslett and Hayes (1998) and Martin (1992) are not appropriate for the multiple variance component situation. There is no easily computable estimate of $V_{(a)}$

unless it is re-estimated. The methods in chapters 3 and 4 offer an intermediate approach for estimating $V_{(a)}$ but it is a two-stage approach and not very computationally frugal.

The methods proposed in chapter 5 are not applicable to these two diagnostic measures either. The Taylor series involved in this approach requires the derivative of the statistic of interest with respect to a variance or autocorrelation parameter. These measures involve the determinants of a matrix. This is not easily computable as the derivative of the determinant of a matrix is not equal to the determinant of its derivative. Thus the methods in chapter 5 are not applicable to these diagnostic measures.

The methods proposed in chapter 4 for updating the autocorrelation parameters following deletion are limited. They are applicable only to AR(p) and MA(q) models. For AR models the approach is computationally expensive for any models of higher order than AR(1). However, for many applications to time series or longitudinal data, AR(1) models are frequently used. The method is method of moments based and while this approach is reasonably accurate for AR models, a REML or MI based approach would be more desirable.

These issues pose challenges and offer opportunities for further research. The focus of this thesis has been on the LMM with normally distributed errors. However, the ideas and methods may be adapted for other models such as generalized linear mixed models. Diagnostic measures for such models are not well developed and it is likely that cheaply computable deletion diagnostics would generate significant interest. The methods developed in this thesis may be adaptable for this purpose.

References

- Analytical Methods Committee (1987) Recommendations for the conduct and interpretation of co-operative trials, *The Analyst*, **112**, 679-686
- Baade, I.A..and Pettitt, A.N. (2000) Multiple and conditional deletion diagnostics for general linear models, *Communications in statistics –Theory and Methods*. **29**(8): 1899-1910
- Banerjee, M and Frees, E.W. (1997) Influence Diagnostics for Linear Longitudinal Models, *Journal of the American Statistical Association*, **92**, 999-1005.
- Banerjee, M. (1998), Cook"s Distance in linear longitudinal models, *Communications in Statistics, Part A -- Theory and Methods*, **27**, 2973-2983
- Battese, G.E. Harter, R.M. and Fuller, M.A. (1988) An error components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association* **83** 28-36
- Beckman, R.J., Nachtsheim, C.J. and Cook. R.D. (1987) Diagnostics for Mixed-Model Analysis of Variance. *Technometrics*. **29**, 413-426
- Belsley, D., Kuh, E. and Welsch, R. (1980). *Regression Diagnostics :Identifying influential data and sources of collinearity*, New York: Wiley.
- Box, G. and Jenkins, G. (1976) *Time Series Analysis: Forecasting and Control* (Revised Edition), San Francisco: Holdenday.
- Bryk, A.S. Raudenbush, S.W. (1992) *Advanced Qualitative Techniques in the Social Sciences No 1: Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publications Inc. (USA)

- Bruce, A.G. and Martin, R.D. (1989) Leave-k-out diagnostics for time series (with discussion). *Journal of the Royal Statistical Society, Ser. B* **51** 363-424.
- Chatterjee, S, and Hadi, A. S. (1986) Influential observations, high leverage points, and outliers in regression. (with discussion). *Statistical Science*, **1** 379-416.
- Chatterjee, S, and Hadi, A. S. (1988) *Sensitivity Analysis in Linear Regression* Wiley.
- Christensen R, (1987) *Plane Answers to Complex Questions*, Springer
- Christensen, R, Pearson, L.M. and Johnson, W. (1992) Case-Deletion Diagnostics for Mixed Models, *Technometrics*, **34**, 38-45.
- Cochran, W.G. (1939) The use of analysis of variance in enumeration by sampling. *Journal of the American Statistical Association*. **34**, 492-510.
- Cook, R.D. (1977) Detection of influential observations in linear regression, *Technometrics*, **19**, 15-18
- Cook, R.D. (1986), Assessment of local influence, *Journal of the Royal Statistical Society, Series B, Methodological*, **48**, 133-169
- Cook, R.D. and Weisberg, S. (1982) *Residuals and Influence in regression*. New York: Chapman and Hall.
- Cox, D. R. and Vermuth, N. (1998) *Multivariate dependencies: models, analysis and interpretation*. Chapman and Hall.
- Crowder, M and Hand, D. (1990) *Analysis of Repeated Measures*. Chapman and Hall, London.
- Cryer, J. D. (1985) *Time Series Analysis*, PWS Publishers
- Demidenko, E. (2004) *Mixed Models Theory and Applications*, Wiley

- Demidenko, E. and Stukel, T.A. (2005) Influence analysis for linear mixed-effects models *Statistics in Medicine*, **6** 893-909
- Dempster, A. P., Rubin, D. B. and Tsutakawa, R. K. (1981). Estimation in covariance component models. *Journal of the American Statistical Association*, **76**, 341-353
- Diggle, P., (1990) *Time Series: A Biostatistical Introduction* Clarendon Press, Oxford.
- Diggle, P., Liang, K-Y and Zeger, SL (1994) *Analysis of Longitudinal Data*, Oxford Univ. Press
- Draper, N. R. and John, J.A..(1981) Influential observations and outliers in regression. *Technometrics* **23**, 21-26
- Goldstein, H. (1987) *Multilevel Statistical Models*, Arnold, London
- Hand, D and Crowder, M. (1996) *Practical Longitudinal Data Analysis*, Chapman and Hall.
- Haslett, J (1999) A Simple Derivation of Deletion Diagnostic Results for the General Linear Model with Correlated Errors, *Journal of the Royal Statistical Society, B*, **61**, 603-609.
- Haslett, J. and Dillane, D., (2004) Application of “delete = replace” to deletion diagnostics for variance component estimation in the linear model, *Journal of the Royal Statistical Society, Ser. B*, **66** 131-143.
- Haslett, J and Hayes, K (1998) Residuals for the Linear Model with General Covariance Structure, *Journal of the Royal Statistical Society (B)*, **60**, 201-215.
- Haslett, S.J. and Haslett, J. (2005) On the residuals for the Linear Model. *Technical Report. Department of Statistics TCD*
- Henderson, C.R. (1953) Estimation of variance and covariance components. *Biometrics*, **9**, 226-252

- Hidreth, C. and Lu, J. Y. (1960) Demand relations with autocorrelated disturbances
Technical Bulletin 276, Department of Agricultural economics, Michigan State University
- Hodges, J. S. (1998), Some algebra and geometry for hierarchical models, applied to diagnostics (Disc: p521-536), *Journal of the Royal Statistical Society, Series B, Methodological*, **60** , 497-521
- Kadiyala, k. R. (1970) Testing for the independence of regression disturbances.
Econometrika 38, 97-117
- Kianaford, F. and Swallow, W. H., (1996) A review of the development and application of recursive residuals in linear models. *Journal of the American Statistical Association*, **91**: 391-400
- Laird, NM and Ware, JH (1982) Random-effects models for longitudinal data,
Biometrics, **38**, 963-974
- Lange, N , and Ryan, L. (1989), Assessing normality in random effects models, *The Annals of Statistics*, **17** , 624-642
- Lawrence, A. J. (1995) Deletion, influence and masking in regression. *Journal of the Royal Statistical Society, Ser. B*, **57**, 181-189.
- Lesaffre, E , and Verbeke, G. (1998), `Local influence in linear mixed models,
Biometrics, **54** , 570-582
- Lesaffre, E , Asefa, M , and Verbeke, G (1999), Assessing the goodness-of-fit of the Laird and Ware model -- An example: The Jimma Infant Survival Differential Longitudinal Study, *Statistics in Medicine*, 18 , 835-854
- Lindley, D.V. and Smith A.M.F. (1972) Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Ser. B*, **34**, 1-18.

Lindsey, J. K. (1993) *Models for Repeated Measurements*, Clarendon Press, Oxford.

Longford, N.T. (1987) A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, **74**, 817-827.

Longford, N.T. (1993), *Random Coefficient Models*, Oxford Univ Press (Oxford)

Martin, R.J. (1992) Leverage, Influence and Residuals in Regression Models when Observations are Correlated, *Communications in Statistics - Theory and Methods*, **21**, 1183-1212.

Mason W.M., Wong, G. W., and Entwisle, B. (1984). Contextual analysis through the multi-level linear model. *Sociological Methodology*, Jossey Bass, 72-103

McCulloch, C.E. and Nelder J.A. (1989). *Generalised linear models*, 2nd edition. New York: Chapman and Hall.

McCulloch, C.E. and Searle, S.R. (2001) *Generalized, Linear and Mixed Models*, Wiley, NY

Nelder J.A. and Wedderburn R.W.M. (1972). Generalised Linear Models. *Journal of the Royal Statistical Society*, Ser. A, **135**, 370-384.

Oman, S. D. (1995), Checking the assumptions in mixed-model analysis of variance: A residual analysis approach, *Computational Statistics and Data Analysis*, **20** , 309-330

Ouwens, Mario J. N. M. , Tan, Frans E. S. , and Berger, Martijn P. F. (2001), Local influence to detect influential data structures for generalized linear mixed models, *Biometrics*, **57** (4) , 1166-1172

- Pan, J.X. and Fang, K.T. (2002) *Growth Curve Models and Statistical Diagnostics*. New York: Springer-Verlag.
- Patterson, H.D. and Thomposon, R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 454-554
- Picford, M., and Haslett, S., (1999) A statistical test of single firm market power, *New Zealand Economic Papers*. **33**, 39-58.
- Pregibon, D. (1981) Logistic Regression Diagnostics. *Annals of Statistics* **4**, 705-724.
- Rao, Poduri S. R. S.(Ed) (1997) *Variance Components Estimation : Mixed Models, Methodologies and Applications*, Chapman and Hall
- Schall, R. and Dunne, T.T. (1988) A unified approach to outliers in the General Linear Model, *Sankya* **50**, Seris B 157-167
- Searle, S.R. Cassella, G and McCullagh, CE (1992) *Variance Components*, Wiley
- Tan, F. E. S., Ouwens, Mario J. N. , and Berger, Martijn P. F. (2001), Detection of influential observations in longitudinal mixed effects regression models, *The Statistician*, **50** (3) , 271-284
- Theil, H. (1965) The analysis of disturbances in linear regression, *Journal of the American Statistical Association*, **60** 1067-1079.
- Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data*, New York: Springer-Verlag.
- Venables, W.N. and Ripley B.D. (1994) *Modern Applied Statistics with S-PLUS*, 2nd ed. Springer

- Wagner, J R, and Thaggard, N A. (1979) Gas-Liquid Chromatographic Determination of Nicotine Contained on Cambridge Filter Pads, *Journal of the Association of Official Analytical Chemists*, 62 229-236
- Waternaux, C. Laird, N.M. and Ware, J.H. (1989) methods for analysis of longitudinal data. *Journal of the American Statistical Association* **84** 33-41
- Weisberg, S. (1985) *Applied Linear Regression*. Wiley
- Weiss, R. E. and Lazaro, C. G. (1992) Residual Plots for repeated measures *Statistics in Medicine*, **11** 115-124
- Whittaker, J. (1990) *Graphical models in applied multivariate statistics*. Wiley
- Winsor, C.P. and Clarke, G.I. (1940) Statistical study of variation in plankton nets. *Sears Foundation Journal of marine research*. **3**, 1-34.

Appendix 1 Derivation of results used to compute proposed diagnostics of chapters three and four.

A1.1 Change in Sum of Squares Terms

The conditional residual $\tilde{e}_{(a)} = Y_a - \tilde{Y}_a(Y_b)$ is available via

$$\begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix} = \begin{pmatrix} D_a & 0 \\ 0 & 0 \end{pmatrix} QY \equiv \bar{D}_a QY, \text{ where } D_a = Q_{aa}^{-1} = \text{Var}(\tilde{e}_{(a)}), \text{ defining } \bar{D}_a \text{ implicitly.}$$

The key identity is $Y_{\tilde{a}} = (I - \bar{D}_a Q)Y \equiv W_a Y$ (implicitly defining W_a).

The sum-of-squares arising in the LM can be written as $\hat{e}^T V^{-1} \hat{e} = Y^T QY$. In the reduced LM we write $Y_b^T Q_{(a)} Y_b$, using the notation $Q_{(a)}$ to refer to the corresponding block of

the reduced matrix $\begin{pmatrix} V_{bb} & X_b \\ X_b^T & 0 \end{pmatrix}^{-1}$. A consequence of the ‘delete = replace’ identity is that

$$Y_b^T Q_{(a)} Y_b = Y_{\tilde{a}}^T QY_{\tilde{a}}. \text{ It is easy to show by direct expansion that (i) } \bar{D}_a Q \bar{D}_a = \bar{D}_a \text{ (ii)}$$

$$Q \bar{D}_a Q = Q - \bar{Q}_{(a)} \text{ where we define } \bar{Q}_{(a)} = \begin{pmatrix} 0 & 0 \\ 0 & Q_{(a)} \end{pmatrix} \text{ and (iii), since } X^T Q = 0, \text{ for any}$$

matrix A we have

$$E[Y_{\tilde{a}}^T Q A Q Y_{\tilde{a}}] = E[Y^T Q A Q Y] - E\left[\begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix}^T Q A Q \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix}\right] \quad (\text{A.1})$$

When $A = V$, A.I is an identity for the arguments; that is $Y_{\tilde{a}}^T Q Y_{\tilde{a}} = Y^T QY - \tilde{e}_{(a)}^T D_a^{-1} \tilde{e}_{(a)}$.

When $A = ZZ^T$ the identity in A.1 results

$$Y^T QZZ^T QY - Y_{\tilde{a}}^T QZZ^T QY_{\tilde{a}} = 2Y^T QZZ^T \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix} - \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix}^T QZZ^T \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix}$$

A1.2 Computation of trace terms based on ‘delete = replace’ results.

The “delete = replace” approach also provides updates for trace terms which are useful for the different methods considered in chapter 3.

From above $Y_{\bar{a}} = (I - \bar{D}Q)Y = W_a Y$ with $\bar{D} = \begin{pmatrix} D_a & 0 \\ 0 & 0 \end{pmatrix}$ where $D_a = Q_{aa}^{-1}$.

If $Y_{\bar{a}}' A Y_{\bar{a}} = Y_b' A_{(a)} Y_b$ and given that $\text{Var}(Y_{\bar{a}}) = (I - \bar{D}Q)V(I - \bar{D}Q)^T$

then

$$\begin{aligned} E(Y_{\bar{a}}' A Y_{\bar{a}}) &= \text{tr}(A(\text{Var}(Y_{\bar{a}}))) \\ &= \text{tr}(AV - 2AVQ\bar{D} + A\bar{D}) \\ &= \text{tr}(A_{(a)}V_{(a)}) = \text{tr}(AV - 2AVQ\bar{D} + A\bar{D}) \end{aligned}$$

If A is symmetric and of the form QBQ then some simplification yields the identity

$$\text{tr}(A_{(a)}V_{(a)}) = \text{tr}(AV - A\bar{D}) = \text{tr}(AV - A\bar{D})$$

Therefore the reduction in the trace of AV from deleting a subset is $\text{tr}(A\bar{D})$. In many applications in the thesis A is of the form $QZZ^T Q$.

Appendix 2: Derivation of results used in Taylor expansion approach outlined in Chapter 5.

A.2 Differentiation Results

Given the model $Y = X\beta + \varepsilon$ with $\text{Var}(Y) = V$ and V is modeled with parameters $\alpha_1, \alpha_2, \dots, \alpha_q$ the proofs below outline the derivation of $\frac{dT}{d\alpha_j}$ where T is the function of interest and $j = 1, \dots, q$

A.2.1 $\frac{dV^{-1}}{d\alpha_j}$

Given that

$$VV^{-1} = I$$

then

$$V \frac{dV^{-1}}{d\alpha_j} + \frac{dV}{d\alpha_j} V^{-1} = 0$$

and thus

$$\frac{dV^{-1}}{d\alpha_j} = -V^{-1} \frac{dV}{d\alpha_j} V^{-1}$$

A.2.2 $\frac{d(X^T V^{-1} X)^{-1}}{d\alpha_j}$

$$\begin{aligned} \frac{d(X^T V^{-1} X)^{-1}}{d\alpha_j} &= -(X^T V^{-1} X)^{-1} \frac{d(X^T V^{-1} X)}{d\alpha_j} (X^T V^{-1} X)^{-1} \\ &= (X^T V^{-1} X)^{-1} X^T V^{-1} \frac{dV}{d\alpha_j} V^{-1} X (X^T V^{-1} X)^{-1} \\ &= B \frac{dV}{d\alpha_j} B^T \end{aligned}$$

A.2.3 $\frac{d\beta}{d\alpha_j}$

$$\frac{d\beta}{d\alpha_j} = \frac{d}{d\alpha_j} (X^T V^{-1} X)^{-1} X^T V^{-1} Y$$

$$\begin{aligned}
&= B \frac{dV}{d\alpha_j} B^T X^T V^{-1} Y - (X^T V^{-1} X)^{-1} X^T V^{-1} \frac{dV}{d\alpha_j} V^{-1} Y \\
&= B \frac{dV}{d\alpha_j} (B^T X^T V^{-1} Y - V^{-1} Y) \\
&= -B \frac{dV}{d\alpha_j} QY
\end{aligned}$$

Similarly

$$\frac{dB}{d\alpha_j} = -B \frac{dV}{d\alpha_j} Q$$

A.2.4 $\frac{dQ}{d\alpha_j}$

$$\begin{aligned}
&\frac{d}{d\alpha_j} (V^{-1} - V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1}) \\
&= -V^{-1} \frac{dV}{d\alpha_j} V^{-1} + V^{-1} \frac{dV}{d\alpha_j} H + H \frac{dV}{d\alpha_j} Q \\
&= -V^{-1} \frac{dV}{d\alpha_j} Q + H \frac{dV}{d\alpha_j} Q \\
&= -Q \frac{dV}{d\alpha_j} Q
\end{aligned}$$

A.2.5 $\frac{dH}{d\alpha_j}$

$$\begin{aligned}
&= \frac{d}{d\alpha_j} V^{-1} X B \\
&= -V^{-1} \frac{dV}{d\alpha_j} H - H \frac{dV}{d\alpha_j} Q
\end{aligned}$$

A.2.6 $\frac{d\hat{e}}{d\alpha_j}$

$$\frac{d\hat{e}}{d\alpha_j} = \frac{d}{d\alpha_j} Y - X \hat{\beta}$$

$$= XB \frac{dV}{d\alpha_j} QY$$

$$\text{A.2.7} \quad \frac{d\hat{\sigma}^2}{d\alpha_j}$$

$$\frac{d\hat{\sigma}^2}{d\alpha_j} = \frac{d}{d\alpha_j} Y^T QY / N - p$$

$$= -Y^T Q \frac{dV}{d\alpha_j} QY / N - p$$

$$\text{A.2.9} \quad \frac{d\hat{\sigma}_{(a)}^2}{d\alpha_j}$$

$$\frac{d\hat{\sigma}_{(a)}^2}{d\alpha_j} = (Y_b^T Q_{(a)} \left(\frac{dV}{d\alpha_j} \right)_{bb} Q_{(a)} Y_b^T) / N - p - r$$

$$= \left(Y^T Q \frac{dV}{d\alpha_j} QY - 2Y^T Q \frac{dV}{d\alpha_j} Q \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix} + \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix}^T Q \frac{dV}{d\alpha_j} Q \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix} \right) / N - p - r$$

$$\text{A.2.10} \quad \frac{dRVC_{(a)}}{d\alpha_j}$$

$$\frac{dRVC_{(a)}}{d\alpha_j} = \frac{\hat{\sigma}^2 \frac{d\hat{\sigma}_{(a)}^2}{d\alpha_j} - \hat{\sigma}_{(a)}^2 \frac{d\hat{\sigma}^2}{d\alpha_j}}{\hat{\sigma}^4}$$

$$\text{A.2.11} \quad \frac{d\tilde{e}_{(a)}}{d\alpha_j}$$

$$\frac{d\tilde{e}_{(a)}}{d\alpha_j} = Q_{aa}^{-1} \left(Q \frac{dV}{d\alpha_j} Q \right)_{aa} Q_{aa}^{-1} (QY)_a - Q_{aa}^{-1} \left(Q \frac{dV}{d\alpha_j} QY \right)_a$$

$$= Q_{aa}^{-1} \left(Q \frac{dV}{d\alpha_j} Q \right)_{aa} \tilde{e}_{(a)} - Q_{aa}^{-1} \left(Q \frac{dV}{d\alpha_j} QY \right)_a$$

$$\text{A.2.12} \quad \frac{d\tilde{e}_{(P)}}{d\alpha_j} \quad \tilde{e}_{(P)} \text{ denotes the (stacked) vector based on } (\tilde{e}_{(a_1)}, \dots, \tilde{e}_{(a_r)})$$

Let $D_a = \text{Var}(\tilde{e}_{(a)}) = (Q_{aa})^{-1}$. Given $D_p^{-1}\tilde{e}_{(p)} = V^{-1}\hat{e} = QY$ where D_p is block diagonal with blocks $(D_{a_1}, \dots, D_{a_r})$. then

$$\frac{dD_p^{-1}}{d\alpha_j}\tilde{e}_{(p)} + D_p^{-1}\frac{d\tilde{e}_{(p)}}{d\alpha_j} = -Q\frac{dV}{d\alpha_j}QY$$

or

$$D_p^{-1}\frac{d\tilde{e}_{(p)}}{d\alpha_j} = -Q\frac{dV}{d\alpha_j}QY - \frac{dD_p^{-1}}{d\alpha_j}\tilde{e}_{(p)}$$

$$\frac{d\tilde{e}_{(p)}}{d\alpha_j} = D_p(-Q\frac{dV}{d\alpha_j}QY - \frac{dD_p^{-1}}{d\alpha_j}\tilde{e}_{(p)})$$

A.2.13 $\frac{d\text{DFBETA}_{(a)}}{d\alpha_j}$

$$\frac{d\text{DFBETA}_{(a)}}{d\alpha_j} = \frac{dB}{d\alpha_j}\begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix} + B\begin{pmatrix} \frac{d\tilde{e}_{(a)}}{d\alpha_j} \\ 0 \end{pmatrix}$$

$$= -B\frac{dV}{d\alpha_j}Q\begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix} + B\begin{pmatrix} \frac{d\tilde{e}_{(a)}}{d\alpha_j} \\ 0 \end{pmatrix}$$

A.2.14 $\frac{d\text{Cook's Distance}}{d\alpha_j}$

Let D_A denote Cook's Distance with $D_A(\hat{\sigma}^2, X^TV^{-1}X) = \tilde{e}_{(a)}^T H_{AA} \tilde{e}_{(a)} / p\hat{\sigma}^2$

$$\frac{dD_A(\hat{\sigma}^2, X^TV^{-1}X)}{d\alpha_j} = \frac{p\hat{\sigma}^2 \frac{d}{d\alpha_j}(\tilde{e}_{(a)}^T H_{AA} \tilde{e}_{(a)}) - \tilde{e}_{(a)}^T H_{AA} \tilde{e}_{(a)} \frac{d}{d\alpha_j} p\hat{\sigma}^2}{(p\hat{\sigma}^2)^2}$$

$$\frac{dD_A(\hat{\sigma}_{(a)}^2, X^TV^{-1}X)}{d\alpha_j} = \frac{p\hat{\sigma}_{(a)}^2 \frac{d}{d\alpha_j}(\tilde{e}_{(a)}^T H_{AA} \tilde{e}_{(a)}) - \tilde{e}_{(a)}^T H_{AA} \tilde{e}_{(a)} \frac{d}{d\alpha_j} p\hat{\sigma}_{(a)}^2}{(p\hat{\sigma}_{(a)}^2)^2}$$

$$\frac{dD_A(\hat{\sigma}_{(a)}^2, X_{(a)}^T V_{(a)}^{-1} X_{(a)})}{d\alpha_j} = \frac{p\hat{\sigma}_{(a)}^2 \frac{d}{d\alpha_j} (\tilde{e}_{(a)}^T (H_{AA} - H_{AA}(V^{AA})^{-1} H_{AA}) \tilde{e}_{(a)}) - \tilde{e}_{(a)}^T (H_{AA} - H_{AA}(V^{AA})^{-1} H_{AA}) \tilde{e}_{(a)} \frac{d}{d\alpha_j} p\hat{\sigma}_{(a)}^2}{(p\hat{\sigma}_{(a)}^2)^2}$$

$$\frac{d\beta}{d\alpha_j} = -B \frac{dV}{d\alpha_j} QY \quad (4.1)$$

$$\frac{dQ}{d\alpha_j} = -Q \frac{dV}{d\alpha_j} Q \quad (4.2)$$

$$\frac{d\hat{e}}{d\alpha_j} = XB \frac{dV}{d\alpha_j} QY \quad (4.3)$$

$$\frac{d\hat{\sigma}^2}{d\alpha_j} = -Y^T Q \frac{dV}{d\alpha_j} QY / N - p \quad (4.4)$$

$$\frac{d\hat{\sigma}_{(a)}^2}{d\alpha_j} = \left(Y^T Q \frac{dV}{d\alpha_j} QY - 2Y^T Q \frac{dV}{d\alpha_j} Q \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix} + \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix}^T Q \frac{dV}{d\alpha_j} Q \begin{pmatrix} \tilde{e}_{(a)} \\ 0 \end{pmatrix} \right) / N - p - r \quad (4.5)$$

$$\frac{dRVC_{(a)}}{d\alpha_j} = \frac{\hat{\sigma}^2 \frac{d\hat{\sigma}_{(a)}^2}{d\alpha_j} - \hat{\sigma}_{(a)}^2 \frac{d\hat{\sigma}^2}{d\alpha_j}}{\hat{\sigma}^4} \quad (4.6)$$

