School Of Public Health

Some Applications of Local Influence Diagnostics

John Siu-Hung Yick

"This thesis is presented as part of the requirements for
the award of the Degree of Doctor of Philosophy
of the
Curtin University of Technology"

May 2000

STATEMENT

This thesis is my own work and has not been submitted previously, in whole or in part, in respect of any other academic award.

John Yick

ACKNOWLEDGMENTS

I wish to express my greatest appreciation to my supervisor, Dr. Andy Lee, for his advice, support and guidance. The completion of this thesis would not have been possible without his continual encouragement, especially during times of difficulty. It is also my pleasure to thank him for going through this thesis so vigilantly and providing much needed suggestions for improvements.

I gratefully acknowledge the receipt of a Northern Territory University Research Scholarship which enabled me to start this project. I also wish to acknowledge Curtin University of Technology for granting me an admission in order to complete this project.

My heartfelt thanks also go to my family and friends for their moral support. Lastly, I express my deepest gratitude to my wife Cherli Low and my mother for their patience, understanding and support.

John Yick

ABSTRACT

The influence of observations on the outcome of an analysis is of importance in statistical data analysis. A practical and well-established approach to influence analysis is case deletion. However, it has its draw-backs when subsets of observations are jointly influential and offset each other's influence. Another approach is local influence proposed by Cook (1986).

The local influence methodology of Cook (1986) is based on the curvature of the likelihood displacement surface formed by model/data perturbations. Wu and Luo (1993a, 1993b) further developed the idea and proposed the study of the perturbation-formed surface of a variable by evaluating the curvature of the surface in addition to its maximum slope. This thesis utilizes the local influence approach to develop influence diagnostic methods for four different topics.

Firstly, we proposed a stepwise confirmatory procedure for the detection of multiple outliers in two-way contingency tables. The procedure begins with the identification of a reliable set of candidate outliers by evaluating the derivatives of the perturbation-formed surface of the Pearson goodness-of-fit statistic. An adding-back iterative algorithm is then applied to the candidate set to assess their relative discordancy. Using two real data sets, the proposed procedure is shown to be less susceptible to both masking and swamping problems than residual based measures. In a Monte Carlo study, the local influence diagnostics are also found to outperform standard residual-based methods in terms of efficiency and other criteria.

Transformations of covariates are commonly applied in regression analysis. When a parametric transformation family is used, the maximum likelihood estimate of the transformation parameter is often sensitive to minor perturbations of the data. Diagnostics based on the local influence approach are derived to assess the influence of observations on the covariate transformation parameter in

generalized linear models. Three numerical examples are presented to illustrate the usefulness of the proposed diagnostics. The need for transformation is also addressed in addition to assessing influence on the transformation parameter.

A common method of choosing the link function in generalized linear models is to specify a parametric link family indexed by unknown parameters. The maximum likelihood estimates of such link parameters, however, often depend on one or several extreme observations. Diagnostics based on the local influence approach are derived to assess the sensitivity of the parametric link analysis. Two examples demonstrate that the proposed diagnostics can identify jointly influential observations on the link even when masking is present. The application of the diagnostics can also assist us in revising the link parameter and hence the form of the model.

The portmanteau statistic is commonly used for testing goodness-of-fit of time series models. However, this lack of fit test may depend on one or several atypical observations in the series. We investigate the sensitivity of the portmanteau statistic in the presence of additive outliers. Diagnostics based on the local influence approach are developed to assess both local and global influence. Three practical examples demonstrate the usefulness of the proposed diagnostics.

PUBLICATIONS FROM THE RESEARCH WORK OF THIS THESIS

Yick, J. S. and Lee, A. H. (1998). Unmasking outliers in two-way contingency tables, *Computational statistics & data analysis*, **29**, 69-79.

Lee, A. H. and Yick, J. S. (1999). A perturbation approach to outlier detection in two-way contingency tables, *Australian & New Zealand journal of statistics*, **41**, 305-314.

Lee, A. H. and Yick J. S. (1999). Covariate transformation diagnostics for generalized linear models, *Annals of the institute of statistical mathematics*, **51**, 383-398.

Yick J. S. and Lee, A. H. (1998). Sensitivity of parametric link functions in generalized linear models, *Journal of biopharmaceutical statistics*, **8**, 391-406.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

1.1 Aims and scope of this thesis

Statistical models are extremely useful devices for extracting and understanding the essential features of a set of data. It is well known that models are always approximate description of the underlying processes and therefore are nearly always not perfect. Because of this inexactness, the study of the variation in the results of an analysis under modest modifications of the problem formulation becomes important. If a minor modification of an approximate description seriously influences key results of an analysis, there is surely cause for concern. On the other hand, if such modifications are found to be unimportant, the sample is robust with respect to the induced perturbation and our ignorance of the precise model will do no harm (Barnard, 1980).

Much of the work on influence assessment of a model perturbation is concerned with the perturbation scheme in which the weights attached to individuals or groups of cases are modified. For the most part, the case-weights are restricted to be either 0 or 1 so that a case is either deleted or retained in full weight. One of the drawback of the single case deletion method is that subsets of observations can be jointly influential or can offset each other's influence. Although influence subsets or multiple outliers can be identified by generalizing deletion statistics formally to subsets of several points, the very large number of subsets usually renders the approach impractical. (There are n!/[p!(n-p)!] subsets of size p).

Cook (1986) proposed the local influence approach to influence assessment. The idea of differentiation instead of deletion is prominent in this approach, a change in paradigm from that of case-deletion. Local influence is based on perturbation of a case or model components and not on its total deletion, and employs a differential comparison of parameter estimates before and after perturbation. This approach gives us the influence of all the cases on the model

simultaneously, not individual cases or subset of cases separately by themselves.

Lawrance (1988) compared Cook's local influence approach with the deletion approach. Important advantages of the local influence approach were highlighted. Several useful schemes of perturbation were considered in that paper with some emphasis on the assessment of the effect of perturbations.

Wu and Luo (1993a, 1993b) examined global and local influence by studying the surface of a variable, such as the maximum likelihood estimate (MLE), formed by perturbation, which they referred as the MLE surface. Their method, which is called the second ordered approach, is based on an assessment of the directions corresponding to large local maximum curvatures of the MLE surface, in addition to the study of the maximum slope directions. It was found that, for a single parameter of interest, the maximum slope of the MLE surface is the same as the direction of the maximum curvature of Cook's likelihood displacement surface. Wu and Luo referred to the maximum slope as the first order approach.

This thesis intends to utilize the local influence approach, both first and second order, to develop influence diagnostic methods for the following topics.

1.1.1 Contingency tables

Diagnostics for measuring model deviations and influence of particular data points are used extensively in modern regression analysis. For contingency tables however, it is not the influence of individual cases which is of interest, but rather the contribution to a lack of fit from the observed counts in a single cell in the table, which must be evaluated. It is possible that a small subset of outlying cells can give rise to significant test statistics when testing a particular model for fit.

To detect outliers and interaction in two-way tables of measurements, one often study residuals that become available after applying some process of fitting the data. However, when dealing with similar matters in two-way contingency tables, one shall encounter more features that need attention (Mosteller and Parunak, 1985). Published methods for location of these cells include residual-based methods and the sequential identification of cells whose deletion leads to the maximum reduction in the statistic for independence (Brown (1974), Perli, Hommel and Lehmacher (1985), Muñoz-García, Moreno-Rebollo and Pascual-Acosta (1987), Andersen (1992), García-Heras, Muñoz-García and Pascual-Acosta (1993)). Also see Barnett and Lewis (1994, Chapter 12) for a general review. Such procedures, however, tend to suffer from the deficiencies of masking and swamping. Simonoff (1988) gave some discussion on the masking and swamping problem in contingency tables.

Local perturbation approach based on the maximum slope and large curvatures associated with the perturbation formed surface of the Pearsons Chi-square statistic will be investigated in Chapter 3 to reduce the masking and swamping problems. A stepwise outlier testing procedure proposed by Lee and Fung (1997) will also be examined and modified. Numerical examples and a Monte Carlo study will be used to assess the effectiveness of the local perturbation diagnostics and the stepwise testing procedure.

1.1.2 Covariate transformations in generalized linear models

Transformation of variables have been traditionally applied to data in an effort to identify an easily interpretable and scientifically meaningful relationship. Deletion diagnostics for assessing the case influence on the transformation parameter estimators in the Box-Cox regression model have been extensively studied in the last decade, see e.g. Cook and Wang (1983), Atkinson (1986), Hinkley and Wang (1988). Recently, the local influence approach have also been employed for diagnosing transformation. Many articles related to the local influence approach deal with the response transformation or the transform-both-sides models, e.g. Lawrance (1988), Tsai and Wu (1992). Few articles address the transformation of the covariates or explanatory variables only.

Wei and Hickernell (1996) presented diagnostics for the transformation of explanatory variables using local influence. Their diagnostics, however, are only applicable to the linear regression setting. This thesis extends the approach to generalized linear models. We shall also address the need for transformation in addition to assessing influence on the transformation parameter in Chapter 4.

1.1.3 Parametric link functions in generalized linear models

The Link function, which relates the linear predictor to the expected value of the response, is a major component of the generalized linear models (McCullagh and Nelder, 1989). The maximum likelihood estimate of the link parameter, and hence the exact form of the link function, may depend crucially on one or a few extreme observations.

Our aim is to propose influence diagnostics for assessing the effect of minor perturbations on the maximum likelihood estimator of the link parameter in generalized linear models. In Chapter 5, three separate approaches based on the analysis of the link parameter surface, profile likelihood displacement, and

partial influence will be investigated. Several numerical examples will be used to illustrate sensitivity of the link analysis. The method enables one to revise the link parameter and hence the form of the model.

1.1.4 Portmanteau statistic in time series

The fact that goodness-of-fit of a time series model is very prone to atypical observations has been a main concern to practitioners and researchers. The major difficulty in detecting multiple outliers in time series is due to the correlation structure of the process giving rise to the so called masking and smearing problems (Bruce and Martin, 1989). Although numerous methods have been proposed in literature on the topic of outlier detection, to our knowledge, no work has been done on assessing the effect of multiple/ consecutive/ patches of outliers on the model adequacy via the Portmanteau statistic. Fox (1972) proposed two types of intervention models to classify two types of outliers, additive and innovational, that might occur in practice. Wu, Hosking and Ravishanker (1993) introduced the notion of reallocation outliers, which are additive outliers whose magnitudes sum to zero. Many other patterns of outliers can possibly take place but our effort is focused on additive outliers.

In Chapter 6 of this thesis, we will adopt the local influence approach to study the slope of the perturbation-formed surface of the Portmanteau statistic. Numerical examples will be used to illustrate the usefulness of the proposed diagnostic in assessing sensitivity of the Portmanteau statistic.

1.2 Notational conventions

Most of the notations used in this thesis should be consistent with those in the literature. Outlined below is the notational system used in this thesis.

The convention $\boldsymbol{x} = \{x_{ij}\}$ indicates that the matrix $\boldsymbol{x}$ has x_{ij} as its typical $(i,j)th$ element. Bold face quantities are reserved for vectors and should be distinguished from ordinary scalars. The transpose of a matrix $\boldsymbol{x}$ is denoted by $\boldsymbol{x}^T$. Parameters to be estimated are often denoted by Greek letters such as θ with corresponding estimate $\hat{\theta}$. The differentiation operator is $\frac{\partial}{\partial\theta}$ throughout this thesis.

For two way contingency tables, x_{i+} denotes the row marginal total and x_{+j} denotes the column marginal total, each summing to the sample size N. p_{ij} represents the probability for the (i,j)th cell x_{ij} with row marginal probability p_{i+} and column marginal probability p_{+j}. x_{IJ} denotes are particular cell where as x_{ij} denotes a general cell. The expectation operator is denoted by $E()$ and e_{ij} denotes the expected cell frequency.

In covariate transformation, $\mathbf{y}$ denotes the responses, $\mathbf{x}$ denotes the covariates and $\mathbf{z}$ is reserved for the covariates to be transformed. The linear predictor is denoted by η and the Likelihood operator is denoted by $L()$. Similar notations apply to the generalized linear models in the link function Chapter.

For time series models, y_t is the tth observed value, $\mathbf{a}$ denotes the residuals, r_k denotes the lag k autocorrelation of the residuals and B represents the backward shift operator.

1.3 Statistical computing

Statistical computations are required extensively for this thesis. Commonly used statistics, parameters in regressions, their standard errors, and simulations of data are carried out using the existing S-plus functions. More complex procedures and the proposed diagnostics are implemented by writing customized functions in S-plus. For brevity, such functions and programs are not listed in this thesis. However, they are readily available from the author upon request or downloaded from the author's home page at

$$http://www.cherli.com/johnyick/Splusprog.html.$$

At times, the results maybe slightly different if another statistical computing software is used. This could be due to the different numerical algorithms facilitated in different softwares. However, it is envisaged that such potential discrepancies will be small and should not grossly affect the end results or conclusions.

CHAPTER 2

LOCAL INFLUENCE METHODOLOGY

2. OVERVIEW OF LOCAL INFLUENCE

Cook (1986) introduced the idea of local influence as a general tool to assess the effect of small perturbations to data. The idea of differentiation instead of deletion is prominent in this approach. Local influence is based using the perturbation of a case and not on its complete deletion, and employs a differential comparison of parameter estimates before and after perturbation. Cook (1986) proposed the local influence approach on the perturbation formed likelihood displacement surface and gave some results in the linear regression setting. The concept was also discussed in Cook (1987).

Lawrance (1988) further developed the idea of local influence in the linear regression setting with a weaker reliance on geometrical curvatures and compared it to the case deletion approach. The paper also extended the perturbation approach to include the linear model assumptions of constant variance and independence, and contrasted to perturbations of the response and explanatory variables. Extensions of the local influence approach to generalized linear models were given by Thomas and Cook (1989, 1990).

Wu and Luo (1993a) proposed to examine global and local influences by studying the surface of a variable, such as the maximum likelihood estimate (MLE), formed by perturbation, which they referred to as the MLE surface. Their method, the so called second-order approach, is based on an assessment of the directions corresponding to relatively large local maximum curvatures of the MLE surface, in addition to the study of the maximum slope directions. It was found that, for a single parameter of interest, the maximum slope of the MLE surface is the same as the direction of the maximum curvature of Cook's likelihood displacement surface. Wu and Luo referred to the maximum slope as the first-order approach. Application of the approach was illustrated via the transformation model. Wu and Luo (1993b) gave further illustrations by applying the approach to residual sum of squares and multiple potentials in

regression are given in Section 2.2.

2.1 Definition of local influence

This section presents a review of the definition and the development of Cook's (1986, 1987) local influence methodology.

Suppose a statistical model with a p-dimensional parameter vector $\boldsymbol{\theta}$ is of interest. General perturbations are introduced into the model through an $n \times 1$ vector $\boldsymbol{\omega}$. Write $L(\boldsymbol{\theta}|\boldsymbol{\omega})$ for a log-likelihood corresponding to the perturbed data. Let $\hat{\boldsymbol{\theta}}$ be the maximum likelihood estimate of $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ be the corresponding maximum likelihood estimate from the perturbed model. The influence of the perturbation $\boldsymbol{\omega}$ can be assessed by the log-likelihood displacement

$$LD(\boldsymbol{\omega}) = 2\left[L(\hat{\boldsymbol{\theta}}) - L\left(\hat{\boldsymbol{\theta}}(\boldsymbol{\omega})\right)\right],$$

with $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{\omega}_0)$ being the parameter estimate under no perturbation. The $LD(\boldsymbol{\omega})$ measures the distance between $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}(\boldsymbol{\omega}_0)$ in terms of the log-likelihood difference. To locate the directions of large change at $LD(\boldsymbol{\omega}_0)$, we approximate the surface by its tangent plane. However, since $LD(\boldsymbol{\omega}_0)$ achieves its maximum at $\hat{\boldsymbol{\theta}}$, the likelihood displacement is nonnegative and achieves its minimum at $\boldsymbol{\omega}_0$, so the tangent plane at $LD(\boldsymbol{\omega}_0)$ is horizontal and carries no diagnostic information about local change. Hence we turn to the normal curvatures at $LD(\boldsymbol{\omega}_0)$.

The normal curvature at $LD(\boldsymbol{\omega}_0)$ in the unit direction $\boldsymbol{d}$ may be expressed as:

$$C(\boldsymbol{d}) = 2|\boldsymbol{d}^T \boldsymbol{\Delta} L^{(2)}(\hat{\boldsymbol{\theta}})^{-1} \boldsymbol{\Delta}^T \boldsymbol{d}|, \tag{2.1}$$

where $\boldsymbol{\Delta} = \partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\theta}^T\partial\boldsymbol{\omega}$ and $L^{(2)}(\hat{\boldsymbol{\theta}}) = \partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}^T$ are both evaluated at $\boldsymbol{\omega}_0$ and $\hat{\boldsymbol{\theta}}$.

Let $\boldsymbol{d}_{max}$ be the direction cosine of maximum normal curvature, which is the perturbation direction that produces the greatest local change in the parameter estimates as measured by the likelihood displacement. The most in-

fluential element of the data may be identified by their large components of the vector $\boldsymbol{d}_{max}$, whereas sensitivity of the perturbation may be indicated by the value $C_{max} = C(\boldsymbol{d}_{max})$, which is the maximum of equation (2.1) over all possible directions. According to matrix theory, C_{max} is the largest eigenvalue of $\boldsymbol{\Delta} L^{(2)}(\hat{\boldsymbol{\theta}})^{-1}\boldsymbol{\Delta}^T$ and $\boldsymbol{d}_{max}$ is the corresponding eigenvector. The directional cosine $\boldsymbol{d}_{max}$ may be used as diagnostics for influence.

When a subset $\boldsymbol{\theta}_1$ from the partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ is of interest, diagnostic for influence can be based on

$$\boldsymbol{\Delta}\left\{[L^{(2)}]^{-1} - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & [L_{22}^{(2)}]^{-1} \end{pmatrix}\right\}\boldsymbol{\Delta}^T \tag{2.2}$$

where $[L_{22}^{(2)}]$ is determined from the partition $[L^{(2)}] = \begin{pmatrix} [L_{11}^{(2)}] & [L_{12}^{(2)}] \\ [L_{21}^{(2)}] & [L_{22}^{(2)}] \end{pmatrix}$ conformably with the partition of $\boldsymbol{\theta}$.

2.2 Second-order approach

Wu and Luo (1993a, 1993b) studied the perturbation-formed MLE surface of a parameter of interest in regression. Unlike the likelihood displacement surface, such a MLE surface does not necessarily have zero first derivative at the null point of no perturbation, so that its slope as well as curvature can be used to examine local influence. They referred to the maximum slope as the first-order approach and directions corresponding to the large normal curvatures as the second-order approach.

Suppose that we have n cases for a given model and the parameter/statistic of interest is η. We introduce small perturbation into the data through a $n \times 1$ vector $\boldsymbol{\omega} = \boldsymbol{\omega}_0 + a\boldsymbol{l} \in \Omega$, where Ω denotes the open set of relevant perturbations, $\boldsymbol{l} = (l_1, ..., l_n)^T$ is a unit direction vector, the scalar a measures the magnitude of the perturbation in direction $\boldsymbol{l}$.

The null point $\boldsymbol{\omega}_0 = (1, ..., 1)^T$ which is equivalent to $a = 0$, represents no perturbation in any direction. If the case k is the only perturbation, then it is located in direction $-\boldsymbol{l}_{<k>}$ from the null point, where $\boldsymbol{l}_{<k>}$ is defined as

$$\boldsymbol{l}_i = \begin{cases} 1 & for \quad i = k, \\ 0 & for \quad i \neq k. \end{cases}$$

Under this structure, the MLE $\hat{\eta}$ of η is a function of $\boldsymbol{\omega} = \boldsymbol{\omega}(a)$. The geometric surface formed by the vector $\boldsymbol{\alpha}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^T, \hat{\eta}(\boldsymbol{\omega}))^T$ contains essential information on the influence of the perturbation in question. Following the derivation of Cook (1986), Wu and Luo obtained the signed normal curvature in direction $\boldsymbol{l}$,

$$C_{\boldsymbol{l}} = \frac{\boldsymbol{l}^T \eta^{(2)} \boldsymbol{l}}{(1 + \eta^{(1)T}\eta^{(1)})^{1/2} \boldsymbol{l}^T (I + \eta^{(1)}\eta^{(1)T})\boldsymbol{l}}, \tag{2.3}$$

with

$$\left.\frac{\partial \hat{\eta}}{\partial \omega^T}\right|_{\omega_0} = \hat{\eta}^{(1)T}\boldsymbol{l},$$

$$\left.\frac{\partial^2 \hat{\eta}}{\partial \omega \partial \omega^T}\right|_{\omega_0} = \boldsymbol{l}^T \hat{\eta}^{(2)T}\boldsymbol{l}.$$

The curvature $C_{\boldsymbol{l}}$ in equation (2.3) takes the form

$$\frac{\boldsymbol{l}^T F \boldsymbol{l}}{\boldsymbol{l}^T M \boldsymbol{l}}. \tag{2.4}$$

According to matrix theory, the local maximum curvatures and corresponding directions are the eigenvalue-eigenvector solution of the equation

$$|F - \lambda M| = 0. \tag{2.5}$$

The diagnostics for the first-order approach are the by-products of this method. The value of maximum slope is $\eta^{(1)T}\boldsymbol{l}$ which is the direction with cosine

$$\frac{\eta_i^{(1)}}{||\eta^{(1)}||} \quad for i = 1, ..., n.$$

Wu and Luo also recommended the study of plots of $\hat{\eta}(\omega)$ against the perturbation scale a for any direction of interest, including the directional cosines associated with the maximum slope and large local curvatures. As mentioned previously, the deletion of case k is related to the perturbation in direction $-\boldsymbol{l}_{<k>}$ from the null point. Therefore, if case k is globally influential, the $\hat{\eta}(\omega)$ curve associated with that direction should have a great change on the $a > 0$ side. Now suppose that a direction $\boldsymbol{l}$ of interest has an outstanding cosine $\boldsymbol{l}_k$ say. If this large $\boldsymbol{l}_k > 0$ then $\hat{\eta}(\omega)$ associated with $\boldsymbol{l}$ should behave like $\hat{\eta}(\omega)$ associated with $\boldsymbol{l}_{<k>}$. Similarly if this large $\boldsymbol{l}_k < 0$ then $\hat{\eta}(\omega)$ associated with $\boldsymbol{l}$ should behave like $\hat{\eta}(\omega)$ associated with $-\boldsymbol{l}_{<k>}$. Therefore, we can investigate global influence by studying these plots.

Wu and Luo applied the local influence method to transformation parameters, residual sum of squares and multiple potentials in regression and demonstrated that it is very effective as a diagnostic tool. The work of this thesis will be based on Wu and Luo's formulation of local influence.

CHAPTER 3

TWO-WAY CONTINGENCY TABLES

3. TWO-WAY CONTINGENCY TABLES

When assessing the goodness-of-fit of the independence model in contingency tables, it is possible that a small subset of deviating cells can give rise to a significant test statistic. The major problem is detection of such cells rather than the subsequent outlier-resistant analysis (Bradu and Hawkins, 1982). Published methods for location of outliers include discordancy criteria and the sequential identification of cells whose omission leads to the maximum reduction in the statistic for independence. An excellent review of the literature can be found in Barnett and Lewis (1994, Chapter 12). Residuals of various forms have often been incorporated into formal detection procedures. Such procedures, however, tend to suffer from the deficiencies of masking and swamping. Here, masking is considered in respect of a measure of outlyingness. A conditional perspective on masking and other aspects of deletion influence in regression are examined by Lawrance (1995).

To identify multiple outliers in the context of a two-way contingency table, we study the perturbation-formed surface of the Pearson goodness-of-fit statistic, by evaluating the tangent plane and curvatures of the surface. Unlike the assessment of local influence (Cook, 1986) on parameters of a model, we are concerned with assessing the goodness-of-fit with respect to the independence assumption. Through an example, it is found that the perturbation approach provides effective diagnostics that are resistant to masking and swamping. A Monte Carlo study was also carried out to assess the effectiveness of this detection technique. Results indicate that the proposed diagnostics outperform standard residual-based methods in terms of efficiency and other criteria. In Section 3.3, we propose a stepwise confirmatory procedure to assess the relative discordancy of the candidate set of outliers. Two examples are given to demonstrate the usefulness of this approach.

3.1 Assessing goodness-of-fit

The problem of outliers detection in contingency tables has generally been considered from the point of view of the independence hypothesis (Barnett and Lewis (1994, p. 432)). Suppose we have a two-way $r \times c$ contingency table giving observed frequencies x_{ij} $(i = 1, ..., r; j = 1, ..., c)$, with row marginal totals x_{i+} and column marginal totals x_{+j}, each summing to the sample size N. The usual assumed null model is of independence between the row and column classifications. An *outlier* is a particular cell whose observed frequency deviates significantly from the corresponding expected frequency on the null model. Let $p_{ij} = E[x_{ij}/N]$. Denoting the entire set of $n = rc$ cells by S and the subset of cells under question by Q, the independence model is

$$H_0: \quad p_{ij} = p_{i+}p_{+j} \qquad \left\{ \sum_i p_{i+} = \sum_j p_{+j} = 1 \right\}, \quad \forall (i,j) \in S.$$

The two most prominent statistics for testing H_0 against all omnibus alternatives are the Pearson goodness-of-fit statistic

$$X^2 = \sum_{i,j} \frac{(x_{ij} - e_{ij})^2}{e_{ij}}$$

and the likelihood ratio goodness-of-fit statistic

$$G^2 = 2 \sum_{i,j} x_{ij} \ln(x_{ij}/e_{ij})$$

where the expected cell frequency $e_{ij} = Np_{ij}$ is estimated by $x_{i+}x_{+j}/N$ under multinomial sampling. Both statistics have a null asymptotic χ^2 distribution with $(r-1) \times (c-1)$ degrees of freedom.

Outliers detection in the above context should be distinguished from the related problem of influence assessment where the main objective is assessing sensitivity of model parameters. To identify influential observations in contingency tables, Andersen (1992) proposed (cell) deletion diagnostics for the parametric multinomial distribution, with particular applications to the Goodman's RC association model (Goodman, 1981).

3.1.1 Adjusted and deleted residuals

When the hypothesis of independence is rejected, it is important to identify any cell(s) which display lack-of-fit and whose removal will lead to acceptance of the quasi-independence model

$$H_1: \quad p_{ij} = p_{i+}p_{+j} \qquad \left\{ \sum_{i,j} p_{i+}p_{+j} = 1 \right\}, \quad \forall (i,j) \notin Q.$$

A variety of data analytic techniques have been proposed, the basic building blocks of which are residuals of different forms; see Barnett and Lewis (1994, Chapter 12) for a review. From the *standardized residual*

$$r_{ij} = (x_{ij} - e_{ij}) / \sqrt{e_{ij}} \ , \tag{3.1}$$

Haberman (1973) constructed the *adjusted residual* as

$$\tilde{r}_{ij} = r_{ij} / \{(1 - x_{i+}/N)(1 - x_{+j}/N)\}^{\frac{1}{2}}. \tag{3.2}$$

Another popular measure is the *deleted residual* (Brown, 1974), which is defined as

$$r^*_{ij} = \left(x_{ij} - e^*_{ij}\right) / \sqrt{e^*_{ij}} \ , \tag{3.3}$$

where

$$e^*_{ij} = (x_{i+} - x_{ij})(x_{+j} - x_{ij})/(N - x_{i+} - x_{+j} + x_{ij}) \tag{3.4}$$

denotes the expected value of the (i,j)th cell under H_1 with the (i,j)th cell *deleted* or treated as missing.

The distributions of $\tilde{r}_{ij}$ in different experimental situations are given by Muñoz-García et al (1987), while Fuchs and Kenett (1980) suggested using the Bonferroni bound to provide critical values for the maximum $\tilde{r}_{ij}$ to test for discordancy. Simonoff (1988) advocated the use of r^*_{ij} coupled with a backwards-stepping algorithm for outlier detection. The null distribution of r^*_{ij}, however, does not appear to be known.

3.1.2 Perturbation on Pearson goodness-of-fit statistic

For a $r \times c$ contingency table, a similar approach to Wu and Luo (1993a) can be applied to the perturbation-formed surface of the Pearson goodness-of-fit statistic. We introduce small changes $\boldsymbol{\omega} = \langle \omega_{11}, ..., \omega_{rc} \rangle^T \in \Omega$, into the assumed probability p_{ij} so that it becomes $\omega_{ij} p_{ij}$, where Ω denotes the open set of relevant perturbations and $\omega_{ij} > 0 \ \forall (i,j)$. Suppose there is a null point $\boldsymbol{\omega}_0 = \langle 1, ..., 1 \rangle^T \in \Omega$ representing no perturbation so that $X^2(\boldsymbol{\omega}_0)$ gives the observed X^2 statistic. The geometric surface of interest is formed by $X^2(\boldsymbol{\omega})$, the Pearson goodness-of-fit statistic under perturbation $\boldsymbol{\omega}$. To find the direction of largest local change, we approximate the surface by its tangent plane at $\boldsymbol{\omega}_0$, which is determined by $\frac{\partial X^2(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^T}$ at $\boldsymbol{\omega}_0$. The direction of largest local change is just the direction of maximum slope on this tangent plane over Ω. It is found that

$$\left. \frac{\partial X^2(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^T} \right|_{\boldsymbol{\omega}_0} = X^{(1)} = \langle X_{ij}^{(1)} \rangle \tag{3.5}$$

where

$$X_{ij}^{(1)} = (x_{ij} - e_{ij}) \left\{ -2 - (x_{ij} - e_{ij})/e_{ij} \right\}. \tag{3.6}$$

In the second order approach, the direction cosines corresponding to large curvatures of the surface of interest are used to supplement the first order diagnostics. The second derivative of $X^2(\boldsymbol{\omega})$ evaluated at $\boldsymbol{\omega}_0$ is a diagonal matrix with entries

$$4x_{ij} - 2e_{ij} + 2\left(x_{ij} - e_{ij}\right)^2 / e_{ij} \ . \tag{3.7}$$

The local maximum curvatures and their corresponding directions are then found via equation (2.5). We denote the resulting direction of largest local maximum curvature by $X^{(2)} = \langle X_{ij}^{(2)} \rangle$, and together with $X^{(1)}$, they form our main diagnostic quantities.

Alternatively, instead of treating X^2 as a data-generated statistic, we consider X^2 as a function of n non-negative but not necessarily integer quantities

x_{ij}. We wish to study the perturbation-formed surface of this function by introducing small changes ω to the variables such that x_{ij} becomes $\omega_{ij}x_{ij}$. Again, the geometric surface of interest $X^2(\omega)$ can be approximated by its tangent plane at $\omega_0 = \langle 1, ..., 1 \rangle^T$. The resulting direction of maximum slope on this tangent plane is

$$\left.\frac{\partial X^2(\omega)}{\partial \omega^T}\right|_{\omega_0} = X' = \left\langle X'_{ij} \right\rangle^T , \tag{3.8}$$

where for a particular entry X'_{IJ},

$$\begin{aligned} X'_{IJ} = & \frac{2\left(x_{IJ} - e_{IJ}\right)}{e_{IJ}} \left\{ x_{IJ} + \frac{x_{I+} + x_{+J} + e_{IJ} - (x_{IJ} - e_{IJ})(x_{I+} + x_{+J} - e_{IJ})}{N} \right\} \\ & + \sum_j \frac{x_{IJ}}{e_{Ij}} \frac{x_{+j}}{N} \left(1 - \frac{x_{I+}}{N}\right) (e_{Ij} - x_{Ij})(e_{Ij} + x_{Ij}) \\ & + \sum_i \frac{x_{IJ}}{e_{iJ}} \frac{x_{i+}}{N} \left(1 - \frac{x_{+J}}{N}\right) (e_{iJ} - x_{iJ})(e_{iJ} + x_{iJ}) \\ & - \sum_{i \neq I, j \neq J} \frac{x_{IJ}}{e_{ij}} \frac{e_{ij}}{N} (e_{ij} - x_{ij})(e_{ij} + x_{ij}) \ . \end{aligned} \tag{3.9}$$

For the second order approach, perturbation directions corresponding to large curvatures of the surface are recommended for assessing local sensitivity, large components of which signify the presence of aberrant cells. Since the curvature is a function of the first and second derivatives, we need to evaluate $\frac{\partial^2 X^2(\omega)}{\partial \omega^2}$ at ω_0 as well. It can be shown that

$$\begin{aligned} \frac{\partial^2 X^2(\omega)}{\partial \omega_{IJ} \partial \omega_{IJ}} = & 2x_{IJ}^2 \left\{ \frac{2x_{IJ} + N}{x_{I+}x_{+J}} + \frac{x_{IJ}^2}{e_{IJ}N} - 2N \left(\frac{1}{x_{I+}^2 x_{+J}} + \frac{1}{x_{I+} x_{+J}^2} \right) \right\} \\ & + 2x_{IJ}^2 \left\{ \sum_j \frac{x_{Ij}^2}{e_{Ij} x_{I+}^2} - \frac{x_{Ij}^2}{e_{Ij} x_{I+} N} + \sum_i \frac{x_{iJ}^2}{e_{iJ} x_{+J}^2} - \frac{x_{iJ}^2}{e_{iJ} x_{+J} N} \right\} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 X^2(\omega)}{\partial \omega_{IJ} \partial \omega_{IK}} = & x_{IJ} x_{IK} \left\{ \frac{x_{IJ}^2}{e_{IJ}^2 N} + \frac{x_{IK}^2}{e_{IK}^2 N} + \frac{x_{IJ}}{e_{IJ} N} + \frac{x_{IK}}{e_{IK} N} - \frac{2x_{IJ}}{e_{IJ} x_{I+}} - \frac{2x_{IK}}{e_{IK} x_{I+}} \right\} \\ & + x_{IJ} x_{IK} \left\{ \sum_j \frac{x_{Ij}^2}{e_{Ij}} \left(\frac{1}{x_{I+}} - \frac{1}{x_{I+} N} \right) - \sum_i \frac{x_{iJ}^2}{e_{iJ} x_{+J} N} - \frac{x_{iK}^2}{e_{iK} x_{+K} N} \right\} \end{aligned}$$

$$\frac{\partial^2 X^2(\boldsymbol{\omega})}{\partial\omega_{IJ}\partial\omega_{LJ}} = x_{IJ}x_{LJ}\left\{\frac{x_{IJ}^2}{e_{IJ}^2 N}+\frac{x_{LJ}^2}{e_{LJ}^2 N}+\frac{x_{IJ}}{e_{IJ}N}+\frac{x_{LJ}}{e_{LJ}N}-\frac{2x_{IJ}}{e_{IJ}x_{+J}}-\frac{2x_{LJ}}{e_{LJ}x_{+J}}\right\}$$
$$+ x_{IJ}x_{LJ}\left\{\sum_i \frac{x_{iJ}^2}{e_{iJ}}\left(\frac{1}{x_{+J}}-\frac{1}{x_{+J}N}\right)-\sum_j \frac{x_{Ij}^2}{e_{Ij}x_{I+}N}-\frac{x_{Lj}^2}{e_{Lj}x_{L+}N}\right\}$$
$$\frac{\partial^2 X^2(\boldsymbol{\omega})}{\partial\omega_{IJ}\partial\omega_{LK}} = x_{IJ}x_{LK}\left\{\frac{2x_{IJ}}{e_{IJ}}+\frac{2x_{LK}}{e_{LK}}-\sum_i \frac{x_{iJ}^2}{e_{iJ}x_{+J}}-\frac{x_{iK}^2}{e_{iK}x_{+K}}\right.$$
$$\left.-\sum_j \frac{x_{Ij}}{e_{Ij}x_{I+}}-\frac{x_{Lj}}{e_{Lj}x_{L+}}\right\} \tag{3.10}$$

The direction cosines of largest local maximum curvature, X'', together with the maximum slope direction X', provide diagnostics additional to those obtained by perturbing probabilities above. Sampling properties of these direction cosines remain to be developed, but do not seem crucial at the diagnostic stage when identification is the principal aim. Warning limits of $\pm\frac{Z_\alpha}{\sqrt{n}}$ have also been suggested for informal calibration (Lawrance, 1988). It should be remarked that the diagnostics from the frequency perturbation approach are appropriate in the context of the independence hypothesis, whereas those based on probability perturbations have the advantage of being generally applicable to any null model.

3.1.3 Perturbation on likelihood ratio goodness-of-fit statistic

The above perturbation schemes can also be applied to the likelihood ratio statistic G^2. However, under probability perturbations,

$$\left.\frac{\partial G^2(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^T}\right|_{\boldsymbol{\omega}_0} \propto \langle x_{ij} \rangle^T ,$$

so that it does not provide any diagnostic information on the relative discordancy.

Formulae for the frequency perturbation approach are given below. For the likelihood ratio statistic

$$G^2 = 2 \sum_{i,j} x_{ij} \ln \left(\frac{x_{ij}}{e_{ij}} \right) ,$$

$$\left.\frac{\partial G^2(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^T}\right|_{\boldsymbol{\omega}_0} = G' = \langle G'_{ij} \rangle^T ,$$

where for a particular entry G'_{IJ},

$$G'_{IJ} = x_{IJ} \left\{ \ln \left(\frac{x_{IJ}}{e_{IJ}} \right) - x_{IJ} \left(\frac{1}{x_{I+}} + \frac{1}{x_{+J}} \right) \right\} .$$

For the second order approach, we also need $\frac{\partial^2 G^2(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^2}|_{\boldsymbol{\omega}_0}$, which elements are given by:

$$\begin{aligned}
\frac{\partial^2 G^2(\boldsymbol{\omega})}{\partial \omega_{IJ} \partial \omega_{IJ}} = {} & 2x_{IJ} \left\{ \frac{x_{IJ}}{x_{I+}^2} + \frac{x_{IJ}}{x_{+J}^2} + \frac{x_{IJ}}{x_{I+}x_{+J}} + \frac{1}{N} - \frac{x_{IJ}}{x_{I+}N} - \frac{x_{IJ}}{x_{+J}N} - \frac{1}{x_{I+}} - \frac{1}{x_{+J}} \right\} \\
& + \frac{x_{IJ}}{x_{I+}x_{+J}} \left\{ \frac{x_{IJ}}{N} - \frac{x_{IJ}}{x_{I+}} - \frac{x_{IJ}}{x_{+J}} + N \right\} \left\{ (x_{I+}x_{+J}) \left(x_{IJ} - \frac{x_{IJ}}{N} + 1 \right) \right\} \\
& + \sum_i x_{IJ}^2 x_{iJ} \left(\frac{2}{x_{+J}^2} - \frac{1}{x_{+J}N} - \frac{1}{x_{+J}} + \frac{1}{N} - \frac{1}{N^2} \right) \\
& + \sum_j x_{IJ}^2 x_{Ij} \left(\frac{2}{x_{I+}^2} - \frac{1}{x_{I+}N} - \frac{1}{x_{I+}} + \frac{1}{N} - \frac{1}{N^2} \right) \\
& - \sum_{i \neq I, j \neq J} x_{IJ}^2 \frac{x_{ij}}{N^2}
\end{aligned}$$

$$\frac{\partial^2 G^2(\omega)}{\partial \omega_{IJ} \partial \omega_{IK}} = x_{IJ} x_{IK} \left\{ \frac{1}{x_{I+}} - \frac{x_{I+}}{N^2} + \frac{2}{N} - \frac{2}{x_{I+}} - \sum_{i \neq I, j \neq J, K} \frac{x_{ij}}{N^2} \right\}$$
$$+ \sum_{i \neq I} x_{iJ} \left(\frac{1}{x_{+J}} - \frac{1}{N} - \frac{1}{x_{+J} N} \right) + x_{iK} \left(\frac{1}{x_{+K}} - \frac{1}{N} - \frac{1}{x_{+K} N} \right)$$

$$\frac{\partial^2 G^2(\omega)}{\partial \omega_{IJ} \partial \omega_{LJ}} = x_{IJ} x_{LJ} \left\{ \frac{1}{x_{+J}} - \frac{x_{+J}}{N^2} + \frac{2}{N} - \frac{2}{x_{+J}} - \sum_{j \neq J, i \neq I, L} \frac{x_{ij}}{N^2} \right\}$$
$$+ \sum_{j \neq J} x_{Ij} \left(\frac{1}{x_{I+}} - \frac{1}{N} - \frac{1}{x_{I+} N} \right) + x_{Lj} \left(\frac{1}{x_{L+}} - \frac{1}{N} - \frac{1}{x_{L+} N} \right)$$

$$\frac{\partial^2 G^2(\omega)}{\partial \omega_{IJ} \partial \omega_{LK}} = \frac{x_{IJ} x_{LK}}{N}$$

In general, we found that perturbation diagnostics derived from G^2 appear to be less sensitive than those of X^2. Therefore, we focus on the X^2 diagnostics in subsequent investigations.

3.1.4 An illustration

Table 3.1

Simonoff (1988) 5 × 5 artificial table

18	41	41	20	21
39	20	20	22	22
24	20	20	16	18
20	20	19	19	19
23	19	20	17	20

Simonoff (1988) invented a 5×5 contingency table with 3 outlying cells (1,2), (1,3) and (2,1) to illustrate the danger of *swamping*. The data set is reproduced in Table 3.1. The Pearson goodness-of-fit statistic is 24.64 on 16 degrees of freedom (p-value = 0.076). The overall fit is neither extremely good nor extremely bad.

Table 3.2

Adjusted and deleted residuals

-3.124	2.532	2.532	-0.977	-1.084
-3.695	3.646	3.646	-1.195	-1.331
2.866	-1.604	-1.604	0.349	-0.012
4.118	-1.938	-1.938	0.440	-0.014
0.595	-0.291	-0.291	-0.151	0.127
0.760	-0.358	-0.358	-0.182	0.155
-0.418	-0.234	-0.506	0.794	0.471
-0.513	-0.288	-0.616	0.988	0.582
0.267	-0.618	-0.348	0.096	0.652
0.337	-0.751	-0.428	0.116	0.814

Table 3.2 gives the adjusted (above) and deleted (below) residuals for each cell. The critical value obtained from the Bonferroni bound is 2.878 at $\alpha = 0.05$. The non-outlying cell (1,1) is wrongly identified for both types of residuals. Contrary to claims in the literature (see e.g. Barnett and Lewis (1994, p. 438)), we have found that the deleted residuals are still vulnerable to swamping.

Table 3.3 shows the normalized maximum slope direction (above) and maximum curvature direction (below) for individual cells, based on the probability perturbation approach. Using either $X^{(1)}$ or $X^{(2)}$, the first three cells identified are (2,1), (1,2) and (1,3), a correct result.

Table 3.3

Maximum slope direction $X^{(1)}$ and maximum curvature direction $X^{(2)}$

0.381	-0.456	-0.456	0.126	0.142
-0.001	0.440	0.440	0.000	0.000
-0.514	0.206	0.206	-0.048	0.002
-0.783	-0.001	-0.001	0.000	0.000
-0.085	0.038	0.038	0.018	-0.016
0.000	-0.000	-0.000	0.000	0.000
0.054	0.031	0.065	-0.104	-0.061
-0.000	-0.000	-0.000	0.000	0.000
-0.037	0.079	0.045	-0.012	-0.087
0.000	-0.000	-0.000	0.000	0.000

The extent of swamping effect on cell (1,1) is much less, especially for $X^{(2)}$. Similar results are obtained when perturbing X^2 through the observed frequencies, whereas results using G^2 are less sensitive.

3.2 Simulation study

In this section, we compare the performances of the outlier detection criteria, in a simulation study. The main concerns are:

1. How often are outlying cells identified as such (efficiency)?
2. How often are outlying cells not identified (masked)?
3. How often are non-outlying cells wrongly identified (swamped)?

In the simulations, tables of size 5×5 are generated using the S-plus `SAMPLE` function. Following Simonoff (1988), frequencies in contaminant cells are generated with slippage probability

$$p_{ij}^{*} = p_{ij}\left(1 + \Delta_{ij} N^{-\frac{1}{2}}\right)\left\{\sum_{(i,j)\in S} p_{ij}^{*}\right\}^{-1},$$

where the slippage constant, Δ_{ij}, is zero for non-contaminant cells. The null probabilities p_{ij} are uniform for a 5×5 table with $N = 500$. According to Kotze and Hawkins (1984), "if the restriction that the number of outlier cells is not more than a fixed proportion of all the cells does not hold, it may be necessary to investigate the table for other reasons for deviation from the independence hypothesis". Consequently, we create three discordant cells occupying the same positions as those of the artificial table in Table 3.1. Four alternatives to independence are considered:

(a) $\Delta_{12} = \Delta_{13} = \Delta_{21} = 30$;

(b) $\Delta_{12} = 40$, $\Delta_{13} = 30$, $\Delta_{21} = 20$;

(c) $\Delta_{12} = \Delta_{13} = \Delta_{21} = -20$;

(d) $\Delta_{12} = -20$, $\Delta_{13} = -20$, $\Delta_{21} = 40$.

Results based on 1,000 replications for each alternative configuration are presented in Table 3.4. The table provides estimates of the following performance measures:

β_1 = probability of exactly correct identification;

β_2 = probability of identifying all planted outliers;

β_3 = probability of at least one non-outlying cell is incorrectly identified;

N_C = average number of outlying cells identified;

N_I = average number of cells incorrectly identified.

Good performance of a criterion will be indicated by high values of β_1, β_2 and N_C but low values of β_3 and N_I. The perturbation approach is compared to the backwards-stepping procedure outlined in Simonoff (1988). The reference limit of the direction cosines is set at 0.386, which corresponds to discordancy testing at the nominal 5% level for a 5×5 table. Outcomes of the simulations show that estimates of β_1, β_2, and β_3 have standard errors typically less than 0.0157; estimates of N_C have standard errors less than 0.0252; whereas estimates of N_I have standard errors less than 0.0244.

It is evident from Table 3.4 that the probability perturbation approach is more efficient, as reflected by reasonably high β_1 probabilities, than the backwards-stepping use of the deleted residuals. Overall, the approach is less affected by the masking problem, as shown by the relatively stable β_2 estimates and N_C closest to 3. Moreover, it gives lower β_3 and N_I values (especially $X^{(2)}$), indicating lower levels of swamping. Simulations for the frequency perturbation approach, which have been omitted for brevity, yield results similar to those of probability perturbation. Our findings confirm that outlier detection based on residuals is vulnerable to both masking and swamping effects, and that neither of the residuals performs uniformly better than the other in the settings considered.

Table 3.4

Simulation results

		β_1	β_2	β_3	N_C	N_I
Alternative	(a)					
Backwards-stepping	$\tilde{r}$	0.025	0.025	0.037	1.082	0.044
	r^*	0.051	0.052	0.084	1.566	0.070
Perturbation	$X^{(1)}$	0.063	0.064	0.059	1.702	0.063
	$X^{(2)}$	0.057	0.057	0.023	1.589	0.023
Alternative	(b)					
Backwards-stepping	$\tilde{r}$	0.011	0.011	0.023	1.483	0.045
	r^*	0.030	0.031	0.031	1.429	0.086
Perturbation	$X^{(1)}$	0.046	0.051	0.003	1.723	0.023
	$X^{(2)}$	0.037	0.037	0.014	1.592	0.014
Alternative	(c)					
Backwards-stepping	$\tilde{r}$	0.067	0.071	0.267	1.598	0.196
	r^*	0.013	0.019	0.221	1.095	0.232
Perturbation	$X^{(1)}$	0.235	0.238	0.178	1.875	0.187
	$X^{(2)}$	0.039	0.039	0.061	1.131	0.145
Alternative	(d)					
Backwards-stepping	$\tilde{r}$	0.059	0.061	0.227	1.613	0.248
	r^*	0.012	0.017	0.201	1.464	0.229
Perturbation	$X^{(1)}$	0.035	0.036	0.223	1.454	0.243
	$X^{(2)}$	0.026	0.026	0.083	1.262	0.234

3.3 Testing for discordancy

Application of the identification criteria will provide a subset Q of possible outlying cells. To test a single cell for discordancy, measures such as $\chi^2 - \chi^2_{IJ}$ or $G^2 - G^2_{IJ}$ on omitting cell (I, J) are recommended (Barnett and Lewis, 1994, p. 435), where χ^2_{IJ} and G^2_{IJ} are respectively the Pearson and likelihood ratio statistics under the quasi-independence model H_1 with cell (I, J) excluded. Both test statistics may be compared to the $\chi^2_{(1)}$ reference distribution (Simonoff, 1988). As pointed out by Barnett and Lewis (1994), the consecutive testing of cells for discordancy, whether it is done on an *inward* or *outward* basis, can still be vulnerable to the masking problem. Although a number of multistage methods exist in the outliers detection literature (see e.g. Davies and Gather (1993) for a review), limited results are available for contingency tables. In the following, a confirmatory procedure is proposed to ascertain the relative discordancy of a group of outlying cells.

Lee and Fung (1997) devised an iterative diagnostic approach for the detection of multiple outliers in generalized linear models and nonlinear regressions. The essence of their approach is to form a clean subset (*included set*) of the data that is presumably free of outliers, and then assess the outlyingness of the remaining observations (*omitted set*) relative to the included set in a stepwise fashion. Unlike the regression setting, one cannot simply 'delete' any suspect cell(s) from a contingency table, some modifications of their procedure become necessary.

Algorithm

Step 1: Locate an initial omitted set Q of m candidate outliers using an appropriate identification criterion.

Step 2: Fit the quasi-independence model to the data without Q as identified in Step 1. Calculate the goodness-of-fit statistic T^Q (in either χ^2 or G^2 form).

Step 3: For each $(I, J) \in Q$, we put x_{IJ} separately back into the included set and refit the quasi-independence model. Evaluate goodness-of-fit statistics T^Q_{IJ} from each of the m tables of imputed values, then find the reductions $T^Q_{IJ} - T^Q$.

Step 4: Compare $T^Q_{IJ} - T^Q$ to the $\chi^2_{(1)}$ critical value. The corresponding cell that has been judged least insignificant is excluded from Q. The remaining cells form a new omitted set Q_* with size $m_* = m - 1$.

Step 5: Repeat Steps 2 to 4 in a stepwise manner until the discordancy tests in Step 4 are all significant. The cells in the eventual omitted set $\tilde{Q}$ are declared outliers.

In Step 4 of the algorithm, we choose to decrease the size of Q by one cell per step, which is more conservative than the approach adopted by Lee and Fung (1997). The critical value for the discordancy tests is also taken conservatively to be the nominal $\chi^2_{(1)}$ 1% cut-off of 6.635. A discussion on the true significant level of backward testing can be found in Simonoff (1988). Based on the evidence from the following two examples, it is recommended to use perturbation diagnostics for identifying the initial omitted set of suspect outliers, in order to minimize any potential masking effect. Finally, our experience with practical data sets indicates that the use of either χ^2 or G^2 form will make little difference in determining the eventual omitted set.

3.4 Examples

3.4.1. Student enrolments data

Table 3.5 presents a 7×8 contingency table of student enrolment figures from seven community schools in the Northern Territory, Australia. The data collections were conducted in eight different periods of the year. The Pearson χ^2 statistic is 85.72 and G^2 has the value 84.69 on 42 degrees of freedom, which implies strong evidence against the independence hypothesis.

Table 3.5

Student enrolments Data

93	96	99	99	147	144	87	87
138	141	141	201	189	153	135	114
42	45	42	48	54	48	45	45
63	63	72	66	78	78	93	63
60	60	54	51	51	45	39	36
174	165	156	156	153	150	156	159
78	69	84	78	54	66	78	78

†*source: data from Northern Territory Department of Education 1995 Student Census.*

Table 3.6 shows results of the identification criteria for selected cells of interest. The critical value obtained from the Bonferroni bound is 3.124 at $\alpha = 0.05$ for the residual diagnostics and the reference benchmark for the perturbation diagnostics is set at $1.96/\sqrt{n} = \pm 0.262$. Cells (1,5), (1,6), (2,4), and (7,5) are suggested to be potential outliers according to the residuals. The perturbation diagnostics $X^{(1)}$ and $X^{(2)}$ together identify the candidate set $Q = \{(1,5),(1,6),(2,4),(2,5)\}$. This same set is obtained under the frequency perturbation approach.

Table 3.6

Selected diagnostic results for student enrolments data

	$\tilde{r}$	r^*	$X^{(1)}$	$X^{(2)}$	X'	X''
(1,5)	3.16	3.87	-0.32	-0.20	-0.31	-0.18
(1,6)	3.66	4.50	-0.38	-0.24	-0.37	0.25
(2,4)	3.81	4.92	-0.44	0.86	-0.45	0.87
(2,5)	2.02	2.55	-0.23	-0.33	-0.24	-0.35
(7,5)	-3.42	-3.74	0.22	0.03	0.20	0.01

Table 3.7

Confirmatory analysis results for student enrolments data

$T_{IJ}^{Q_0} - T^{Q_0}$	$X^{(1)}, X^{(2)}$	$\tilde{r}, r^*$
(1,5)	21.46	10.02
(1,6)	17.14	17.41
(2,4)	16.84	13.12
(2,5)	13.25	—
(7,5)	—	10.43

We next apply the iterative algorithm to assess the relative discordancy of the suspect outliers. Table 3.7 presents the result of the confirmatory analysis. Both sets identified by the respective measures are confirmed to be discordant. Upon application of the backwards-stepping method via the change in G^2 (Simonoff 1988) also produces the discordant set (1,5), (1,6), (2,4), (7,5).

An inspection of the school records found that during collection periods 5 and 6, a group of transient seasonal fruit picker families have moved into the farming community near school 1, thus inflating this school's enrolments momentarily. Moreover, at collections 4 and 5, a significant population from a different region has migrated into the community of school 2 to attend a traditional funeral procession. As with most aboriginal funerals, the procession lasted for around three months and hence increased the enrolments of school 2 over those two periods. Apparently, cell (2,5) was not detected by the residuals due to masking in the identification stage, whereas the non-aberrant cell (7,5)

was judged discordant possibly owing to the absence of cell (2,5) in the initial omitted set. The perturbation approach appears to be resistant to such masking and swamping effects.

On fitting the quasi-independence model with the confirmed outlying set $\{(1,5),\ (1,6),\ (2,4),(2,5)\}$ results in $\chi^2 = 33.91$ and $G^2 = 34.27$, suggesting no departure from the quasi-independence assumption. The respective reductions in goodness-of-fit statistics are clearly significant on 4 degrees of freedom.

3.4.2 Archaeological data

As a second illustration of the confirmatory analysis, we consider the 19×6 contingency table from Mosteller and Parunak (1985), which gives the counts of 19 kinds of artifacts found at 6 distance categories from permanent water in southern Ruby Valley, Nevada. The data are reproduced in Table 3.8. The aim of the original study was to investigate the association between site types as defined by the kind of artifacts present, and the proximity to water. Following Mosteller and Parunak, the ordering of the distances is not considered.

The Pearson goodness-of-fit statistic is 190.03 on 90 degrees of freedom, indicating significant departure from independence. We prefer not to apply the likelihood ratio test due to the presence of sampling zeros in the table.

Table 3.8

Archaeological data †

20	102	54	38	29	3
33	136	86	58	56	7
27	122	68	51	53	0
2	10	8	5	4	0
11	82	34	35	30	2
10	53	25	17	17	3
39	185	88	100	58	13
34	179	70	78	60	11
26	78	24	26	14	6
24	88	32	41	26	3
8	44	16	28	39	3
15	75	30	35	27	8
11	32	5	11	21	2
12	28	5	18	7	0
2	10	4	2	6	0
3	8	4	6	8	0
1	2	0	3	3	0
13	5	3	9	7	0
20	36	19	20	28	1

† source: data from Casjens (1974)

Table 3.9

Selected diagnostic results for archaeological data

	$\tilde{r}$	r^*	$X^{(1)}$	$X^{(2)}$	X'	X''
(2,3)	2.95	3.63	-0.33	-0.10	-0.31	-0.09
(3,3)	1.86	2.23	-0.19	0.04	-0.19	0.05
(7,4)	1.92	2.37	-0.23	0.07	-0.22	0.07
(8,2)	1.26	1.78	-0.17	-0.87	-0.17	-0.85
(11,5)	4.48	5.49	-0.37	0.21	-0.38	0.18
(18,1)	5.379	6.81	-0.32	0.11	-0.30	0.09
(18,2)	-3.16	-3.41	-0.09	0.01	-0.09	0.00

Table 3.9 shows results of the identification criteria for selected cells of interest. The critical value obtained from the Bonferroni bound is 3.327 at $\alpha = 0.05$ (Fushs and Kennett, 1980). The reference points for perturbation diagnostics are taken to be $1.96/\sqrt{n} = \pm 0.184$. Cells (11,5) and (18,1) have the most extreme adjusted residuals. The deleted residuals also bring attention to cells (2,3) and (18,2) in addition. With respect to the perturbation diagnostics, the candidate set $Q = \{(2,3),(3,3),(7,4),(8,2),(11,5),(18,1)\}$ is obtained.

If we apply the algorithm to the set $\{(2,3),(11,5),(18,1),(18,2)\}$ from the residuals, the first three cells are retained as outlying. The same result is obtained by the backwards-stepping method. The candidate set Q from the perturbation approach is next considered. Table 3.10 lists the sequential changes in χ^2 during successive fittings of the quasi-independence models.

The eventual discordant set obtained is $\tilde{Q} = \{(2,3),(11,5),(18,1)\}$. Upon fitting the quasi-independence model with $\tilde{Q}$ identified as such, the resultant drop in $\chi^2 = 52.22$ is highly significant on 3 degrees of freedom. It is interesting to remark that in their original analysis of the data, Mosteller and Parunak also found (11,5) and (18,1) discordant, with detailed justification provided. Although an archaeological interpretation has been given for cell (2,3), their exploratory and residual-based methods apparently rejected this somewhat marginal outlier.

Table 3.10

Confirmatory analysis results for archaeological data

	$T_{IJ}^{Q} - T^{Q}$	$T_{IJ}^{Q_1} - T^{Q_1}$	$T_{IJ}^{Q_2} - T^{Q_2}$	$T_{IJ}^{Q_3} - T^{Q_3}$
(2,3)	8.77	9.15	9.60	8.03
(3,3)	3.75	3.98	4.24	—
(7,4)	2.53	2.74	—	—
(8,2)	0.92	—	—	—
(11,5)	17.06	17.33	17.70	18.08
(18,1)	24.18	24.33	24.66	24.89

3.5 Discussion

The perturbation approach is preferable to identification based on residuals because it provides a more reliable set of candidate outliers and is more effective in dealing with the masking and swamping problems. Another advantage of the approach is that it only requires the independence assumption and does not rely on the explicit formulation of a parametric model which is often required by other diagnostic methods. The stepwise confirmatory procedure is also shown to be effective.

CHAPTER 4

COVARIATE TRANSFORMATION DIAGNOSTICS IN GENERALIZED LINEAR MODELS

4. COVARIATE TRANSFORMATION DIAGNOSTICS IN GLM

Transformations of variables have often been applied to data in statistical modeling. Parametric transformation families, such as the Box-Cox power transformation, is commonly used. Various diagnostics have been proposed to assess the sensitivity of the maximum likelihood estimate (MLE) of the transformation parameter; see e.g. Cook and Wang (1983), Atkinson (1986, 1988), Wang (1987), Tsai and Wu (1990). Most of these methods are concerned with transformation of the response or simultaneous transformation (transform-both-sides model). Transformation diagnostics for the covariates, however, have been studied to a lesser extent. Ezekiel and Fox (1959) introduced the partial residual plot. Box and Tidwell (1962) suggested constructed variables and added variable plots to assist the selection of suitable transformations for covariates. A review of such procedures can be found in Cook and Weisberg (1982) and Chatterjee and Hadi (1988).

Traditionally, transformation diagnostics are derived using the case deletion approach (Cook and Weisberg (1982), Wei and Shih (1994a)). Since Cook (1986) developed the local influence methodology as a general tool for assessing the effect of small departures from model assumptions, there is a large body of literature dealing with response transformation and simultaneous transformations based on this approach (Lawrance (1988), Hinkley and Wang (1988), Tsai and Wu (1992), Shih (1993), Wei and Shih (1994b), Shih and Wei (1995)). In contrast, limited diagnostics are available for analyzing the transformation of covariates. Cook (1987) used a subset formula from local influence to derive diagnostics for partially nonlinear models, which include transformation of a single covariate as a special case. Wei and Hickernell (1996) considered further extensions to several covariates based on profile likelihood displacement and found

that their diagnostics are related to those of Cook (1987). Nevertheless, all of the above methods are devoted exclusively to the linear regression setting.

In this chapter we present local influence diagnostics for assessing the effect of minor perturbations on the MLE of the covariate transformation parameters in generalized linear models. Two separate approaches based on analysis of the transformation parameter surface or profile likelihood displacement, and partial influence, are proposed in the next section. Specific perturbation schemes are outlined in Section 4.2 to examine the different aspects of influence. Three illustrative examples are provided in Section 4.3.

4.1 Profile likelihood displacement and local influence

We assume the responses $\boldsymbol{y} = \langle y_1, \cdots, y_n \rangle^T$ have a density or mass function of the form

$$f_y(y_i; \boldsymbol{\theta}) = \exp\left\{ \left[y_i \theta_i - b(\theta_i) \right] / \mathtt{a}(\phi) + c(y_i, \phi) \right\}$$

with $\theta_i = k(\eta_i)$, where η_i is the linear predictor and $\mathtt{a}(\cdot)$, $b(\cdot)$, $c(\cdot)$ are known functions. Without loss of generality the dispersion parameter ϕ is assumed known or may be replaced by an estimate $\hat{\phi}$ and write $\hat{\mathtt{a}} = \mathtt{a}(\hat{\phi})$, which gives an exponential-family density with natural parameter $\boldsymbol{\theta}$. The log-likelihood function is then

$$\hat{\mathtt{a}}^{-1} \sum_{i=1}^{n} \left[y_i k(\eta_i) - b\{k(\eta_i)\} \right] .$$

Goodness-of-fit of a generalized linear model may often be improved by transforming one or more covariates $\boldsymbol{z}$ of $X = \langle \boldsymbol{x}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}_{(1)}, \cdots, \boldsymbol{x}_{(p)}, \boldsymbol{z}_{(1)}, \cdots, \boldsymbol{z}_{(q)} \rangle$. Let the linear predictor of the transformation model be

$$\boldsymbol{\eta} = \boldsymbol{x}\boldsymbol{\delta} + G(\boldsymbol{z}, \boldsymbol{\lambda})\boldsymbol{\xi} \ ,$$

where the $n \times q$ matrix $G(\boldsymbol{z}, \boldsymbol{\lambda}) = \langle g_1(\boldsymbol{z}_{(1)}, \lambda_1), \cdots, g_q(\boldsymbol{z}_{(q)}, \lambda_q) \rangle$, and

$$g_j(\boldsymbol{z}_{(j)}, \lambda_j) = \langle g_j(z_{1j}, \lambda_j), \cdots, g_j(z_{nj}, \lambda_j) \rangle^T$$

represents a known, twice continuously differentiable transformation family indexed by λ_j $(j = 1, \cdots, q)$. Here the parameter vector $\boldsymbol{\lambda}$ is of special interest.

Let $\tilde{\boldsymbol{\delta}}(\boldsymbol{\lambda})$, $\tilde{\boldsymbol{\xi}}(\boldsymbol{\lambda})$ be functions that maximize $L(\boldsymbol{\lambda}; \boldsymbol{\delta}, \boldsymbol{\xi})$ for fixed $\boldsymbol{\lambda}$ and denote the corresponding profile log-likelihood for $\boldsymbol{\lambda}$ by $L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}(\boldsymbol{\lambda}), \tilde{\boldsymbol{\xi}}(\boldsymbol{\lambda}))$. To assess the global influence of individual cases on the MLE $\hat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$, one can adopt the case deletion approach of Cook and Weisberg (1982). The difference between $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\lambda}}_{[i]}$, the MLE of $\boldsymbol{\lambda}$ without case i, can be measured through the profile likelihood displacement

$$LD_i = 2[L(\hat{\boldsymbol{\lambda}}) - L(\hat{\boldsymbol{\lambda}}_{[i]})] \tag{4.1}$$

where $L(\boldsymbol{\lambda}) = L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}(\boldsymbol{\lambda}), \tilde{\boldsymbol{\xi}}(\boldsymbol{\lambda}))$. For the special case of linear regression, Wei and Hickernell (1996) obtained approximations to simplify computations of LD_i. A large value of LD_i indicates that $\hat{\boldsymbol{\lambda}}$ is likely to be dependent on case i.

4.1.1 First and second order approach

For the MLE surface of the transformation parameter $\boldsymbol{\lambda}$, we introduce small changes into our model through an $n \times 1$ vector $\boldsymbol{\omega} \in \Omega$, where Ω denotes the open set of relevant perturbations. Suppose there is a null point $\boldsymbol{\omega}_0$ in Ω representing no perturbation so that $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}_0} = \hat{\boldsymbol{\lambda}}$. In the manner of Wu and Luo (1993a), the MLE surface of $\boldsymbol{\lambda}$ is the geometric influence graph formed by $\boldsymbol{\alpha}(\boldsymbol{\omega}) = \langle \boldsymbol{\omega}^T, \hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}} \rangle$, where $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}$ is the estimate of $\boldsymbol{\lambda}$ under perturbation $\boldsymbol{\omega}$.

To find the direction of largest local change, we approximate the MLE surface by its tangent plane at $\boldsymbol{\omega}_0$, which is determined by $\frac{\partial \hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}}{\partial \boldsymbol{\omega}^T}$ at $\boldsymbol{\omega}_0$. The direction of largest local change is just the direction of maximum slope on this tangent plane over Ω. The derivations below are similar to those of Shih (1993, p.414) for the transform-both-sides model. Write $L(\boldsymbol{\lambda}|\boldsymbol{\omega}) = L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}(\boldsymbol{\lambda}|\boldsymbol{\omega}), \tilde{\boldsymbol{\xi}}(\boldsymbol{\lambda}|\boldsymbol{\omega})|\boldsymbol{\omega})$ for the profile log-likelihood corresponding to the perturbed model, where $\tilde{\boldsymbol{\delta}}(\boldsymbol{\lambda}|\boldsymbol{\omega})$ and $\tilde{\boldsymbol{\xi}}(\boldsymbol{\lambda}|\boldsymbol{\omega})$ are functions that maximize $L(\boldsymbol{\lambda}; \boldsymbol{\delta}, \boldsymbol{\xi}|\boldsymbol{\omega})$ for fixed $\boldsymbol{\lambda}$ and $\boldsymbol{\omega}$. Then $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}$ satisfies the following equation:

$$\frac{\partial L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \boldsymbol{\lambda}} = \mathbf{0} \, .$$

Differentiating with respect to ω_i yields

$$\frac{\partial^2 L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T} \left(\frac{\partial \boldsymbol{\lambda}_{\boldsymbol{\omega}}}{\partial \omega_i} \right) + \frac{\partial^2 L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \omega_i \partial \boldsymbol{\lambda}^T} = \mathbf{0} \, .$$

Therefore,

$$\frac{\partial \boldsymbol{\lambda}_{\boldsymbol{\omega}}}{\partial \omega_i} = \left[\frac{\partial^2 L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T} \right]^{-1} \left(\frac{-\partial^2 L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial \omega_i \partial \boldsymbol{\lambda}^T} \right) \, . \tag{4.2}$$

Let $L_1^{(1)}$ and $L_2^{(1)}$ be the derivatives of $L(\boldsymbol{\lambda}; \boldsymbol{\delta}, \boldsymbol{\xi}|\boldsymbol{\omega})$ with respect to $\boldsymbol{\lambda}$ and $(\boldsymbol{\delta}, \boldsymbol{\xi})$ respectively, with superscript (t) denoting the tth derivative of the function. It

can be shown that

$$\frac{\partial^2 L(\boldsymbol{\lambda};\tilde{\boldsymbol{\delta}},\tilde{\boldsymbol{\xi}}|\boldsymbol{\omega})}{\partial\omega_i\partial\boldsymbol{\lambda}^T} = L^{(2)}_{1\omega i} + L^{(2)}_{12}\frac{\partial(\tilde{\boldsymbol{\delta}}(\boldsymbol{\lambda}|\boldsymbol{\omega}),\tilde{\boldsymbol{\xi}}(\boldsymbol{\lambda}|\boldsymbol{\omega}))}{\partial\omega_i} = L^{(2)}_{1\omega i} - L^{(2)}_{12}\left[L^{(2)}_{22}\right]^{-1}L^{(2)}_{2\omega i}\ .$$

The partitions of $L^{(2)}$ are

$$L^{(2)}_{11} = -\sum_{i=1}^{n} S2_i G^{(1)}(\boldsymbol{z}_i,\boldsymbol{\lambda})^T\boldsymbol{\xi}^T\boldsymbol{\xi}G^{(1)}(\boldsymbol{z}_i,\boldsymbol{\lambda}) + S1_i diag(\boldsymbol{\xi})G^{(2)}(\boldsymbol{z}_i,\boldsymbol{\lambda})$$

$$L^{(2)}_{12} = \sum_{i=1}^{n}\langle S2_i diag(\boldsymbol{\xi})G^{(1)}(\boldsymbol{z}_i,\boldsymbol{\lambda})^T\boldsymbol{x}_i,$$
$$S2_i G(\boldsymbol{z}_i,\boldsymbol{\lambda})^T G^{(1)}(\boldsymbol{z}_i,\boldsymbol{\lambda})diag(\boldsymbol{\xi}) + S1_i diag(G^{(1)}(\boldsymbol{z}_i,\boldsymbol{\lambda}))\rangle$$

$$L^{(2)}_{21} = [L^{(2)}_{12}]^T$$

$$L^{(2)}_{22} = \sum_{i=1}^{n} S2_i\,\langle\boldsymbol{x}_i, G(\boldsymbol{z}_i,\boldsymbol{\lambda})\rangle^T\langle\boldsymbol{x}_i, G(\boldsymbol{z}_i,\boldsymbol{\lambda})\rangle$$

where $\boldsymbol{x}_i = \langle x_{i1},\cdots,x_{ip}\rangle$ denotes the ith row of $\boldsymbol{x}$,

$$G^{(1)}(\boldsymbol{z}_i,\boldsymbol{\lambda}) = \left\langle\frac{g_1(z_{i1},\lambda_1)}{\partial\lambda_1},\cdots,\frac{g_q(z_{iq},\lambda_q)}{\partial\lambda_q}\right\rangle$$

is a $1\times q$ vector,

$$G^{(2)}(\boldsymbol{z}_i,\boldsymbol{\lambda}) = \frac{\partial G(\boldsymbol{z}_i,\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^T}$$

is a $q\times q$ matrix,

$$S1_i = y_i k^{(1)}(\eta_i) - b^{(1)}(k(\eta_i))k^{(1)}(\eta_i)\ ,$$

$$S2_i = y_i k^{(2)}(\eta_i) - \{b^{(2)}(k(\eta_i))[k^{(1)}(\eta_i)]^2 + b^{(1)}(k(\eta_i))k^{(2)}(\eta_i)$$

while $L^{(2)}_{1\omega i} = \dfrac{\partial^2 L(\boldsymbol{\lambda};\boldsymbol{\delta},\boldsymbol{\xi}|\boldsymbol{\omega})}{\partial\omega_i\partial\boldsymbol{\lambda}^T}$ and $L^{(2)}_{2\omega i} = \dfrac{\partial^2 L(\boldsymbol{\lambda};\boldsymbol{\delta},\boldsymbol{\xi}|\boldsymbol{\omega})}{\partial\omega_i\partial(\boldsymbol{\delta},\boldsymbol{\xi})^T}$ are entries on the corresponding columns of

$$\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\boldsymbol{\omega}};\boldsymbol{\delta}_{\boldsymbol{\omega}},\boldsymbol{\xi}_{\boldsymbol{\omega}}) = \frac{\partial^2 L(\boldsymbol{\lambda};\boldsymbol{\delta},\boldsymbol{\xi}|\boldsymbol{\omega})}{\partial\boldsymbol{\omega}\partial(\boldsymbol{\lambda};\boldsymbol{\delta},\boldsymbol{\xi})^T} \tag{4.3}$$

and further derived for various types of perturbations in Section 4.2. All of the above quantities are evaluated at $\boldsymbol{\omega}_0$ and $\hat{\boldsymbol{\lambda}}$.

To compute the maximum slope direction at the null point, $\boldsymbol{l}^{\hat{\boldsymbol{\lambda}}}_{slope}$, we evaluate $\dfrac{\partial \hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}}{\partial \boldsymbol{\omega}^T}$ at $\boldsymbol{\omega}_0$ and $\hat{\boldsymbol{\lambda}}$. For second order local influence, the direction $\boldsymbol{l}^{\hat{\boldsymbol{\lambda}}}_{max}$ which corresponds to the maximum normal curvature of the MLE surface is the main diagnostic quantity to study the combined effect on $\hat{\boldsymbol{\lambda}}$. As stated in Chapter 2, Section 2.2, the local maximum curvatures and the corresponding directions are the eigen value-vector solution of the equation

$$|F - \boldsymbol{\lambda} M| = 0$$

where the matrices

$$F = \frac{\partial^2 \hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^T} = \left[\frac{\partial \hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}}{\partial \boldsymbol{\omega}^T}\right]^T \left(\frac{\partial^2 L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\xi}} | \boldsymbol{\omega})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T}\right) \left[\frac{\partial \hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}}{\partial \boldsymbol{\omega}^T}\right] \tag{4.4}$$

and

$$M = I + \frac{\partial \hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}}{\partial \boldsymbol{\omega}^T} \left[\frac{\partial \hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}}{\partial \boldsymbol{\omega}^T}\right]^T$$

are evaluated at $\boldsymbol{\omega}_0$. Alternatively, we can apply the subset formulation of Cook (1987) to the profile likelihood displacement

$$LD(\boldsymbol{\omega}) = 2\left[L(\hat{\boldsymbol{\lambda}}) - L(\hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}})\right] . \tag{4.5}$$

An equivalent form to (4.4) is then

$$\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\boldsymbol{\omega}}; \boldsymbol{\delta}_{\boldsymbol{\omega}}, \boldsymbol{\xi}_{\boldsymbol{\omega}}) \left\{[L^{(2)}]^{-1} - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & [L^{(2)}_{22}]^{-1} \end{pmatrix}\right\} \boldsymbol{\Delta}^T(\boldsymbol{\lambda}_{\boldsymbol{\omega}}; \boldsymbol{\delta}_{\boldsymbol{\omega}}, \boldsymbol{\xi}_{\boldsymbol{\omega}}) . \tag{4.6}$$

If λ is a scalar parameter (transformation of a single covariate, i.e. $q = 1$), then $\frac{\partial^2 L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\xi}} | \boldsymbol{\omega})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T}$ becomes a scalar quantity so that $\boldsymbol{l}^{\hat{\lambda}}_{max}$ is proportional to $\boldsymbol{l}^{\hat{\lambda}}_{slope} = (\partial \hat{\lambda}_{\boldsymbol{\omega}} / \partial \boldsymbol{\omega}^T)^T$ and

$$\frac{\partial \hat{\lambda}_{\boldsymbol{\omega}}}{\partial \boldsymbol{\omega}^T} \propto L^{(2)}_{1\boldsymbol{\omega}} - L^{(2)}_{12} \left[L^{(2)}_{22}\right]^{-1} L^{(2)}_{2\boldsymbol{\omega}} , \tag{4.7}$$

4.1.2 Partial influence approach

Let the linear predictor before transformation of covariate z be given by

$$\eta_0 = x\beta + z\gamma \, . \tag{4.8}$$

The model after transformation is

$$\eta = x\delta + G(z, \lambda)\xi \, , \tag{4.9}$$

so that $G(z, \lambda_0) = z$ represents no transformation. A test of the hypothesis H_0: $\lambda = \lambda_0$ can be based on $D_0 - D$, the reduction in deviance from model (4.8) to model (4.9). It is important to judge whether any particular observation has an undue impact on this test. Denote $D_{0[i]}$ and $D_{[i]}$ for the deviances of (4.8) and (4.9) respectively after deleting case i. In the manner of Lee (1988), a partial influence measure for the impact of case i on the transformation can be formulated as

$$d_i = (D_0 - D) - (D_{0[i]} - D_{[i]}) \, , \tag{4.10}$$

which represents the change in deviance due to the transformation of z when the i-th observation is excluded.

Consider the log-likelihood $L(\lambda; \delta, \xi)$ of the transformation model (4.9). The full MLE of λ, δ, ξ are denoted by $\hat{\lambda}$, $\hat{\delta}$, $\hat{\xi}$ respectively. Similarly, let $L_0(\beta, \gamma)$ be the log-likelihood of model (4.8), with MLEs $\hat{\beta}$ and $\hat{\gamma}$. Under minor perturbations, the respective log-likelihoods become $L(\lambda; \delta, \xi|\omega)$ and $L_0(\beta, \gamma|\omega)$, with associated MLEs $(\hat{\lambda}_\omega, \hat{\delta}_\omega, \hat{\xi}_\omega)$ and $(\hat{\beta}_\omega, \hat{\gamma}_\omega)$. Suppose that $L(\lambda; \delta, \xi|\omega_0) = L(\lambda; \delta, \xi)$ and $L_0(\beta, \gamma|\omega_0) = L_0(\beta, \gamma)$. The partial influence on the transformation due to perturbation ω can be assessed by

$$d(\omega) = 2\left\{\left[L_0(\hat{\beta}, \hat{\gamma}) - L(\hat{\lambda}; \hat{\delta}, \hat{\xi})\right] - \left[L_0(\hat{\beta}_\omega, \hat{\gamma}_\omega) - L(\hat{\lambda}_\omega; \hat{\delta}_\omega, \hat{\xi}_\omega)\right]\right\} \, . \tag{4.11}$$

Analogous to (4.10) in case-deletion, the log-likelihood displacement $d(\omega)$ measures the local effect on the transformation parameter with respect to the contours of the unperturbed deviance reduction.

Let $F = \frac{1}{2}d(\omega)$. We obtain the normal curvature at $F(\omega_0)$ along the direction $\boldsymbol{l}$ as

$$C(\boldsymbol{l}) = 2|\boldsymbol{l}^T F^{(2)} \boldsymbol{l}| \ ,$$

$$F^{(2)} = \frac{\partial^2 L(\hat{\boldsymbol{\lambda}}_\omega; \hat{\boldsymbol{\delta}}_\omega, \hat{\boldsymbol{\xi}}_\omega)}{\partial\omega\partial\omega^T} - \frac{\partial^2 L_0(\hat{\boldsymbol{\beta}}_\omega, \hat{\boldsymbol{\gamma}}_\omega)}{\partial\omega\partial\omega^T} \ .$$

Applying the chain rule of differentiation yields

$$\begin{aligned} F^{(2)} =& \boldsymbol{\Delta}(\boldsymbol{\lambda}_\omega; \boldsymbol{\delta}_\omega, \boldsymbol{\xi}_\omega) \left[L^{(2)}(\boldsymbol{\lambda}; \boldsymbol{\delta}, \boldsymbol{\xi})\right]^{-1} \boldsymbol{\Delta}^T(\boldsymbol{\lambda}_\omega; \boldsymbol{\delta}_\omega, \boldsymbol{\xi}_\omega) \\ & - \nabla(\boldsymbol{\beta}_\omega, \boldsymbol{\gamma}_\omega) \left[L_0^{(2)}(\boldsymbol{\beta}, \boldsymbol{\gamma})\right]^{-1} \nabla^T(\boldsymbol{\beta}_\omega, \boldsymbol{\gamma}_\omega) \end{aligned} \tag{4.12}$$

where $\boldsymbol{\Delta}(\boldsymbol{\lambda}_\omega; \boldsymbol{\delta}_\omega, \boldsymbol{\xi}_\omega)$ is defined in (4.3),

$$\nabla(\boldsymbol{\beta}_\omega, \boldsymbol{\gamma}_\omega) = \frac{\partial^2 L_0(\boldsymbol{\beta}, \boldsymbol{\gamma}|\omega)}{\partial\omega\partial(\boldsymbol{\beta}, \boldsymbol{\gamma})^T} \ ,$$

evaluated at ω_0, $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\delta}}$, $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\lambda}}$. Expressions for $L_0^{(2)}(\boldsymbol{\beta}, \boldsymbol{\gamma})$ and $\nabla(\boldsymbol{\beta}_\omega, \boldsymbol{\gamma}_\omega)$ are derived under the original generalized linear model (4.8), which can be found in Thomas and Cook (1989). Let $\boldsymbol{l}_{max}^d$ be the direction cosines of maximum normal curvature, which is the perturbation direction that produces the greatest local change in $\boldsymbol{\lambda}$ as measured by (4.11). Similar to the derivation in Chapter 2 Section 2.2, $\boldsymbol{l}_{max}^d$ is just the eigenvector associated with the largest eigenvalue of

$$\boldsymbol{\Delta}[L^{(2)}]^{-1}\boldsymbol{\Delta}^T - \nabla[L_0^{(2)}]^{-1}\nabla^T \ . \tag{4.13}$$

The most influential elements of the data on the transformation may be identified by their large components of $\boldsymbol{l}_{max}^d$. We also recommend to plot $\hat{\boldsymbol{\lambda}}_\omega$ against the perturbation scale for each local direction $\boldsymbol{l}$ of interest. The characteristics of such curves should be informative for further investigation on the relationship between local and global influences.

4.2 Perturbation schemes

A number of perturbation schemes have been suggested to examine the different aspects of influence (Cook (1987), Thomas and Cook (1989), Wu and Wan (1994)). In the following, we consider relevant perturbation schemes and derive the corresponding $\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\boldsymbol{\omega}}; \boldsymbol{\delta}_{\boldsymbol{\omega}}, \boldsymbol{\xi}_{\boldsymbol{\omega}})$ quantity.

4.2.1 Perturbation of case weights

We define a vector of weights $\boldsymbol{\omega} = \langle \omega_1, \cdots, \omega_n \rangle^T$, $\omega_i \geq 0$, to perturb the contribution of each case to the log- likelihood. The point representing no perturbation is $\boldsymbol{\omega}_0 = \langle 1, \cdots, 1 \rangle^T$. We obtain $\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\boldsymbol{\omega}}; \boldsymbol{\delta}_{\boldsymbol{\omega}}, \boldsymbol{\xi}_{\boldsymbol{\omega}}) = \langle \Delta_i^w \rangle$ evaluated at $\boldsymbol{\omega}_0$ and $\hat{\boldsymbol{\lambda}}$, where

$$\Delta_i^w = S1_i \left\langle \boldsymbol{x}_i, G(\boldsymbol{z}_i, \boldsymbol{\lambda}), G^{(1)}(\boldsymbol{z}_i, \boldsymbol{\lambda}) diag(\boldsymbol{\xi}) \right\rangle \ . \tag{4.14}$$

The case weight perturbation scheme actually generalizes case deletion, where ω_i is limited to the values 0 or 1. Furthermore, if the deletion of the ith case is of interest (as revealed by $\boldsymbol{l}_{max}$), it may be considered as the perturbation located in direction $\boldsymbol{l}_{[i]}$ from the null point, where $\boldsymbol{l}_{[i]}$ is the direction cosines with i-th component -1 but zeros elsewhere (Wu and Luo (1993a, 1993b)). A plot of $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}$ in the direction $\boldsymbol{l}_{[i]}$ can then monitor the global effects of downweighting the ith case.

4.2.2 Perturbation of individual covariates $\boldsymbol{x}$

We modify the j-th individual covariate, $\boldsymbol{x}_{(j)}$ of $\boldsymbol{x}$, to $\boldsymbol{x}_{(j)}(\boldsymbol{\omega}) = \boldsymbol{x}_{(j)} + t\boldsymbol{\omega}$, as long as the covariate is not an indicator variable. Here, t is the scaling factor used to convert the generic perturbation $\boldsymbol{\omega}$ to the appropriate size and units, and $\boldsymbol{\omega}_0 = \langle 0, \cdots, 0 \rangle^T$ represents no perturbation. It can be verified that $\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\boldsymbol{\omega}}; \boldsymbol{\delta}_{\boldsymbol{\omega}}, \boldsymbol{\xi}_{\boldsymbol{\omega}}) = \langle \Delta_i^x \rangle$ where

$$\Delta_i^x = S2_i \delta_j \left\langle \boldsymbol{x}_i, G(\boldsymbol{z}_i, \boldsymbol{\lambda}) + S1_i \boldsymbol{u}_j, G^{(1)}(\boldsymbol{z}_i, \boldsymbol{\lambda}) diag(\boldsymbol{\xi}) \right\rangle \ , \tag{4.15}$$

4.2 Perturbation schemes

A number of perturbation schemes have been suggested to examine the different aspects of influence (Cook (1987), Thomas and Cook (1989), Wu and Wan (1994)). In the following, we consider relevant perturbation schemes and derive the corresponding $\boldsymbol{\Delta}(\boldsymbol{\lambda}_\omega; \boldsymbol{\delta}_\omega, \boldsymbol{\xi}_\omega)$ quantity.

4.2.1 Perturbation of case weights

We define a vector of weights $\boldsymbol{\omega} = \langle \omega_1, \cdots, \omega_n \rangle^T$, $\omega_i \geq 0$, to perturb the contribution of each case to the log- likelihood. The point representing no perturbation is $\boldsymbol{\omega}_0 = \langle 1, \cdots, 1 \rangle^T$. We obtain $\boldsymbol{\Delta}(\boldsymbol{\lambda}_\omega; \boldsymbol{\delta}_\omega, \boldsymbol{\xi}_\omega) = \langle \Delta_i^w \rangle$ evaluated at $\boldsymbol{\omega}_0$ and $\hat{\boldsymbol{\lambda}}$, where

$$\Delta_i^w = S1_i \left\langle \boldsymbol{x}_i, G(\boldsymbol{z}_i, \boldsymbol{\lambda}), G^{(1)}(\boldsymbol{z}_i, \boldsymbol{\lambda}) diag(\boldsymbol{\xi}) \right\rangle \ . \tag{4.14}$$

The case weight perturbation scheme actually generalizes case deletion, where ω_i is limited to the values 0 or 1. Furthermore, if the deletion of the ith case is of interest (as revealed by $\boldsymbol{l}_{max}$), it may be considered as the perturbation located in direction $\boldsymbol{l}_{[i]}$ from the null point, where $\boldsymbol{l}_{[i]}$ is the direction cosines with i-th component -1 but zeros elsewhere (Wu and Luo (1993a, 1993b)). A plot of $\hat{\boldsymbol{\lambda}}_\omega$ in the direction $\boldsymbol{l}_{[i]}$ can then monitor the global effects of downweighting the ith case.

4.2.2 Perturbation of individual covariates $\boldsymbol{x}$

We modify the j-th individual covariate, $\boldsymbol{x}_{(j)}$ of $\boldsymbol{x}$, to $\boldsymbol{x}_{(j)}(\boldsymbol{\omega}) = \boldsymbol{x}_{(j)} + t\boldsymbol{\omega}$, as long as the covariate is not an indicator variable. Here, t is the scaling factor used to convert the generic perturbation $\boldsymbol{\omega}$ to the appropriate size and units, and $\boldsymbol{\omega}_0 = \langle 0, \cdots, 0 \rangle^T$ represents no perturbation. It can be verified that $\boldsymbol{\Delta}(\boldsymbol{\lambda}_\omega; \boldsymbol{\delta}_\omega, \boldsymbol{\xi}_\omega) = \langle \Delta_i^x \rangle$ where

$$\Delta_i^x = S2_i \delta_j \left\langle \boldsymbol{x}_i, G(\boldsymbol{z}_i, \boldsymbol{\lambda}) + S1_i \boldsymbol{u}_j, G^{(1)}(\boldsymbol{z}_i, \boldsymbol{\lambda}) diag(\boldsymbol{\xi}) \right\rangle \ , \tag{4.15}$$

δ_j being the regression coefficient associated with $\boldsymbol{x}_{(j)}$ and $\boldsymbol{u}_j$ denotes a $1 \times q$ row vector with j-th component 1 but zeros elsewhere.

4.2.3 Perturbation of individual transformed covariates z

We perturb the j-th transformed covariate $\boldsymbol{z}_{(j)}$ to $\boldsymbol{z}_{(j)}(\boldsymbol{\omega}) = \boldsymbol{z}_{(j)} + t\boldsymbol{\omega}$. Again, t is the appropriate scaling factor and $\boldsymbol{\omega}_0 = \mathbf{0}$ indicates no perturbation. We find that $\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\boldsymbol{\omega}}; \boldsymbol{\delta}_{\boldsymbol{\omega}}, \boldsymbol{\xi}_{\boldsymbol{\omega}}) = \langle \Delta_i^z \rangle$, where

$$\begin{aligned} \Delta_i^z = \Big\langle & S2_i \xi_j \boldsymbol{x}_i, S2_i \xi_j G(\boldsymbol{z}_i, \boldsymbol{\lambda}) + S1_i G^{(1)}(z_{ij}, \boldsymbol{\lambda}) \boldsymbol{u}_j, \\ & S2_i \xi_j G^{(1)}(\boldsymbol{z}_i, \boldsymbol{\lambda}) diag(\boldsymbol{\xi}) + S1_i \xi_j G^{(2)}(z_{ij}, \boldsymbol{\lambda}) \boldsymbol{u}_j \Big\rangle \end{aligned} \tag{4.16}$$

with ξ_j being the regression coefficient associated with $\boldsymbol{z}_{(j)}$, and $G^{(1)}(z_{ij}, \boldsymbol{\lambda})$ is the j-th entry of $G^{(1)}(\boldsymbol{z}_i, \boldsymbol{\lambda})$, $G^{(2)}(z_{ij}, \boldsymbol{\lambda})$ is the (j, j)-th entry of $G^{(2)}(\boldsymbol{z}_i, \boldsymbol{\lambda})$.

4.2.4 Perturbation of responses

We then consider altering the responses by taking $\boldsymbol{y}(\boldsymbol{\omega}) = \boldsymbol{y} + t\boldsymbol{\omega}$ where $t = diag\{[\hat{a}b^{(2)}(k(\eta_i))]^{\frac{1}{2}}\}$. As with other additive perturbations, $\boldsymbol{\omega}_0 = \mathbf{0}$ gives the unperturbed state. It follows that $\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\boldsymbol{\omega}}; \boldsymbol{\delta}_{\boldsymbol{\omega}}, \boldsymbol{\xi}_{\boldsymbol{\omega}}) = \langle \Delta_i^y \rangle$, where

$$\Delta_i^y = k^{(1)}(\eta_i) \left\langle \boldsymbol{x}_i, G(\boldsymbol{z}_i, \boldsymbol{\lambda}), G^{(1)}(\boldsymbol{z}_i, \boldsymbol{\lambda}) diag(\boldsymbol{\xi}) \right\rangle . \tag{4.17}$$

Note that this perturbation scheme may not be meaningful for discrete response, such as those in binary logistic regression.

4.3 Examples

4.3.1 Snow geese data

Consider the snow geese data as reported by Weisberg (1985) and further analyzed in Wei and Hickernell (1996). The data set consists of observations on the response y = true flock size as obtained by count from aerial photographs and covariate x = visually estimated flock size for a sample of $n = 45$ flocks of snow geese.

The original fitted regression model is

$$\hat{y}_i = 26.65 + 0.883 x_i$$

$$(8.61) \quad (0.08)$$

with standard errors of the coefficients enclosed in parentheses. In view of the heteroscedasticity evident in the data, Wei and Hickernell (1996) proposed the following covariate transformation model

$$y_i = \delta + \left(\frac{x_i^{\lambda} - 1}{\lambda} \right) \xi + \varepsilon_i \ .$$

Parameter estimates for δ and ξ are -35.759 (10.83) and 8.604 (0.63) respectively, and $\hat{\lambda} = 0.538$.

Figure 4.1

Case deletion diagnostics (rescaled) d_i and $\hat{\lambda}_{[i]} - \hat{\lambda}$ for snow geese data.

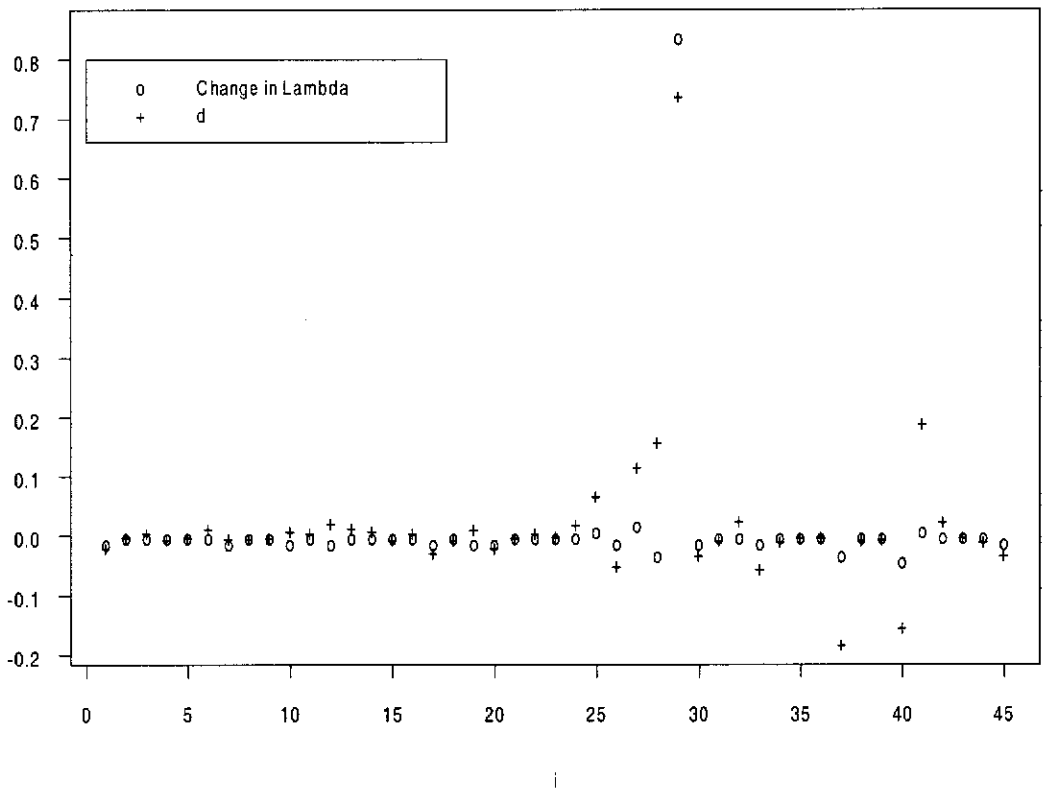
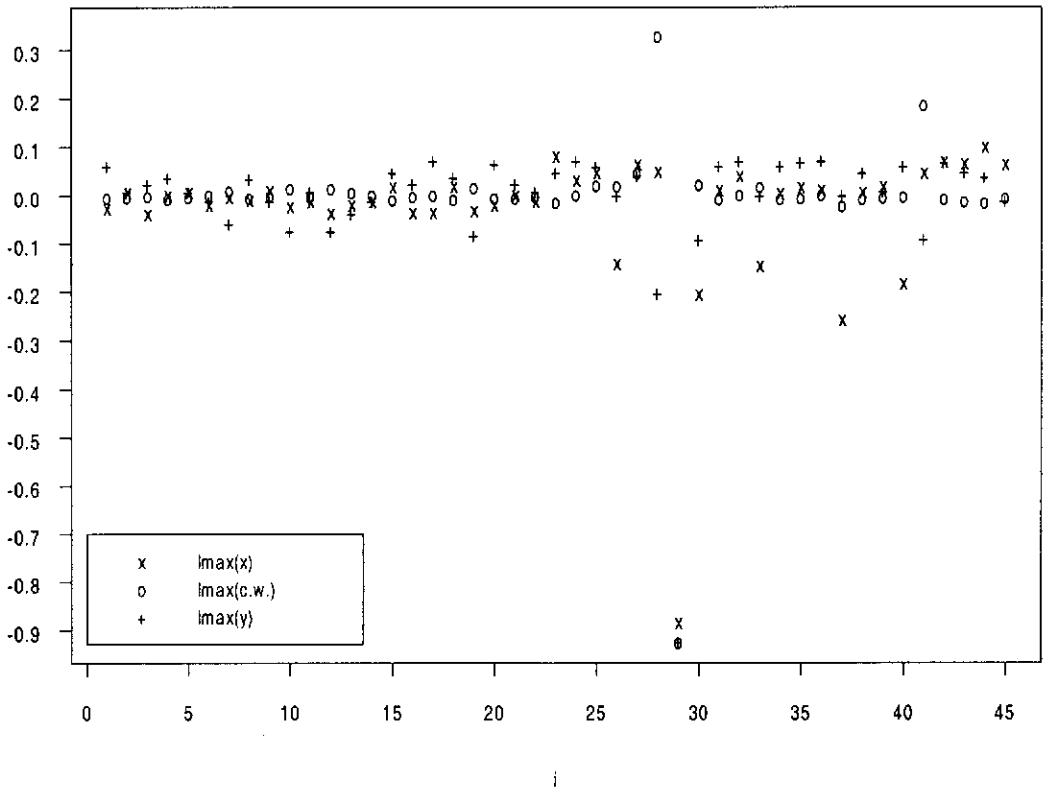


Figure 4.1 shows the (re-scaled) partial influence measure d_i and $\hat{\lambda}_{[i]} - \hat{\lambda}$, based on case deletions. Case 29 is the most influential observation, which is consistent with the index plot of profile likelihood displacement LD_i (Wei and Hickernell (Figure 2)). It affects the estimate of λ significantly, $\hat{\lambda}_{[29]} = 1.38$. Indeed, case 29 is a leverage point recording the highest observer count of 500 birds.

We next examine the local effect of each case on $\hat{\lambda}$. Based on the partial influence approach, the direction cosines $\boldsymbol{l}^d_{max}$ from perturbing case weight, response, and transformed covariate are plotted against case index in Figure 4.2.

Figure. 4.2

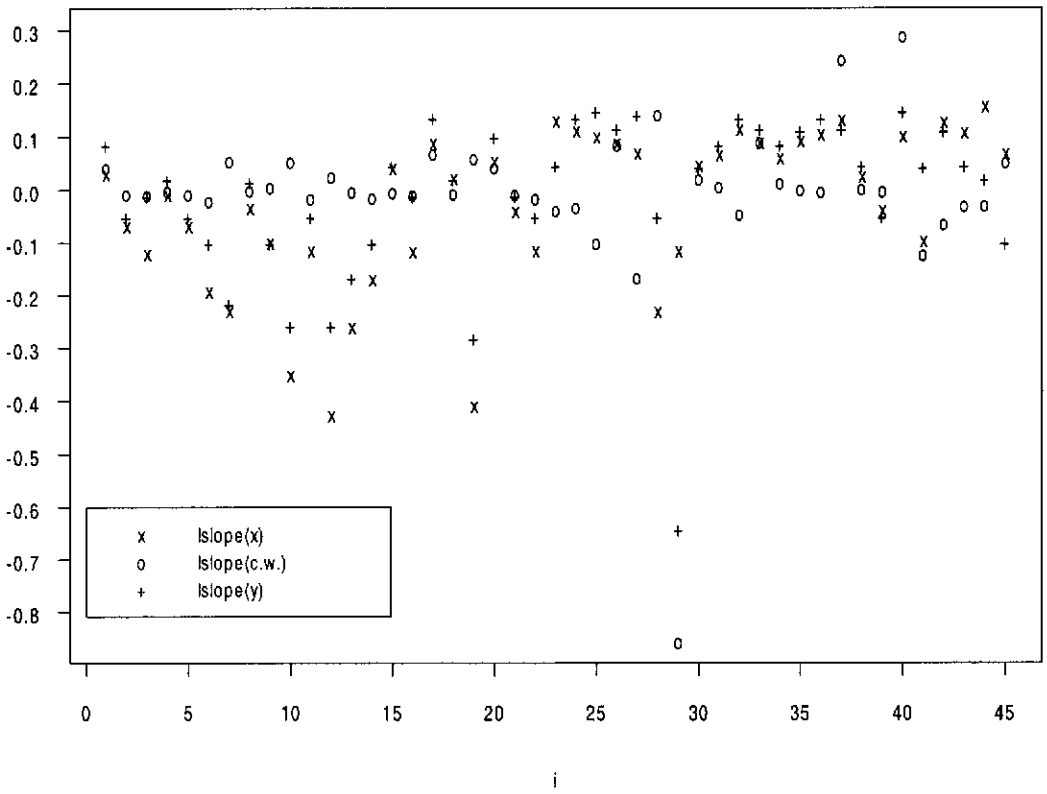
Direction cosines l^d_{max} from local perturbations for snow geese data.



It is evident that the greatest local change in $\hat{\lambda}$ depends essentially on case 29. This result is consistent with the $l^{\hat{\lambda}}_{slope}$ vectors displayed in Figure 4.3 under case weight and response perturbation schemes. However, no cosine in $l^{\hat{\lambda}}_{slope}$ appears to be outlying with respect to perturbations of the transformed covariate. Since λ is a scalar parameter, $l^{\hat{\lambda}}_{max} \propto l^{\hat{\lambda}}_{slope}$. It is interesting to note that Wei and Hickernell had to resort to an alternative perturbation scheme (proportional instead of additive) before case 29 becomes discordant.

Figure. 4.3

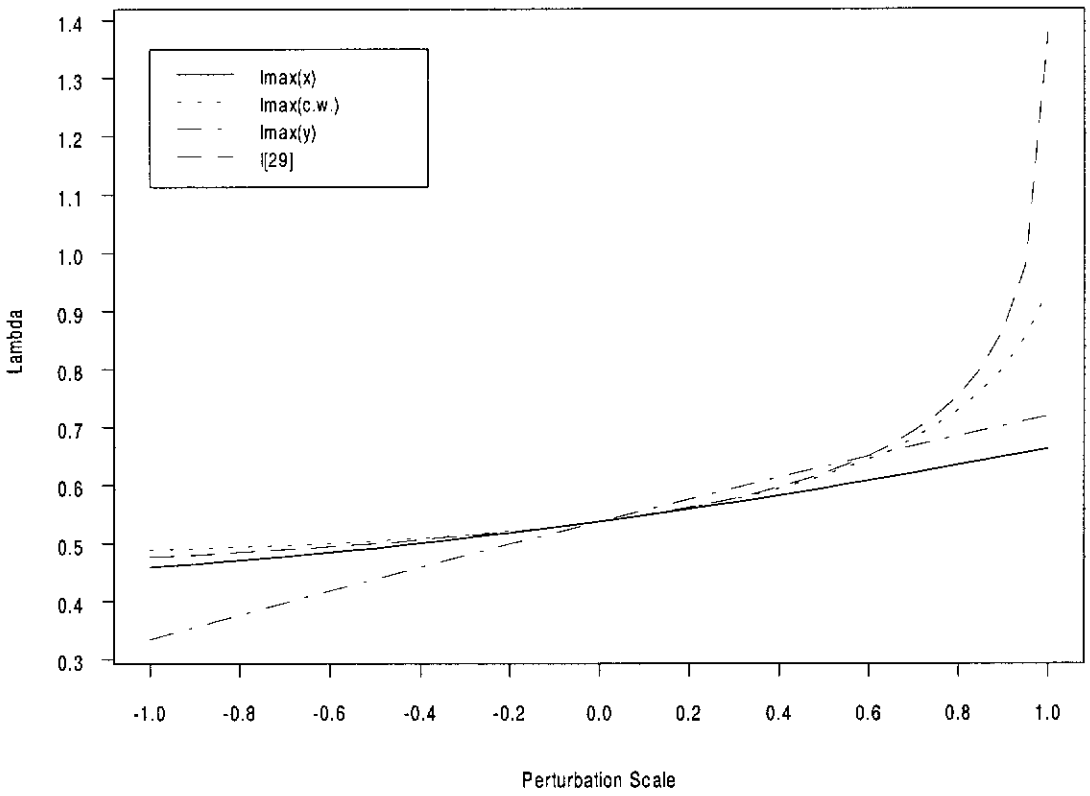
Direction cosines $\boldsymbol{l}^{\hat{\lambda}}_{slope}$ from local perturbations for snow geese data.



To confirm the indications of the proposed local influence diagnostics, we plot $\hat{\lambda}_{\boldsymbol{\omega}}$ against the perturbation scale for each maximizing local direction $\boldsymbol{l}^{d}_{max}$ in Figure 4.4. The effects of downweighting case 29 (in direction $\boldsymbol{l}_{[29]}$) are almost the same as those of simultaneously perturbing all case weights, except when the perturbation scale approaches 1, where $\hat{\lambda}$ increases rapidly as the contribution of case 29 is downweighted to zero. It is worth noting that while the curve associated with response perturbations has the greatest slope at the null state where the perturbation scale is 0, the analysis is quite insensitive to minor modifications in the covariate.

Figure. 4.4

$\hat{\lambda}$ in directions of local influence for snow geese data.



4.3.2 Erythrocyte sedimentation rate data

We next illustrate the proposed diagnostics with data from Collett (1991, p. 8) relating the chronic disease state y_i (0 = healthy; 1 = unhealthy) of 32 individuals, judged from the erythrocyte sedimentation rate (ESR) reading, to the plasma fibrinogen level x_i (in gm/ℓ). The fitted logistic regression model is

$$\begin{aligned} \text{logit}(\hat{\mu}_i) &= -\,6.845 + 1.827 x_i \\ &\qquad (2.764)\ (0.899) \end{aligned}$$

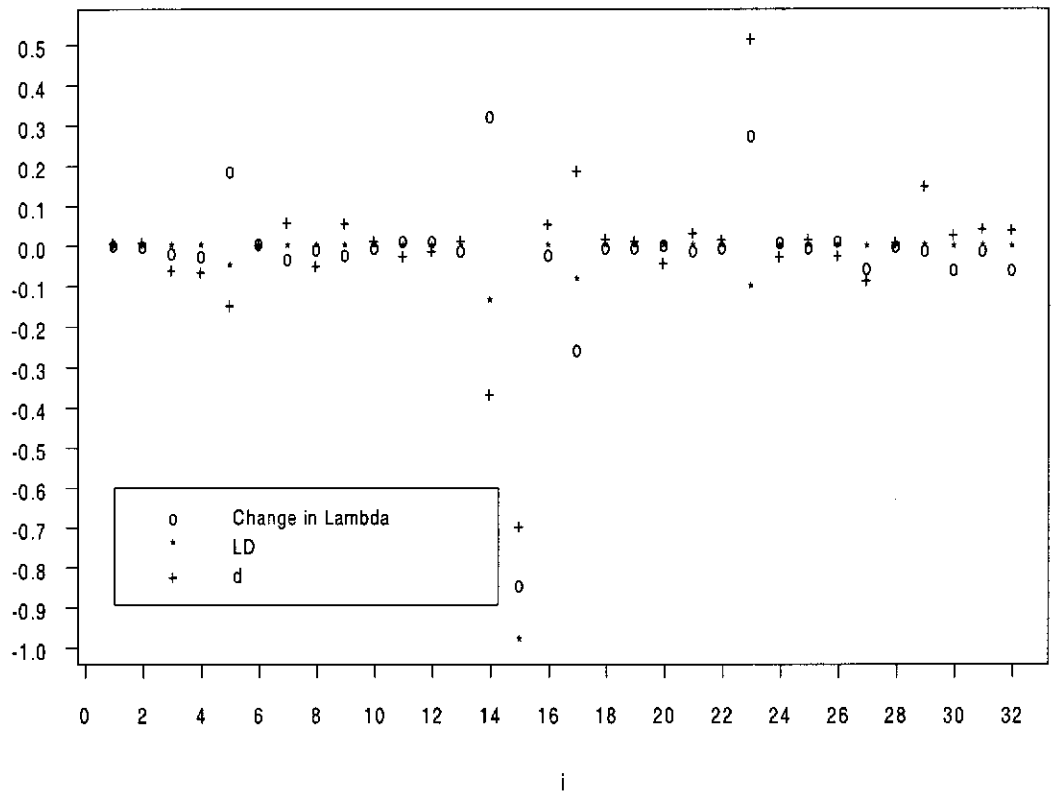
with deviance 24.84 on 30 d.f. A constructed variable plot suggests cases 15 and 23 are outliers and that a non-linear transformation of x_i is required; see Collett (1991, p. 167). Collett then proceeded to include a quadratic term in the model. Alternatively, we consider fitting the Box-Cox transformation model:

$$\text{logit}(\hat{\mu}_i) = \delta_0 + x_i \delta_1 + \left(\frac{x_i^{\lambda} - 1}{\lambda} \right) \xi \ .$$

The resulting reduction in deviance, 8.06, is significant at the 5% level. The MLE for λ is 6.81, while parameter estimates for δ_0, δ_1 and ξ are 20.731 (10.094), -10.425 (5.095), 0.023 (0.013), respectively. Case deletion diagnostics (re-scaled) LD_i and $\hat{\lambda}_{[i]} - \hat{\lambda}$ displayed in Figure 4.5 identify case 15 only, but large values of d_i are found for case 15 and to a certain extent, case 23.

Figure. 4.5

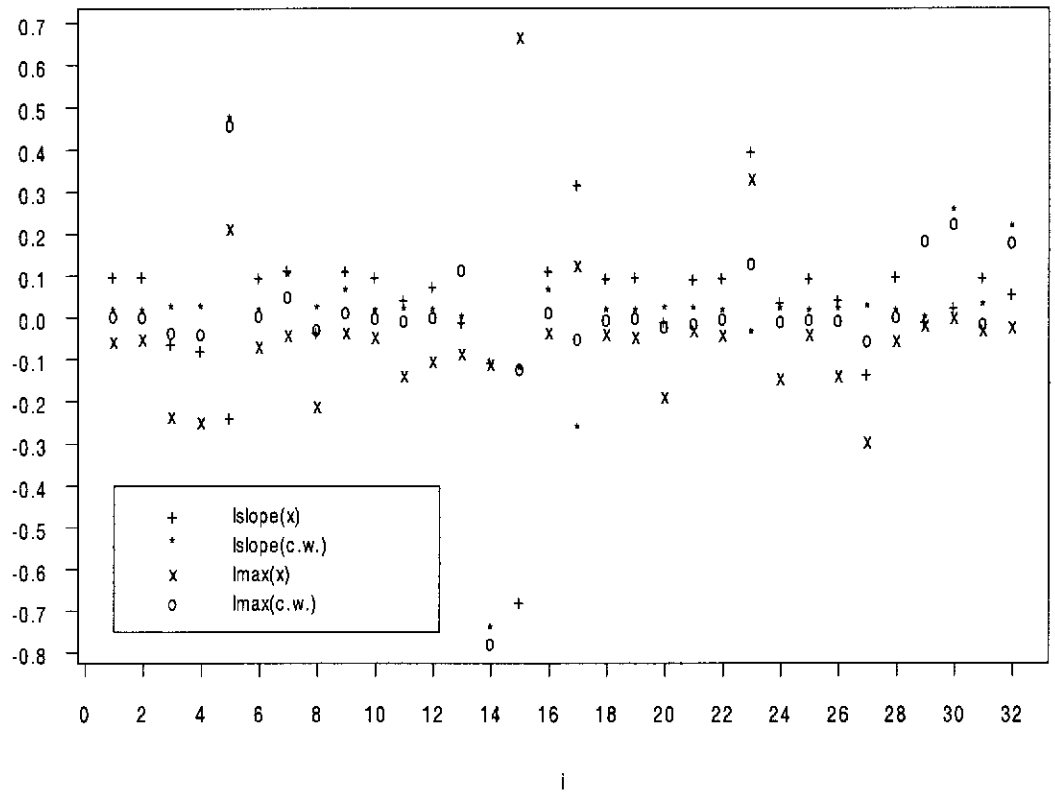
Case deletion diagnostics (rescaled) LD_i, d_i and $\hat{\lambda}_{[i]} - \hat{\lambda}$ for ESR data.



The direction cosines $l^{\hat{\lambda}}_{slope}$ and l^{d}_{max} from minor perturbations are plotted against case index in Figure 4.6. Results based on the first/second order approach are consistent with those of the partial influence approach. Upon perturbing the transformed covariate, case 15, followed by case 23, have components that are separated from those of the other individuals. An inspection of the data reveals that these two observations correspond to unhealthy patients with unusually low plasma fibrinogen counts. Under perturbation of case weight, cases 5 and 14 emerge as influential on the transformation. We note that case 5 recorded the highest fibrinogen level compared to other healthy individuals in the sample. On the other hand, case 14 has near average fibrinogen level among the unhealthy group, yet its standardized deviance residual is the second largest (after case 15) on fitting a quadratic logistic regression (Collett (1991, p. 168)).

Figure. 4.6

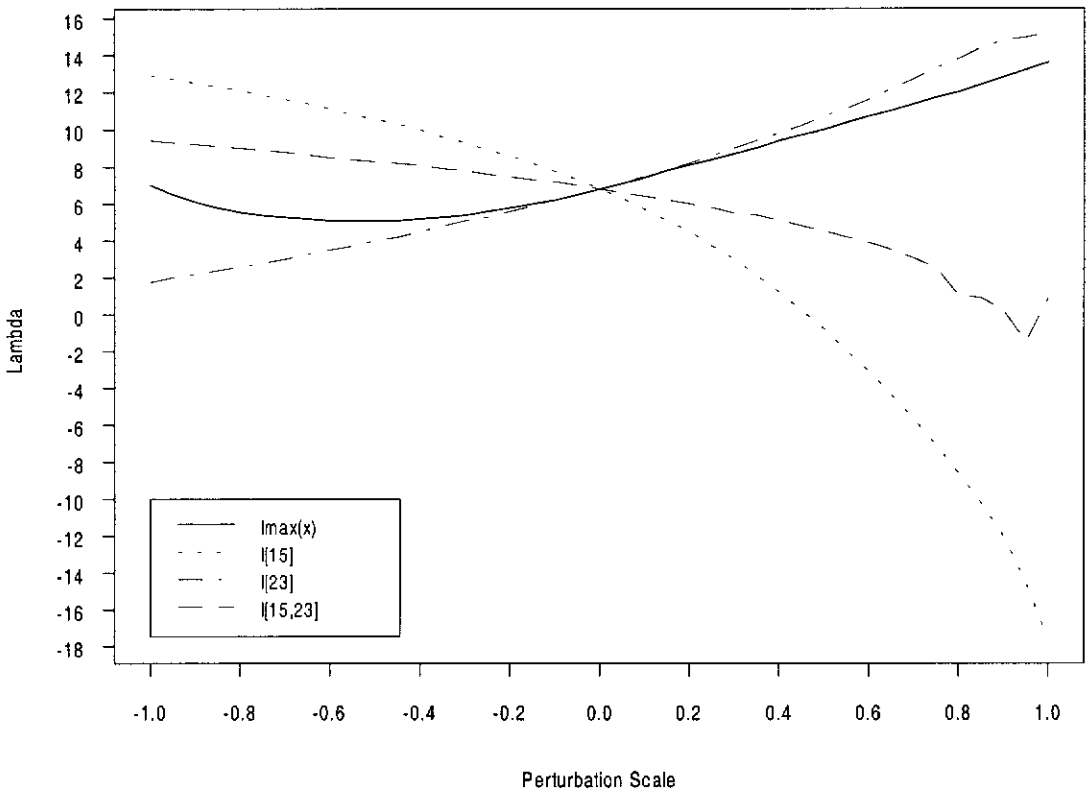
Direction cosines $l^{\hat{\lambda}}_{slope}$ and l^{d}_{max} from local perturbations for ESR data.



To further assess the extent of the perturbation effects, we plot the actual $\hat{\lambda}$ in Figures 4.7 and 4.8 for selected local directions $\boldsymbol{l}$ of interest, including those related to the deletion of cases. The curve $\boldsymbol{l}_{[15]}$ associated with the downweighting of case 15 alone has the greatest slope at the state of no perturbation, in addition to producing the maximum global change in $\hat{\lambda}$. We also found that as the contributions of cases 15 and 23 are being reduced to zero (direction $\boldsymbol{l}_{[15,23]}$), $\hat{\lambda}$ approaches 1, representing no transformation. Therefore, once these two observations are removed, there is no evidence for covariate transformation.

Figure. 4.7

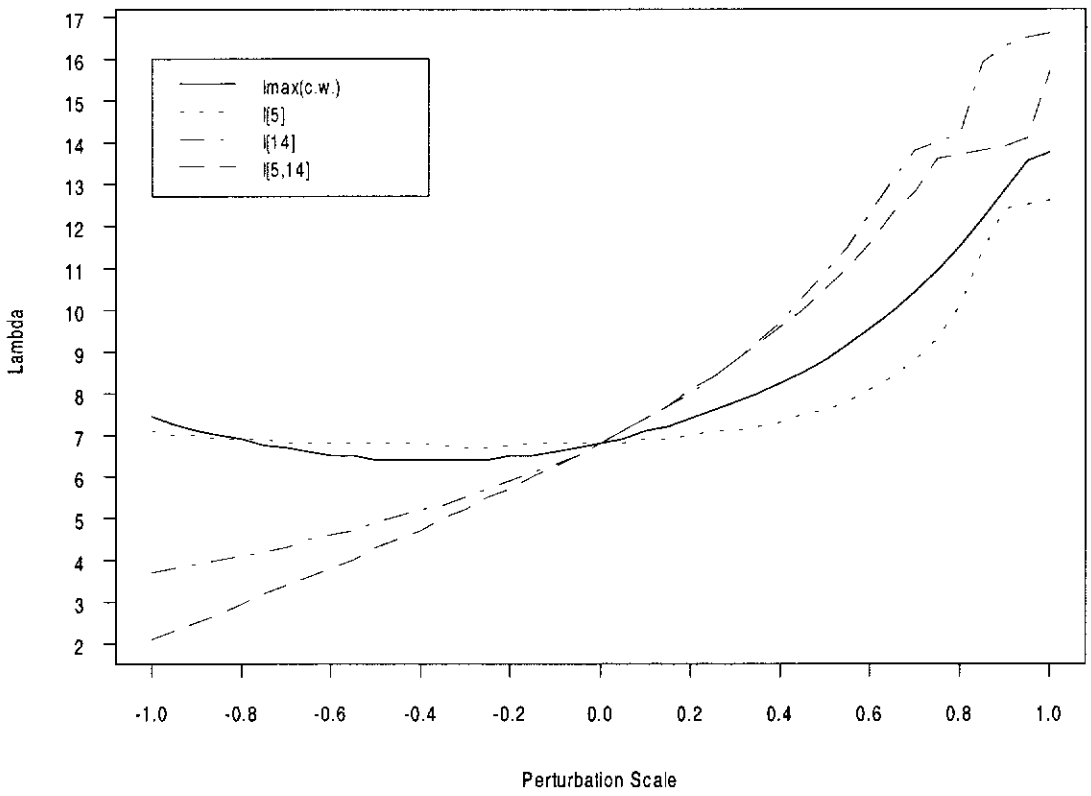
$\hat{\lambda}$ in local influence directions associated with transformed covariate perturbations for ESR data.



The net overall change due to covariate perturbations is not dramatic since the impact exerted by case 15 apparently has been compensated by the other cases. Besides, perturbations in this direction give similar effects as those of simultaneously modifying all case weights. Figure 4.8 also confirms that sensitivity of the transformation parameter depends considerably on the weights attached to cases 5 and 14.

Figure. 4.8

$\hat{\lambda}$ in local influence directions associated with case weight perturbations for ESR data.



4.3.3 Tree data

To provide a numerical illustration of the diagnostics when $\boldsymbol{\lambda}$ is a vector quantity of interest, we consider the tree data from Ryan, Joiner and Ryan (1976, p. 278). The data consist of measurements on tree volume y (in ft^3), tree height x_2 (in ft), and tree diameter x_1 (in *inches*) at $4.5ft$ above ground level for a sample of $n = 31$ black cherry trees. The following covariate transformation model,

$$y_i = \delta + \left(\frac{x_{i1}^{\lambda_1} - 1}{\lambda_1}\right)\xi_1 + \left(\frac{x_{i2}^{\lambda_2} - 1}{\lambda_2}\right)\xi_2 + \varepsilon_i \ ,$$

suggested by Wei and Hickernell (1996), will be adopted in our analysis. The deviance of the fitted model is 188.247 with $\hat{\boldsymbol{\lambda}} = (2.583, 1.738)^T$.

Figure. 4.9

Case deletion diagnostics (rescaled) d_i, $\hat{\lambda}_{[i]1} - \hat{\lambda}_1$ *and* $\hat{\lambda}_{[i]2} - \hat{\lambda}_2$ *for Tree data.*

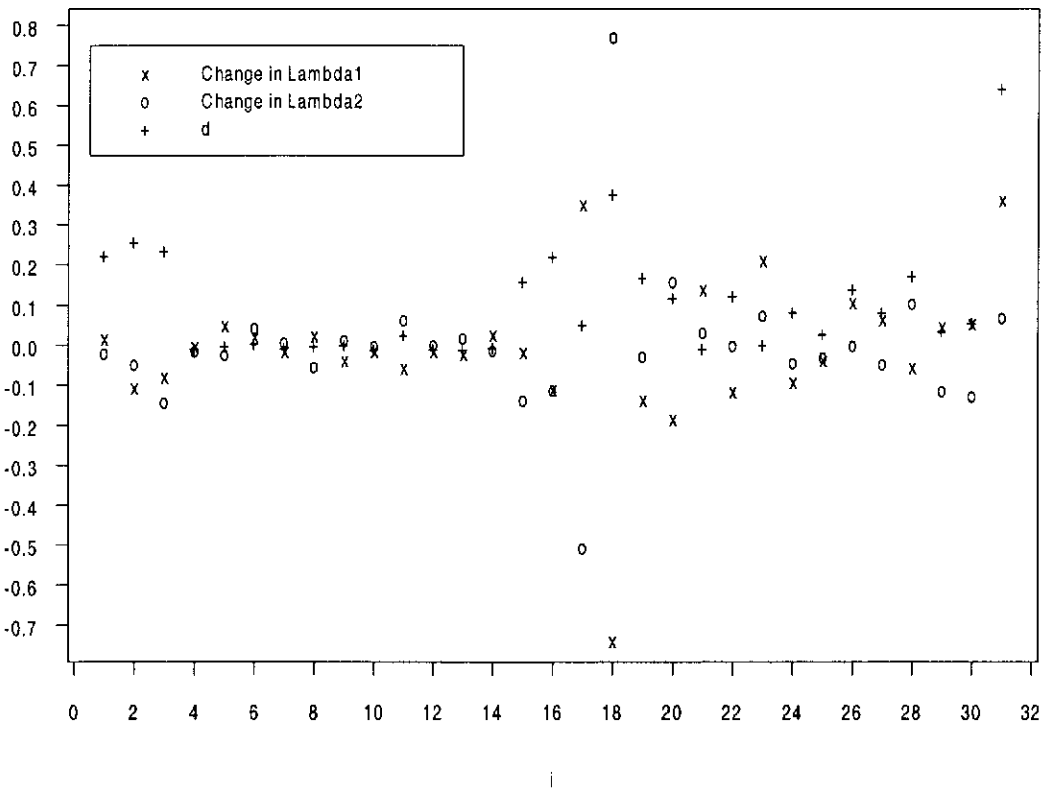
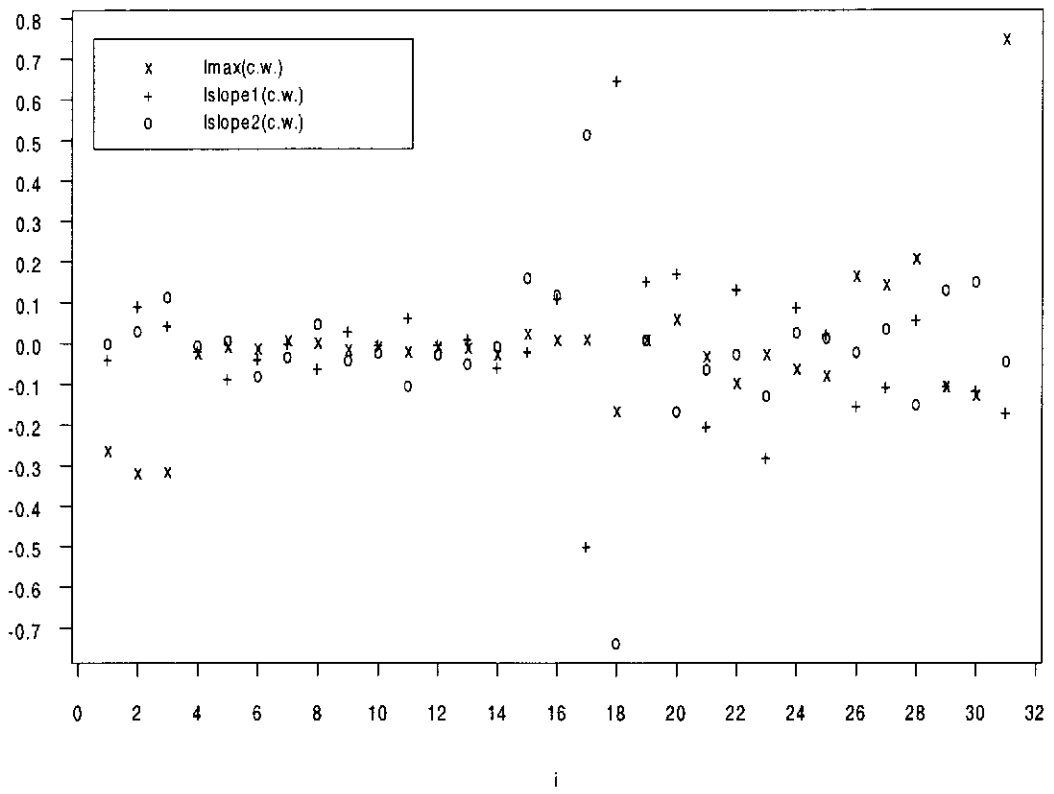


Figure 4.9 gives (re-scaled) $\hat{\lambda}_{[i]1} - \hat{\lambda}_1$, $\hat{\lambda}_{[i]2} - \hat{\lambda}_2$, and d_i. In addition to having the most extreme LD_i values (see Wei and Hickernell (Figure 8)), cases 17 and 18 induce substantial changes in the transformation parameter estimates upon their deletion, $\hat{\boldsymbol{\lambda}}_{[17]} = (2.71, -1.137)^T$, $\hat{\boldsymbol{\lambda}}_{[18]} = (2.325, 6.25)^T$. Meanwhile, the partial influence measure shows that the evidence for covariate transformation depends mainly on case 31.

Figure. 4.10

Direction cosines $\boldsymbol{l}^{\hat{\lambda}_1}_{slope}$, $\boldsymbol{l}^{\hat{\lambda}_2}_{slope}$ and $\boldsymbol{l}^{d}_{max}$ from case weight perturbations for Tree data.



The local influence diagnostics under case weight perturbations are plotted in Figure 4.10. Cases 17 and 18 are clearly influential according to the first order diagnostics $\boldsymbol{l}^{\hat{\lambda}_1}_{slope}$ and $\boldsymbol{l}^{\hat{\lambda}_2}_{slope}$, whereas case 31 is also influential due to its large component of $\boldsymbol{l}^{d}_{max}$. These direction cosines provide different diagnostic infor-

mation to that of the $l^{\hat{\lambda}}_{max}$ vector (Wei and Hickernell (Figure 9)). On the other hand, case 31 emerges as the only influential observation under transformed covariate perturbations. It should be remarked that while case 31 corresponds to the largest tree in the sample, cases 17 and 18 are medium sized trees but with distinctive height of 85 ft and 86 ft respectively.

CHAPTER 5

PARAMETRIC LINK FUNCTIONS IN GENERALIZED LINEAR MODELS

5. PARAMETRIC LINK FUNCTIONS IN GLM

The link function, which relates the linear predictor to the expected value of the response, is a major component of a GLM (McCullagh and Nelder, 1989). When the link function is unknown, it may be assumed as a member of a parametric family indexed by $\boldsymbol{\lambda}$, choices of which lead to different link specifications (Pregibon, 1980). The parameter(s) $\boldsymbol{\lambda}$ is usually estimated by the method of maximum likelihood (Scallan, Gilchrist and Green (1984), Kaiser (1997)). Several parametric families of link functions have been considered in the literature, see e.g. Prentice (1976), Aranda-Ordaz (1981), and Pregibon (1980, 1985).

The maximum likelihood estimate (MLE) of $\boldsymbol{\lambda}$, $\hat{\boldsymbol{\lambda}}$, and hence the exact form of the link function, may depend crucially on one or a few extreme observations. In Box-Cox transformation models, a variety of diagnostic techniques has been developed for the transformation parameter; see Cook and Wang (1983), Atkinson (1988), Tsai and Wu (1990). The local influence methodology of Cook (1986) has also been applied to assess the sensitivity of the transformation parameter estimator (Lawrance (1988), Tsai and Wu (1992)). However, link modification should not be confused with response transformation, the latter typically assumes both linearity and normality on the transformed response-scale.

In this chapter we present influence diagnostics for assessing the effect of minor perturbations on the MLE of the link parameter in GLM. Two separate approaches based on analysis of the link parameter surface and partial influence are proposed in the next section. Specific perturbation schemes are outlined in Section 5.2 to examine the different aspects of influence. Two numerical examples illustrating sensitivity of the link analysis are provided in Section 5.3. It is shown that application of the diagnostics can assist us in revising $\boldsymbol{\lambda}$ and hence the form of the model.

5.1 Likelihood displacement and local influence

We consider n independent observations with responses $\boldsymbol{y} = \langle y_1, \cdots, y_n \rangle^T$ and an $n \times p$ matrix of covariates $\boldsymbol{x}$. The responses $\boldsymbol{y}$ are distributed according to the exponential family

$$f_y(y_i; \boldsymbol{\theta}) = \exp \left\{ \left[y_i \theta_i - b(\theta_i) \right] / \mathsf{a}(\phi) + c(y_i, \phi) \right\}$$

with θ-link function $\theta_i = k(\eta_i, \boldsymbol{\lambda})$, where $\eta_i = \boldsymbol{x}_i \boldsymbol{\beta}$ denotes the linear predictor, and $\mathsf{a}(\cdot)$, $b(\cdot)$, $c(\cdot)$ are known functions. The θ-link function corresponds to a particular member of a parametric link family $g(\boldsymbol{\mu}, \boldsymbol{\lambda})$ indexed by $\boldsymbol{\lambda} = \langle \lambda_1, \cdots, \lambda_q \rangle^T$, where $\boldsymbol{\mu} = E[\boldsymbol{y}]$. Without loss of generality the dispersion parameter ϕ is assumed known or may be replaced by an estimate $\hat{\phi}$ and write $\hat{\mathsf{a}} = \mathsf{a}(\hat{\phi})$. Here the unknown parameter $\boldsymbol{\lambda}$ is of special interest. The log-likelihood function is then given by

$$L(\boldsymbol{\lambda}; \boldsymbol{\beta}) = \hat{\mathsf{a}}^{-1} \sum_{i=1}^{n} \left[y_i k(\eta_i, \boldsymbol{\lambda}) - b\{k(\eta_i, \boldsymbol{\lambda})\} \right] .$$

Goodness-of-fit of a GLM may often be improved by link modification. Let $\tilde{\boldsymbol{\beta}}(\boldsymbol{\lambda})$ be the function that maximizes $L(\boldsymbol{\lambda}; \boldsymbol{\beta})$ for fixed $\boldsymbol{\lambda}$ and denote the corresponding profile log-likelihood for $\boldsymbol{\lambda}$ by $L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\beta}}(\boldsymbol{\lambda}))$. To assess the global influence of individual cases on the MLE $\hat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$, one can adopt the case deletion approach of Cook and Weisberg (1982). The difference between $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\lambda}}_{[i]}$, the MLE of $\boldsymbol{\lambda}$ without case i, can be measured through the profile likelihood displacement

$$LD_i = 2[L(\hat{\boldsymbol{\lambda}}) - L(\hat{\boldsymbol{\lambda}}_{[i]})] \tag{5.1}$$

where $L(\boldsymbol{\lambda}) = L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\beta}}(\boldsymbol{\lambda}))$. A large value of LD_i indicates that $\hat{\boldsymbol{\lambda}}$ is likely to be dependent on case i.

5.1.1 First and second order approach

For the MLE surface of the link parameter $\boldsymbol{\lambda}$, we introduce small changes into our model through an $n \times 1$ vector $\boldsymbol{\omega} = \boldsymbol{\omega}_0 + \mathbf{a}\boldsymbol{l} \in \Omega$, where Ω denotes the open set of relevant perturbations, $\boldsymbol{l} = \langle l_1, \cdots, l_n \rangle^T$ is a unit direction vector and the quantity $\mathbf{a}$ measures the magnitude of the perturbations along the direction $\boldsymbol{l}$. The null point $\boldsymbol{\omega}_0 \in \Omega$ represents no perturbation so that $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}_0} = \hat{\boldsymbol{\lambda}}$.

To find the direction of largest local change, we approximate the MLE surface by its tangent plane at $\boldsymbol{\omega}_0$, which is determined by $\frac{\partial \hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}}{\partial \boldsymbol{\omega}^T}$ at $\boldsymbol{\omega}_0$. The direction of largest local change is just the direction of maximum slope on this tangent plane over Ω.

Similar to the previous chapter, we write $L(\boldsymbol{\lambda}|\boldsymbol{\omega}) = L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\beta}}(\boldsymbol{\lambda}|\boldsymbol{\omega})|\boldsymbol{\omega})$ for the profile log-likelihood corresponding to the perturbed model, where $\tilde{\boldsymbol{\beta}}(\boldsymbol{\lambda}|\boldsymbol{\omega})$ is the function that maximizes $L(\boldsymbol{\lambda}; \boldsymbol{\beta}|\boldsymbol{\omega})$ for fixed $\boldsymbol{\lambda}$ and $\boldsymbol{\omega}$. Then $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}$ satisfies the following equation:

$$\frac{\partial L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\beta}}|\boldsymbol{\omega})}{\partial \boldsymbol{\lambda}} = \mathbf{0} \, .$$

Differentiating with respect to ω_i yields

$$\frac{\partial^2 L(\boldsymbol{\lambda};, \tilde{\boldsymbol{\beta}}|\boldsymbol{\omega})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T} \left(\frac{\partial \boldsymbol{\lambda}_{\boldsymbol{\omega}}}{\partial \omega_i} \right) + \frac{\partial^2 L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\beta}}|\boldsymbol{\omega})}{\partial \omega_i \partial \boldsymbol{\lambda}^T} = \mathbf{0} \, .$$

Therefore,

$$\frac{\partial \boldsymbol{\lambda}_{\boldsymbol{\omega}}}{\partial \omega_i} = \left[\frac{\partial^2 L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\beta}}|\boldsymbol{\omega})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T} \right]^{-1} \left(\frac{-\partial^2 L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\beta}}|\boldsymbol{\omega})}{\partial \omega_i \partial \boldsymbol{\lambda}^T} \right) \, , \tag{5.2}$$

and

$$\frac{\partial^2 L(\boldsymbol{\lambda}; \tilde{\boldsymbol{\beta}}|\boldsymbol{\omega})}{\partial \omega_i \partial \boldsymbol{\lambda}^T} = L^{(2)}_{1\omega i} + L^{(2)}_{12} \frac{\partial \tilde{\boldsymbol{\beta}}(\boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \omega_i},$$

which is equivalent to

$$L^{(2)}_{1\omega i} - L^{(2)}_{12} \left[L^{(2)}_{22} \right]^{-1} L^{(2)}_{2\omega i} \, .$$

The partitions of $L^{(2)}$ are

$$L_{11}^{(2)} = \sum_{i=1}^{n} y_i k_{11}^{(2)}(\eta_i, \boldsymbol{\lambda}) - b^{(2)}(k(\eta_i, \boldsymbol{\lambda})) \left[k_1^{(1)}(\eta_i, \boldsymbol{\lambda})\right]^T k_1^{(1)}(\eta_i, \boldsymbol{\lambda}) - b^{(1)}(k(\eta_i, \boldsymbol{\lambda})) k_{11}^{(2)}(\eta_i, \boldsymbol{\lambda})$$

$$L_{12}^{(2)} = \sum_{i=1}^{n} y_i k_{12}^{(2)}(\eta_i, \boldsymbol{\lambda}) - b^{(2)}(k(\eta_i, \boldsymbol{\lambda})) \left[k_2^{(1)}(\eta_i, \boldsymbol{\lambda}) k_1^{(1)}(\eta_i, \boldsymbol{\lambda})\right]^T - b^{(1)}(k(\eta_i, \boldsymbol{\lambda})) k_{12}^{(2)}(\eta_i, \boldsymbol{\lambda})$$

$$L_{21}^{(2)} = [L_{12}^{(2)}]^T$$

$$L_{22}^{(2)} = \sum_{i=1}^{n} y_i k_{22}^{(2)}(\eta_i, \boldsymbol{\lambda}) - b^{(2)}(k(\eta_i, \boldsymbol{\lambda})) \left[k_2^{(1)}(\eta_i, \boldsymbol{\lambda})\right]^T k_2^{(1)}(\eta_i, \boldsymbol{\lambda}) - b^{(1)}(k(\eta_i, \boldsymbol{\lambda})) k_{22}^{(2)}(\eta_i, \boldsymbol{\lambda}).$$

Quantities $L_{1\omega i}^{(2)} = \dfrac{\partial^2 L(\boldsymbol{\lambda}; \boldsymbol{\beta}|\boldsymbol{\omega})}{\partial \omega_i \partial \boldsymbol{\lambda}^T}$ and $L_{2\omega i}^{(2)} = \dfrac{\partial^2 L(\boldsymbol{\lambda}; \boldsymbol{\beta}|\boldsymbol{\omega})}{\partial \omega_i \partial \boldsymbol{\beta}^T}$ are entries on the corresponding columns of $\boldsymbol{\Delta}(\boldsymbol{\lambda}_\omega; \boldsymbol{\beta}_\omega) = \frac{\partial^2 L(\boldsymbol{\lambda}; \boldsymbol{\beta}|\boldsymbol{\omega})}{\partial \boldsymbol{\omega} \partial (\boldsymbol{\lambda}; \boldsymbol{\beta})^T}$. It is further derived for various types of perturbations in Section 5.2.

The calculations for the maximum slope direction at the null point, $\boldsymbol{l}_{slope}^{\hat{\boldsymbol{\lambda}}}$, and the direction $\boldsymbol{l}_{max}^{\hat{\boldsymbol{\lambda}}}$ corresponding to the maximum normal curvature of the MLE surface, are the same as those in Chapter 4 Section 4.1.1.

5.1.2 Partial influence approach

Let the original model with the hypothesized link function be given by

$$g(\boldsymbol{\mu}, \boldsymbol{\lambda}_0) = \boldsymbol{x}\boldsymbol{\delta} \; . \tag{5.3}$$

Suppose the underlying parametric link family is actually

$$g(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \boldsymbol{x}\boldsymbol{\beta} \; . \tag{5.4}$$

A test of the hypothesis H_0: $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$ can be based on $D_0 - D$, the reduction in deviance from model (5.3) to model (5.4). Denote $D_{0[i]}$ and $D_{[i]}$ for the deviance

of (5.3) and (5.4) respectively after deleting case i. Similar to Section 4.1.2, a partial influence measure for the impact of case i on the link can be formulated as

$$d_i = (D_0 - D) - (D_{0[i]} - D_{[i]}) \ , \tag{5.5}$$

which represents the change in deviance due to link modification when the ith observation is excluded. A large positive d_i indicates the ith observation is contributing substantially to the rejection of the assumed link, whereas a large negative d_i implies that deletion of the observation actually strengthens the evidence for link modification.

Consider the log-likelihood $L(\boldsymbol{\lambda};\boldsymbol{\beta})$ of model (5.4). The full MLE of $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$ are denoted by $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\lambda}}$ respectively. Similarly, let $L_0(\boldsymbol{\delta})$ be the log-likelihood of model (5.3), with MLE $\hat{\boldsymbol{\delta}}$. Under minor perturbations, the respective log-likelihood becomes $L(\boldsymbol{\lambda};\boldsymbol{\beta}|\boldsymbol{\omega})$ and $L_0(\boldsymbol{\delta}|\boldsymbol{\omega})$, with associated MLEs $(\hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})$ and $\hat{\boldsymbol{\delta}}_{\boldsymbol{\omega}}$. Suppose that $L(\boldsymbol{\lambda};\boldsymbol{\beta}|\boldsymbol{\omega}_0) = L(\boldsymbol{\lambda};\boldsymbol{\beta})$ and $L_0(\boldsymbol{\delta}|\boldsymbol{\omega}_0) = L_0(\boldsymbol{\delta})$. The partial influence on the link due to perturbation $\boldsymbol{\omega}$ can be assessed by

$$d(\boldsymbol{\omega}) = 2\left\{\left[L_0(\hat{\boldsymbol{\delta}}) - L(\hat{\boldsymbol{\lambda}};\hat{\boldsymbol{\beta}})\right] - \left[L_0(\hat{\boldsymbol{\delta}}_{\boldsymbol{\omega}}) - L(\hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}};\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})\right]\right\} \ . \tag{5.6}$$

Analogous to (5.5) in case-deletion, the log-likelihood displacement $d(\boldsymbol{\omega})$ measures the local effect on the link parameter with respect to the contours of the unperturbed deviance reduction.

Following the same steps given in Section 4.1.2, to obtain the normal curvature at $F(\boldsymbol{\omega}_0)$ along the direction $\boldsymbol{l}$, we have

$$F^{(2)} = \boldsymbol{\Delta}(\boldsymbol{\lambda}_{\boldsymbol{\omega}};\boldsymbol{\beta}_{\boldsymbol{\omega}})\left[L^{(2)}(\boldsymbol{\lambda};\boldsymbol{\beta})\right]^{-1}\boldsymbol{\Delta}^T(\boldsymbol{\lambda}_{\boldsymbol{\omega}};\boldsymbol{\beta}_{\boldsymbol{\omega}}) - \nabla(\boldsymbol{\delta}_{\boldsymbol{\omega}})\left[L_0^{(2)}(\boldsymbol{\delta})\right]^{-1}\nabla^T(\boldsymbol{\delta}_{\boldsymbol{\omega}}) \tag{5.7}$$

where $\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\boldsymbol{\omega}};\boldsymbol{\beta}_{\boldsymbol{\omega}}) = \frac{\partial^2 L(\boldsymbol{\lambda};\boldsymbol{\beta}|\boldsymbol{\omega})}{\partial\boldsymbol{\omega}\partial(\boldsymbol{\lambda};\boldsymbol{\beta})^T}$ and $\nabla(\boldsymbol{\delta}_{\boldsymbol{\omega}}) = \frac{\partial^2 L_0(\boldsymbol{\delta}|\boldsymbol{\omega})}{\partial\boldsymbol{\omega}\partial\boldsymbol{\delta}^T}$, is evaluated at $\boldsymbol{\omega}_0$, $\hat{\boldsymbol{\delta}}$, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\lambda}}$. Expressions for $L_0^{(2)}(\boldsymbol{\delta})$ and $\nabla(\boldsymbol{\delta}_{\boldsymbol{\omega}})$ are derived under the original generalized linear model (5.3), which can be found in Thomas and Cook (1989).

Again, $\boldsymbol{l}_{max}^{d}$, the perturbation direction that produces the greatest local change in $\boldsymbol{\lambda}$ as measured by (5.6) is just the eigenvector associated with the largest eigenvalue of

$$\boldsymbol{\Delta}[L^{(2)}]^{-1}\boldsymbol{\Delta}^{T} - \nabla[L_0^{(2)}]^{-1}\nabla^{T} \ . \tag{5.8}$$

The most influential elements of the data on the link parameter may be identified by their large components of $\boldsymbol{l}_{max}^{d}$. To assess the global effects of the local perturbations, one may plot $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}_0+\mathrm{a}\boldsymbol{l}}$ against the perturbation size a for each local direction $\boldsymbol{l}$ of interest. As in the previous chapter, the characteristics of such curves should be informative for further investigation on the relationship between local and global influences.

5.2 Perturbation schemes

As with covariate transformation, we consider relevant perturbation schemes and derive the corresponding $\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\boldsymbol{\omega}}; \boldsymbol{\beta}_{\boldsymbol{\omega}})$ quantity below.

5.2.1 Perturbation of case weights

We define a vector of weights $\boldsymbol{\omega} = \langle \omega_1, \cdots, \omega_n \rangle^T$, $\omega_i \geq 0$, to perturb the contribution of each case to the log-likelihood. The point representing no perturbation is $\boldsymbol{\omega}_0 = \langle 1, \cdots, 1 \rangle^T$. We obtain $\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\boldsymbol{\omega}}; \boldsymbol{\beta}_{\boldsymbol{\omega}}) = \langle \Delta_i^w \rangle$ evaluated at $\boldsymbol{\omega}_0$ and $\hat{\boldsymbol{\lambda}}$, where

$$\Delta_i^w = \left\{ y_i - b^{(1)}(k(\eta_i, \boldsymbol{\lambda})) \right\} \left\langle k_2^{(1)}(\eta_i, \boldsymbol{\lambda}), k_1^{(1)}(\eta_i, \boldsymbol{\lambda}) \right\rangle \ . \tag{5.9}$$

Similar to covariate transformation, the case weight perturbation scheme actually generalizes case deletion, where ω_i is limited to the values 0 and 1. Furthermore, if the deletion of the ith case is of interest (as revealed by $\boldsymbol{l}_{max}$), it may be considered as the perturbation located in direction $\boldsymbol{l}_{[i]}$ from the null point, where $\boldsymbol{l}_{[i]}$ are the direction cosines with ith component -1 but zeros elsewhere. A plot of $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\omega}}$ in the direction $\boldsymbol{l}_{[i]}$ can then monitor the global effects of downweighting the ith case.

5.2.2 Perturbation of individual covariates x

We modify the j-th individual covariate, $\boldsymbol{x}_{(j)}$ of $\boldsymbol{x}$, to $\boldsymbol{x}_{(j)}(\boldsymbol{\omega}) = \boldsymbol{x}_{(j)} + t\boldsymbol{\omega}$ $(j = 1, \cdots, p)$, as long as the covariate is not an indicator variable. Here, t is the scaling factor used to convert the generic perturbation $\boldsymbol{\omega}$ to the appropriate size and units, and $\boldsymbol{\omega}_0 = \langle 0, \cdots, 0 \rangle^T$ represents no perturbation. It can be verified

that $\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\boldsymbol{\omega}}; \boldsymbol{\beta}_{\boldsymbol{\omega}}) = \langle \Delta_i^x \rangle$ where

$$\begin{aligned}\Delta_i^x =& \beta_j \Big\langle \Big\{ y_i k_{33}^{(2)}(\eta_i, \boldsymbol{\lambda}) - b^{(2)}(k(\eta_i, \boldsymbol{\lambda}))[k_3^{(1)}(\eta_i, \boldsymbol{\lambda})]^2 \\ & -b^{(1)}(k(\eta_i, \boldsymbol{\lambda}))k_{33}^{(2)}(\eta_i, \boldsymbol{\lambda}) \Big\} \boldsymbol{x}_i, \\ & \Big\{ y_i k_{13}^{(2)}(\eta_i, \boldsymbol{\lambda}) - b^{(2)}(k(\eta_i, \boldsymbol{\lambda}))k_1^{(1)}(\eta_i, \boldsymbol{\lambda})k_3^{(1)}(\eta_i, \boldsymbol{\lambda}) \\ & -b^{(1)}(k(\eta_i, \boldsymbol{\lambda}))k_{13}^{(2)}(\eta_i, \boldsymbol{\lambda}) \Big\} \boldsymbol{\lambda}^T \Big\rangle \\ & + \Big\{ y_i k_3^{(1)}(\eta_i, \boldsymbol{\lambda}) - b^{(1)}(k(\eta_i, \boldsymbol{\lambda}))k_3^{(1)}(\eta_i, \boldsymbol{\lambda}) \Big\} \boldsymbol{u}_j \ .\end{aligned} \tag{5.10}$$

Here, $k_3^{(1)}(\eta_i, \boldsymbol{\lambda})$ denotes the derivative of $k(\eta_i, \boldsymbol{\lambda})$ with respect to η_i, β_j is the regression coefficient associated with $\boldsymbol{x}_{(j)}$, and $\boldsymbol{u}_j$ is a $1 \times p$ row vector with jth component 1 but zeros elsewhere.

5.2.3 Perturbation of responses

We then consider altering the responses by taking $\boldsymbol{y}(\boldsymbol{\omega}) = \boldsymbol{y} + t\boldsymbol{\omega}$, t being the appropriate scaling factor. As with other additive perturbations, $\boldsymbol{\omega}_0 = \mathbf{0}$ gives the unperturbed state. We find that $\boldsymbol{\Delta}(\boldsymbol{\lambda}_{\boldsymbol{\omega}}; \boldsymbol{\beta}_{\boldsymbol{\omega}}) = \langle \Delta_i^y \rangle$, where

$$\Delta_i^y = \Big\langle k_2^{(1)}(\eta_i, \boldsymbol{\lambda}), k_1^{(1)}(\eta_i, \boldsymbol{\lambda}) \Big\rangle \ . \tag{5.11}$$

Note that this perturbation scheme may not be meaningful for discrete response, such as those in binary logistic regression.

5.3 Examples

5.3.1 Leukemia data

It is well known that leukemia is a type of cancer characterized by an excess of white blood cells. Cook and Weisberg (1982, p.179) reported the survival times in weeks and associated white blood cell counts (WBC) as measured at the time of diagnosis of 17 patients who died of *acute myelogenous leukemia*. Cook (1986) used this data set to illustrate likelihood displacement local influence in GLM, and assumed that the survival time y_i follows an exponential distribution with mean $\exp\{\delta_0 + \delta_1 x_i\}$, where $x_i = \log_{10}(\text{WBC}_i)$.

The assumed model

$$\log(\mu_i) = \delta_0 + \delta_1 x_i \tag{5.12}$$

can be embedded within the power transformation family

$$g(\mu_i, \lambda) = \beta_0 + \beta_1 x_i \ , \tag{5.13}$$

where

$$g(\mu, \lambda) = \begin{cases} \mu^\lambda & \text{if } \lambda \neq 0; \\ \log(\mu) & \text{if } \lambda = 0. \end{cases}$$

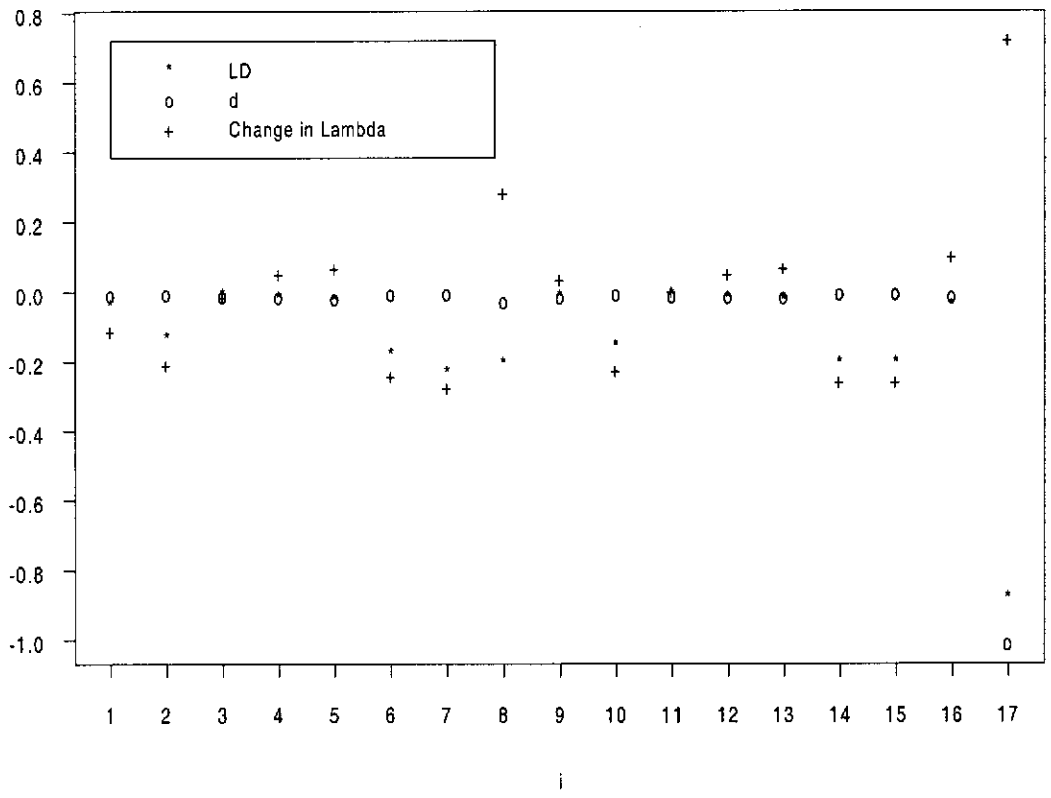
This family incorporates the identity, logarithmic, and reciprocal links as special cases, and is particularly applicable to positive responses. The MLE of δ_0 and δ_1 for model (5.12) are 8.48 (1.612), -1.11 (0.386), while the MLE of β_0, β_1 and λ for model (5.13) are 3.24 (0.499), -0.33 (0.117) and 0.16 (0.002), respectively, with asymptotic standard errors enclosed in parentheses.

Figure 5.1 shows the (re-scaled) profile log-likelihood displacement LD_i, partial influence measure d_i, and $\hat{\lambda}_{[i]} - \hat{\lambda}$, based on case deletions. Large values of all three measures are found for case 17. Its deletion ($\hat{\lambda}_{[17]} = 0.61$) actually strengthens the evidence for link modification. This is not surprising since case 17 corresponds to the only patient who survived a relatively long time (65 weeks)

with a high WBC count of 100,000. The local influence analysis by Cook (1986) also brought attention to case 17 as the single outlier in the sample.

Figure 5.1

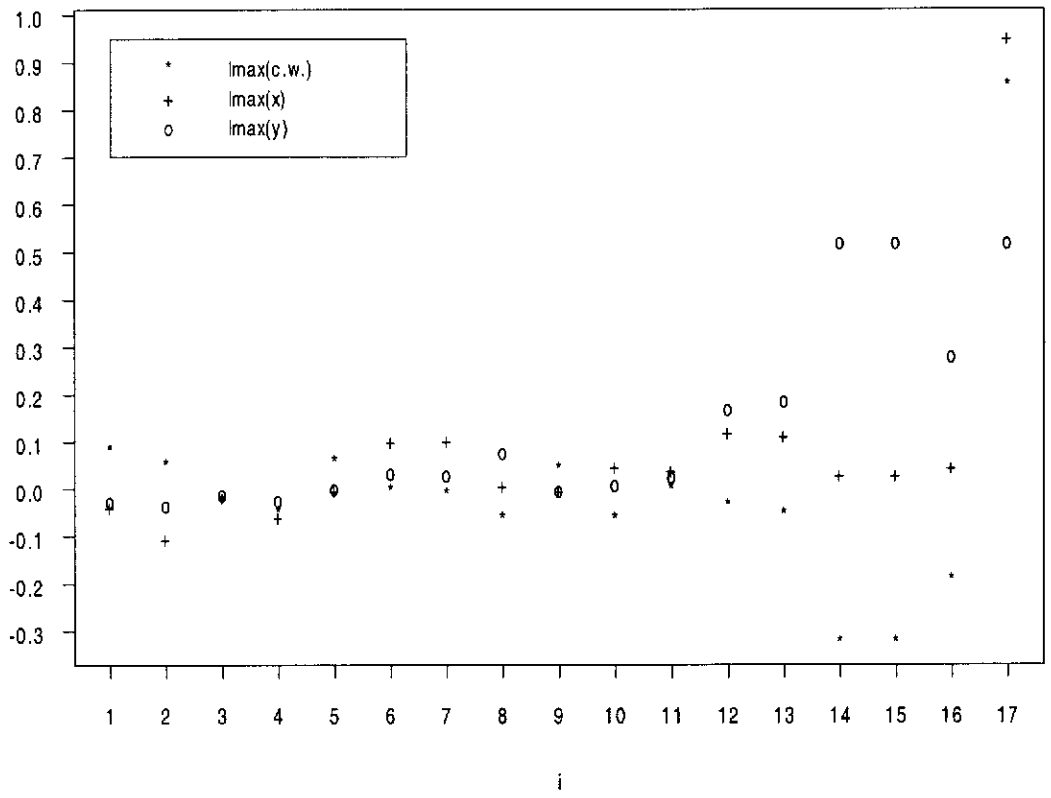
Case deletion diagnostics (rescaled) LD_i, d_i and $\hat{\lambda}_{[i]} - \hat{\lambda}$ for leukemia data.



We next examine the local effect of each case on $\hat{\lambda}$. Based on the partial influence approach, the direction cosines l^d_{max} from minor perturbations are plotted against case index in Figure 5.2.

Figure 5.2

Directional cosines $\boldsymbol{l}^d_{max}$ from local perturbations for leukemia data.

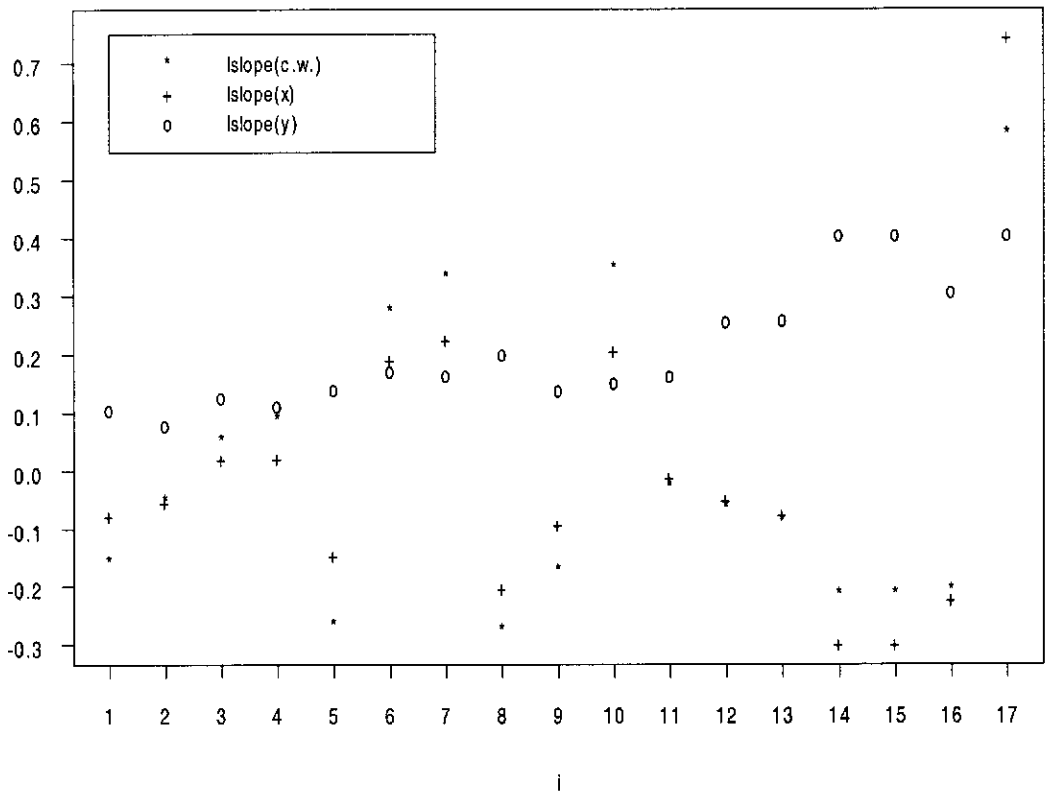


It is evident that the greatest local change in $\hat{\lambda}$ depends essentially on case 17 under case weight and covariate perturbation schemes. However, cases 14 and 15 also become discordant with respect to response perturbations. This result is consistent with the $\boldsymbol{l}_{slope}$ vectors displayed in Figure 5.3 using the first order approach, although the $\boldsymbol{l}_{slope}$ components appear to be less sensitive. Since λ is a scalar parameter, $\boldsymbol{l}_{slope} = \boldsymbol{l}^{\hat{\lambda}}_{max}$. It is interesting to note that cases 14 and 15 are identical, having exactly the same WBC count as case 17 but recording the shortest survival time of one week. Each of these two cases tends to mask the other's effect, so that their joint influence on the link parameter would not be apparent from case deletion diagnostics.

When either observation is excluded from the fit, $\hat{\lambda}$ changes from 0.16 to 0.01, but if both cases are removed, $\hat{\lambda}$ drops quite dramatically to -0.32.

Figure 5.3

Directional cosines $\boldsymbol{l}_{slope}$ from local perturbations for leukemia data.

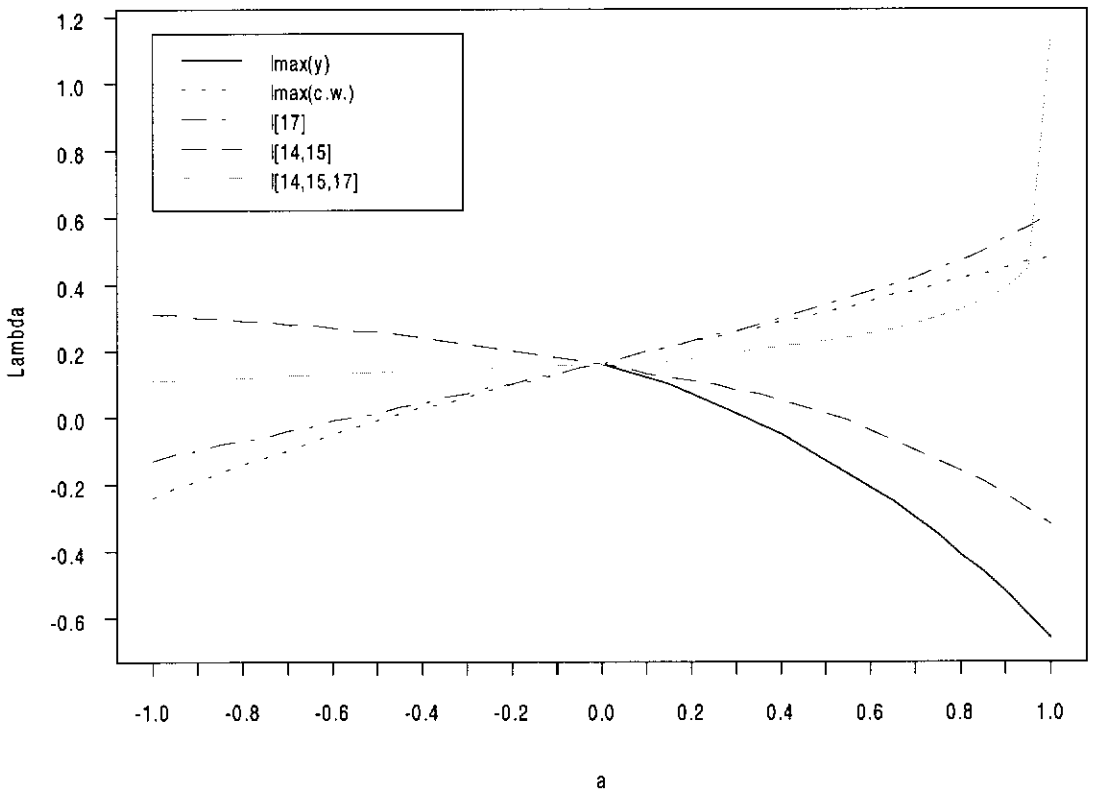


To further assess the global sensitivity of the link parameter, we plot the actual $\hat{\lambda}$ in Figure 5.4 for each local direction $\boldsymbol{l}$ of interest, including those related to the deletion of cases. It should be remarked that both curves associated with $\boldsymbol{l}^d_{max}$ have the greatest slope at $\mathrm{a} = 0$ (no perturbation). The curve resulting from response perturbations is only defined for $\mathrm{a} > 0$ such that the survival times are always positive. Furthermore, downweighting case 17 alone (in direction $\boldsymbol{l}_{[17]}$) produces opposite effects (yet similar in magnitude) to those of simultaneously modifying cases 14 and 15 (in direction $\boldsymbol{l}_{[14,15]}$). The maximum global change in $\hat{\lambda}$ is attained when the contributions of all three cases are reduced to zero (in direction $\boldsymbol{l}_{[14,15,17]}$). Upon their complete omission, $\hat{\lambda} = 1.15$, i.e. closed to the identity link function. The combined impact on $\hat{\lambda}$ is significant due to their identical and extreme WBC count. Fitting an identity link model to the

remaining data also results in a significant reduction in deviance of 12.5 (3 d.f.) Therefore, the exact form of the link function depends crucially on the extreme observations in this example.

Figure 5.4

$\hat{\lambda}$ in directions of local influence for leukemia data.



5.3.2 Erythrocyte sedimentation rate data

We further illustrate the proposed diagnostics with data from Collett (1991, p. 8) relating the chronic disease state y_i (0 = healthy; 1 = unhealthy) of 32 individuals, judged from the *erythrocyte sedimentation rate* (ESR) reading, to the plasma fibrinogen level x_i (in gm/ℓ). The assumed logistic regression model

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \delta_0 + \delta_1 x_i \tag{5.14}$$

can be embedded within the Aranda-Ordaz (1981) family

$$g(\mu_i, \lambda) = \beta_0 + \beta_1 x_i \ , \tag{5.15}$$

where the link function takes the form

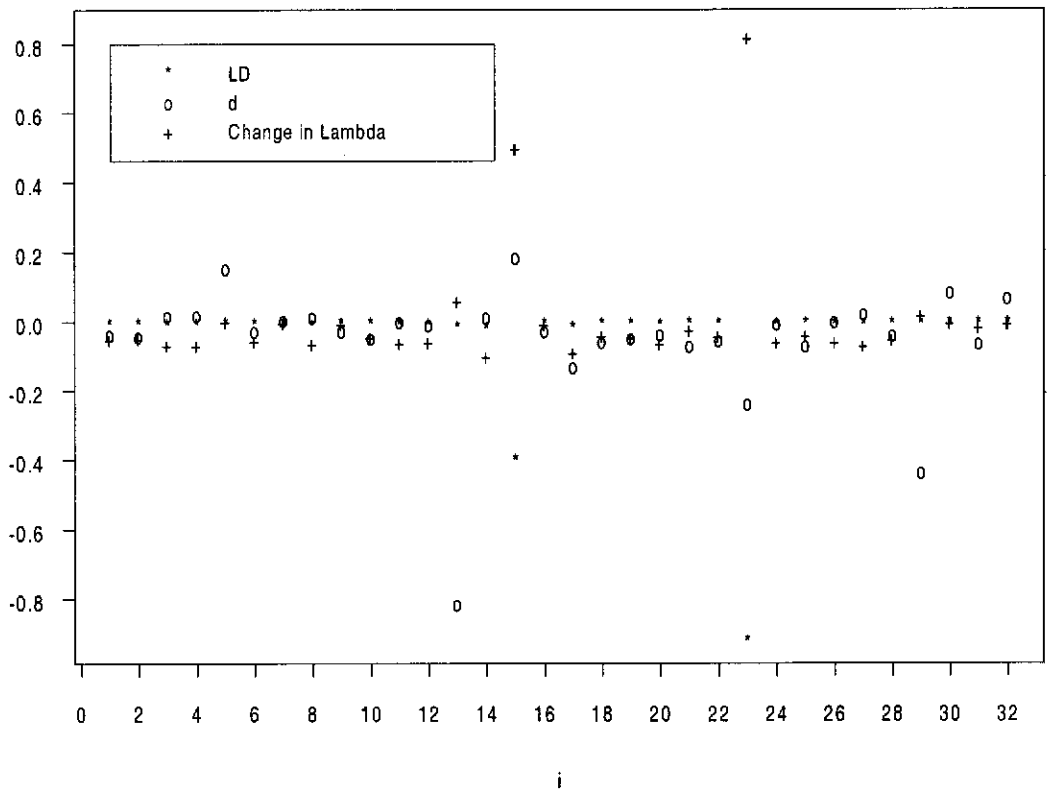
$$g(\mu, \lambda) = \log\left\{\frac{(1-\mu)^{-\lambda} - 1}{\lambda}\right\} \ .$$

This family incorporates the logistic ($\lambda = 1$) and complementary log-log ($\lambda = 0$) links as special cases. The MLE of δ_0 and δ_1 for model (5.14) are -6.845 (2.764), 1.827 (0.899), while the MLE of β_0, β_1 and λ for model (5.15) are -4.796 (2.218), 1.106 (0.727) and 0.552 (0.012), respectively, with asymptotic standard errors enclosed in parentheses.

Figure 5.5 shows the (re-scaled) profile log-likelihood displacement LD_i, partial influence measure d_i, and $\hat{\lambda}_{[i]} - \hat{\lambda}$, based on case deletions. Outstanding values of LD_i and $\hat{\lambda}_{[i]} - \hat{\lambda}$ are observed for cases 15 and 23, whereas d_i brought our attention to case 13 as well.

Figure 5.5

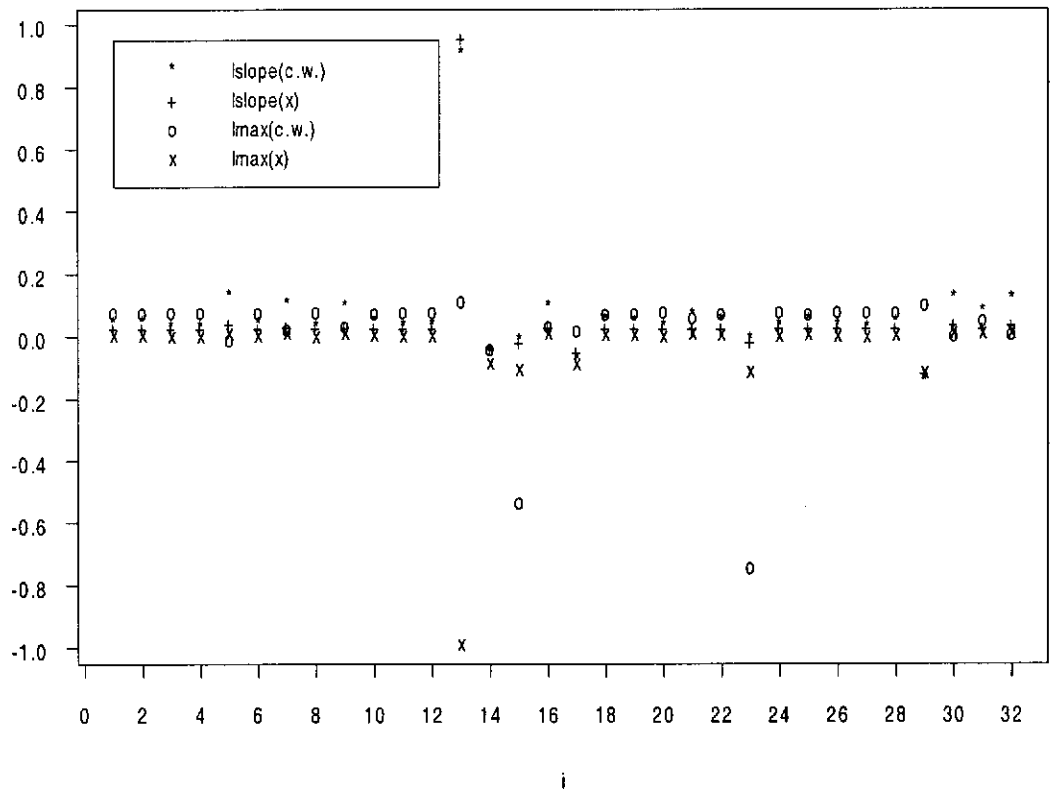
Case deletion diagnostics (rescaled) LD_i, d_i and $\hat{\lambda}_{[i]} - \hat{\lambda}$ for ESR data.



We next examine local influence on the link parameter. The direction cosines from local perturbations are plotted against case index in Figure 5.6. We note that case 13 has outlying cosines according to the first/second order diagnostics. Based on the partial influence approach, however, cases 15 and 23 have large l^d_{max} components when case weights are perturbed. An inspection of the data reveals that case 13 is a leverage point recording the highest fibrinogen counts of 5.06 gm/ℓ in the sample. On the other hand, case 15 and case 23 correspond to two unhealthy patients with relatively low fibrinogen levels, despite the plasma protein concentrations will generally rise under inflammatory disease conditions.

Figure 5.6

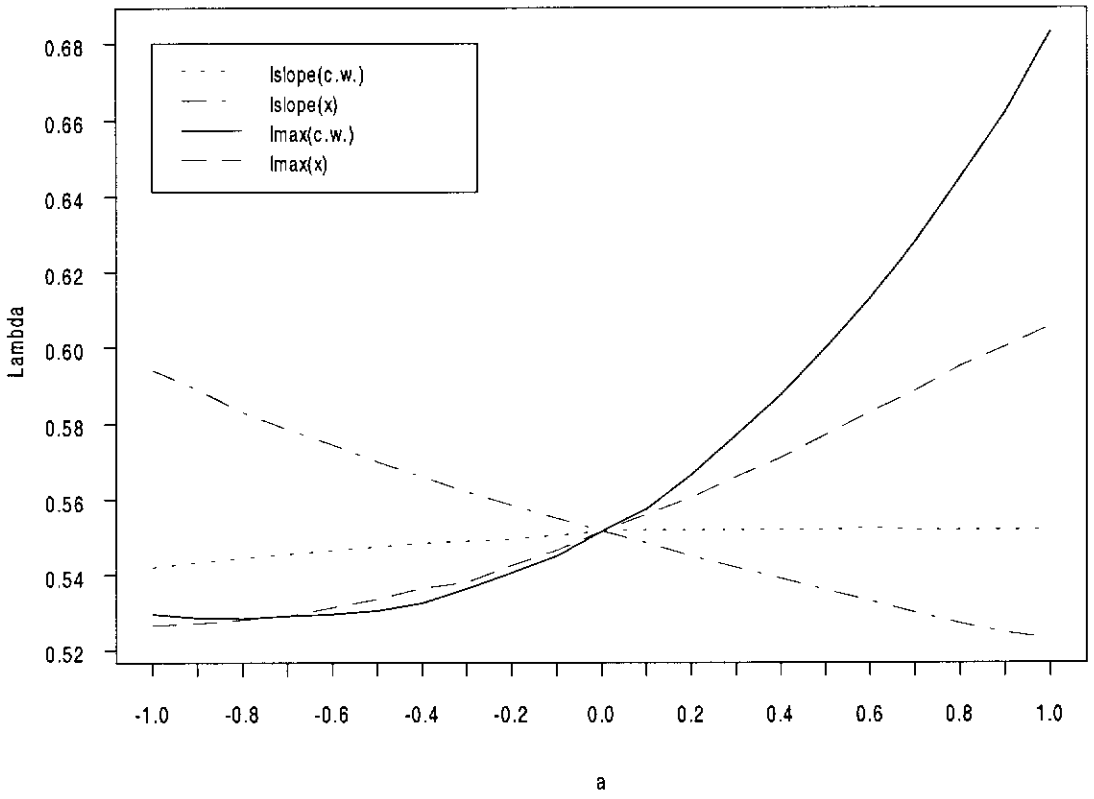
Directional cosines $\boldsymbol{l}_{slope}$ and $\boldsymbol{l}^{d}_{max}$ from local perturbations for ESR data.



To assess the extent of the local perturbations, we plot $\hat{\lambda}$ against a in Figures 5.7 and 5.8 for various local directions of interest. It is found that the effects of simultaneously modifying all case weights are less than those of directly down-weighting the aberrant cases. Moreover, the curves associated with the $\boldsymbol{l}_{slope}$ directions induce minor global changes. The net overall change due to covariate perturbations is not dramatic because the impact exerted by case 13 apparently has been compensated by the other cases.

Figure 5.7

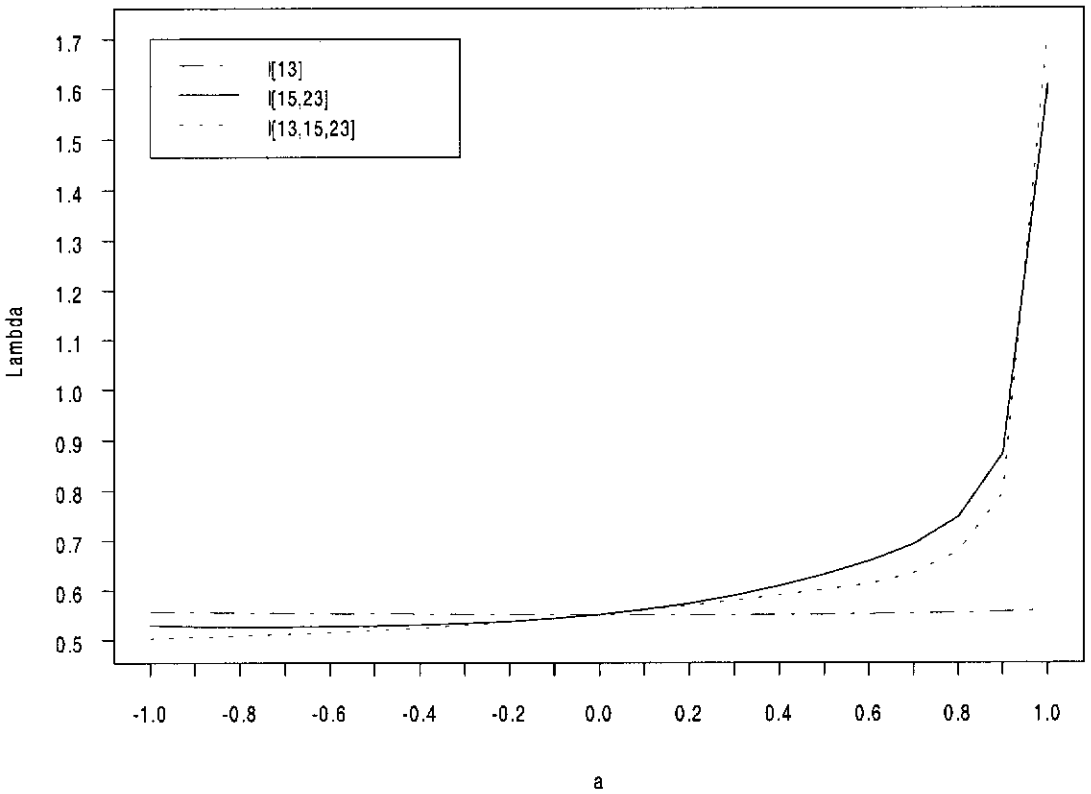
$\hat{\lambda}$ in directions of local influence for ESR data.



Indeed, the link parameter is quite insensitive to the downweighting of case 13 alone, as confirmed by the curve in direction $\boldsymbol{l}_{[13]}$. Finally, fitting model (5.15) without cases 15 and 23 results in a significant reduction in deviance of 19.17 (2 d.f.), with $\hat{\lambda} = 1.608$ (0.016), $\hat{\beta}_0 = -39.78$ (20.6), and $\hat{\beta}_1 = 11.5$ (6.06). The link analysis thus provides useful insights for model revision.

Figure 5.8

$\hat{\lambda}$ in directions of downweighting specific cases for ESR data.



CHAPTER 6

PORTMANTEAU STATISTIC IN TIME SERIES

6. PORTMANTEAU STATISTIC IN TIME SERIES

In many applications of time series, sample data are used to estimate the parameters of the assumed model, and structural relationships are tested statistically. It is common that estimation and the overall goodness-of-fit of an ARMA model may be affected by one or several atypical observations. The unrecognized abnormality will lead to poor forecasts based on the estimated model (Ledolter (1989), Chen and Liu, (1993)). The outliers also affect the sensitivity analysis where the effects of minor changes to the data are monitored. It is therefore important to determine whether the conformance of the hypothesized model is achieved throughout the series or distorted by a few particular observations. The aim of this chapter is to present effective diagnostics for assessing the effects of aberrant observations on the portmanteau statistic.

The investigation of residuals has been well established in regression diagnostics. In time series analysis it is the residual autocorrelations that should be examined. A widely used diagnostic for checking overall model adequacy is the portmanteau statistic Q (Box and Pierce, 1970) which accumulates the lag K squared residual autocorrelations. However, the portmanteau statistic and its variant such as the Ljung-Box-Pierce statistic (Ljung and Box, 1978) are expected to be prone to outliers. It has been documented in the literature that the existence of additive outliers can seriously bias the model coefficients, whereas innovational outliers in general have a much smaller effect; see Ljung (1993). Therefore, we limit our scope to additive outliers and assess the effect of perturbations on the portmanteau statistic in ARMA models. Two separate approaches based on global deletion and local influence analysis are proposed in the evaluations. The resulting diagnostics are then demonstrated using several examples.

6.1 Assessing goodness-of-fit

We are concerned with diagnostic methods for assessing the influence of multiple and/or consecutive outliers on the goodness-of-fit through the portmanteau statistic. Consider the stationary and invertible ARMA model

$$\phi(B)y_t = \theta(B)a_t$$

where B denotes the backward shift operator,

$$\phi(B) = 1 - \phi_1 B - ... - \phi_p B^p \quad \text{and} \quad \theta(B) = 1 - \theta_1 B - ... - \theta_q B^q,$$

$\phi(B)$ and $\theta(B)$ have all their roots outside the unit circle, and a_t is Gaussian white noise with zero mean and innovation variance σ^2. The well known *portmanteau statistic* (Box and Pierce, 1970) for testing goodness or lack of fit is

$$Q = n \sum_{k=1}^{K} r_k^2[\hat{a}]$$

where $r_k[\hat{a}]$ is the lag k autocorrelation of the residuals $\hat{a}$'s. If the orders (p, q) are correctly specified and $n > K$, then Q is distributed asymptotically as χ^2 with degrees of freedom $K - p - q$. To improve the χ^2 approximation of its null distribution, several variants of the portmanteau statistic have been proposed in the literature, including the Ljung-Box-Pierce statistic (Ljung and Box, 1978)

$$Q^* = n(n+2) \sum_{k=1}^{K} (n-k)^{-1} r_k^2[\hat{a}].$$

The additional terms involving n and k, however, may be regarded as nuisance parameters in the assessment of local influence in Section 6.1.2. Without loss of generality, we shall focus on the standard form Q in subsequent investigations. The effect of a change in innovation variance on Q has been studied by Inclán (1992). Meanwhile, robustified modifications of the portmanteau statistic for testing model adequacy have also been proposed (Li (1988), Chan (1994), Jiang, Hui and Zheng (1999)).

6.1.1 Global influence

A practical approach to sensitivity analysis involves the removal of individual cases; see for example Cook and Weisberg (1982) for a review. However, ordinary case deletion is inappropriate in time series. The problem can be effectively handled by treating the observation as missing data prior to parameter estimation. To predict the missing values, we follow the state space representation of Kohn and Ansley (1986) which provides consistent and asymptotically efficient estimates. Alternatively, the E-M Algorithm or other methods can be used to impute the missing values. A summary of such procedures can be found in Basu and Reinsel (1996). Based on the missing data approach, global influence for the portmanteau statistic $Q_{(t)}$ is evaluated from the incomplete series $\langle y_1, ..., y_{t-1}, y_{t+1}, ..., y_n \rangle$. It can be quantified with respect to the asymptotic χ^2 reference distribution.

6.1.2 Local influence

Ledolter (1990) applied the local influence method to outliers detection in time series via the normal curvature of the likelihood displacement surface. It was found that the resulting diagnostic for which the curvature is maximized is given by the vector of differences between observed and interpolated values. Instead of relying on the likelihood displacement criterion, we study the effect of additive-outlier perturbations on the portmanteau statistic. Consider the additive outlier perturbation ARMA model

$$y_t = z_t + d\omega_t \qquad \phi(B)z_t = \theta(B)a_t$$

where y_t is the observed value, $\boldsymbol{\omega} = \langle \omega_1, ..., \omega_n \rangle^T \in \Omega$ are the perturbations with scale d, and Ω denotes the open set of relevant perturbations. In the manner of Wu and Luo (1993a), the geometric surface of interest is formed by $Q(\boldsymbol{\omega})$, the portmanteau statistic under perturbation $\boldsymbol{\omega}$. Here the null point

$\omega_0 = \langle 0, ..., 0 \rangle^T \in \Omega$ represents no perturbation so that $Q(\omega_0)$ gives the observed Q statistic. Unlike the likelihood displacement surface, such a perturbation-formed portmanteau statistic surface $Q(\omega)$ does not have zero first derivative at ω_0, so that its slope can be used to examine local influence. To locate the direction of largest local change, we approximate $Q(\omega)$ by its tangent plane at ω_0. The desired direction is then given by the direction of maximum slope on this tangent plane over Ω. Such a direction vector, $Q' = \partial Q(\omega)/\partial \omega^T$ evaluated at ω_0, will serve as our diagnostic tool. It can be shown that

$$\langle Q'_t \rangle = \sum_{k=1}^{K} \left(S[k]_t \sum_{i}^{n} \hat{a}_i^2 - 2\hat{a}_t \sum_{i}^{n} \hat{a}_i \hat{a}_{t-k} \right) / \left(\sum_{i}^{n} \hat{a}_i^2 \right)^2$$

where the $S[k]_t$ terms are model dependent, formulae of which are given below. Sampling properties of these direction vectors remain to be developed, but do not seem crucial at the diagnostic stage. Warning limits of $\pm Z_\alpha/\sqrt{n}$ may be used for informal calibration (Lawrance, 1988).

AR(1) model with n observations

$S[k]_1 = -\hat{\phi}\hat{a}_{k+2}$

For $1 < t < k+1$, $S[k]_t = \hat{a}_{t+k} - \hat{\phi}(\hat{a}_{t+k+1})$

$S[k]_{k+1} = \hat{a}_{2k+1} - \hat{\phi}(\hat{a}_{2k+2} + \hat{a}_2)$

For $k+1 < t < n-k$, $S[k]_t = \hat{a}_{t-k} + \hat{a}_{t+k} - \hat{\phi}(\hat{a}_{t-k+1} + \hat{a}_{t+k+1})$

$S[k]_{n-k} = \hat{a}_{n-2k} + \hat{a}_n - \hat{\phi}(\hat{a}_{n-2k+1})$

For $n-k < t < n$, $S[k]_t = \hat{a}_{t-k} - \hat{\phi}(\hat{a}_{t-k+1})$

$S[k]_n = \hat{a}_{n-k}$

MA(1) model with n observations

$S[k]_1 = --$

For $1 < t < k+1$, $S[k]_t = \hat{\theta} S[k]_{t+1} + \hat{a}_{t+k} - \hat{a}_{t+k+1}$

$S[k]_{k+1} = \theta S[k]_{k+2} + \hat{a}_{2k+1} - \hat{a}_{2k+2} - \hat{a}_2$

For $k+1 < t < n-k$, $S[k]_t = \hat{\theta} S[k]_{t+1} + \hat{a}_{t-k} + \hat{a}_{t+k} - \hat{a}_{t-k+1} - \hat{a}_{t+k+1}$

$S[k]_{n-k} = \hat{\theta} S[k]_{n-k+1} + \hat{a}_{n-2k} + \hat{a}_n - \hat{a}_{n-2k+1}$

For $n-k < t < n$, $S[k]_t = \hat{\theta} S[k]_{t+1} + \hat{a}_{t-k} - \hat{a}_{t-k+1}$

$S[k]_n = \hat{a}_{n-k}$

ARMA(1, 1) model with n observations

$S[k]_1 = --$

For $1 < t < k+1$, $S[k]_t = \hat{\theta} S[k]_{t+1} + \hat{a}_{t+k} - \hat{\phi}(\hat{a}_{t+k+1})$

$S[k]_{k+1} = \hat{\theta} S[k]_{k+2} + \hat{a}_{2k+1} - \hat{\phi}(\hat{a}_{2k+2} + \hat{a}_2)$

For $k+1 < t < n-k$, $S[k]_t = \hat{\theta} S[k]_{t+1} + \hat{a}_{t-k} + \hat{a}_{t+k} - \hat{\phi}(\hat{a}_{t-k+1} + \hat{a}_{t+k+1})$

$S[k]_{n-k} = \hat{\theta} S[k]_{n-k+1} + \hat{a}_{n-2k} + \hat{a}_n - \hat{\phi}(\hat{a}_{n-2k+1})$

For $n-k < t < n$, $S[k]_t = \hat{\theta} S[k]_{t+1} + \hat{a}_{t-k} - \hat{\phi}(\hat{a}_{t-k+1})$

$S[k]_n = \hat{a}_{n-k}$

6.1.3 Global effect of local perturbations

Since Q_t' signifies how to perturb the postulated model to obtain the greatest local change in Q, the sensitivity of Q to the induced perturbations in direction Q_t' can be further assessed by profiling $Q(\omega)$ against the perturbation scale d. The characteristics of this curve should be informative for further investigation on the relationship between local and global influences.

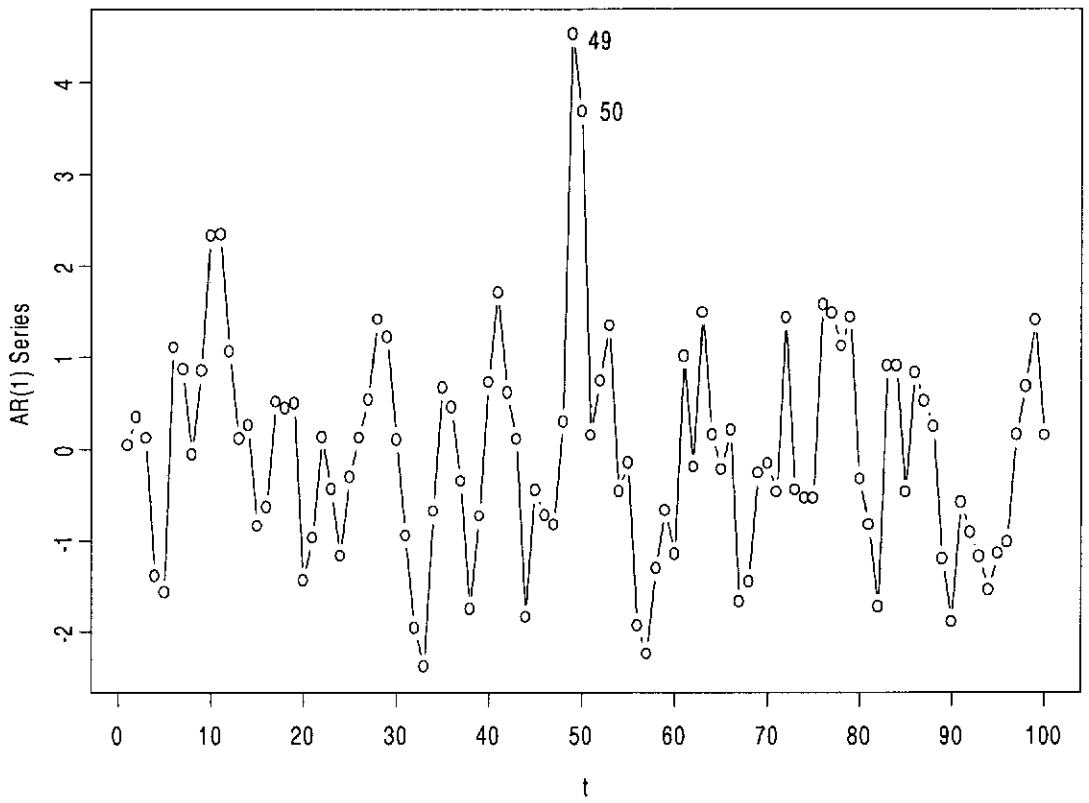
6.2 Applications

6.2.1 AR(1) with consecutive outliers: Simulated series

An artificial AR(1) series of $n = 100$ observations is generated using S-plus function `arima.sim` with $\phi = 0.5$ and innovation variance $\sigma^2 = 1$.

Figure 6.1

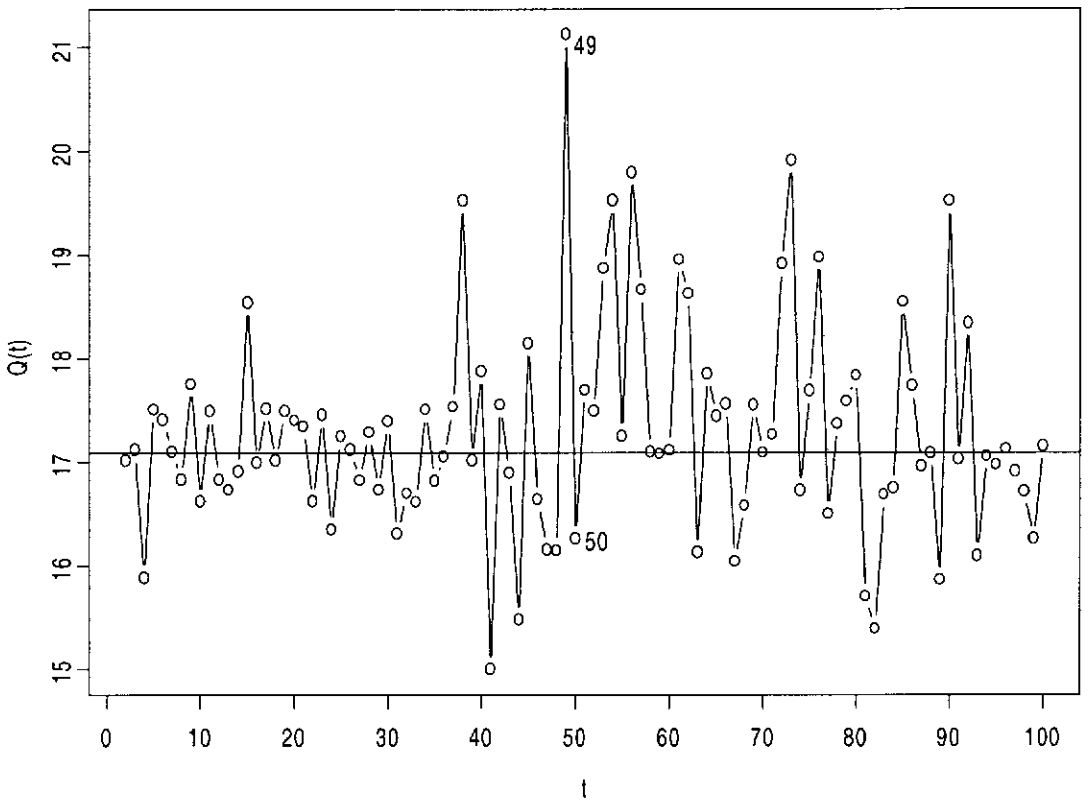
Contaminated AR(1) series.



Fitting the AR(1) model gives $\hat{\phi} = 0.497$ (0.087), $\hat{\sigma}^2 = 0.851$. The portmanteau statistic $Q = 16.24$ at $k = 20$ is readily available from `arima.diag`. To ensure stability of Q, henceforth K is taken to be 20 lags. Two consecutive additive outliers are created by adding 3 to the 49th and 50th observations. The contaminated series, plotted in Figure 6.1, has $\hat{\phi} = 0.507$ (0.087), $\hat{\sigma}^2 = 1.074$, and $Q(\omega_0) = 17.09$.

Figure 6.2

$Q_{(t)}$ for contaminated AR(1) series.



To assess global influence, missing values are introduced one-at-a-time. The resulting likelihood is maximized based on Kalman filtering applied to its state space representation (Kohn and Ansley, 1986). The method of initializing the Kalman filter recursions is that given by Bell and Hillmer (1987). Figure 6.2 shows $Q_{(t)}$ with a horizontal reference line drawn at the null state value $Q(\omega_0)$. As expected $Q_{(49)}$ is relatively large in the time series plot but case 50 does not appear prominent, an artifact of the masking effect due to the adjacent outliers.

Figure 6.3

Q'_t for contaminated AR(1) series.

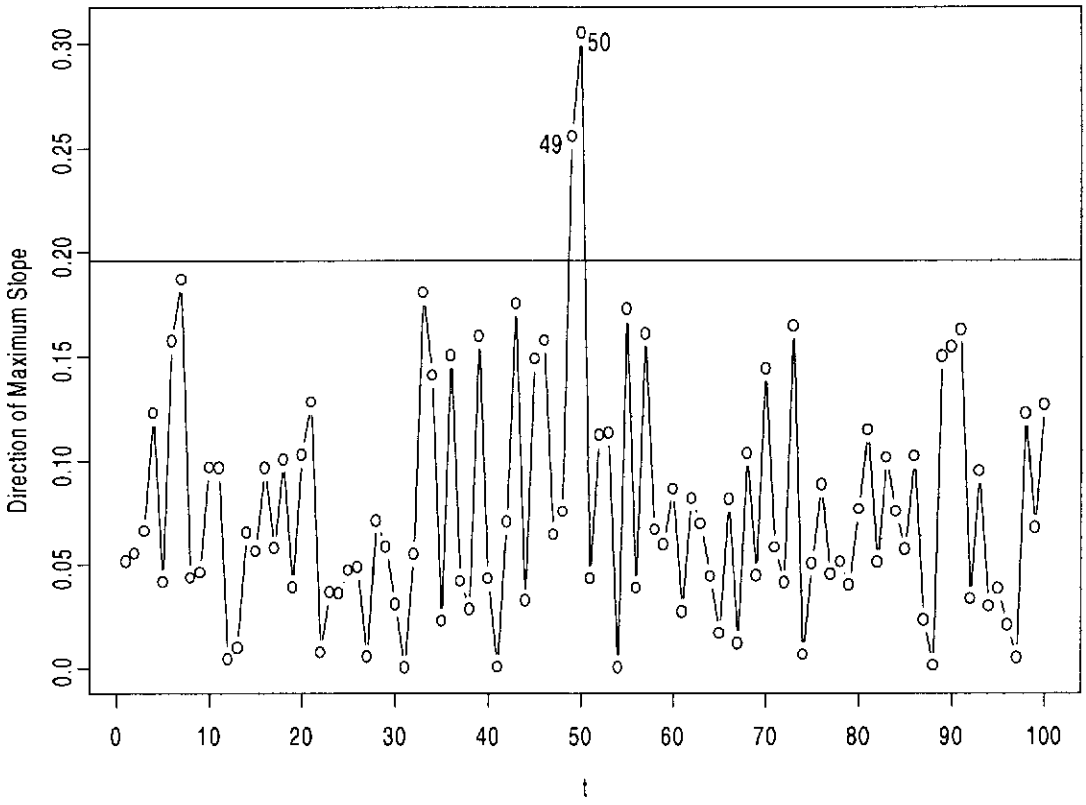
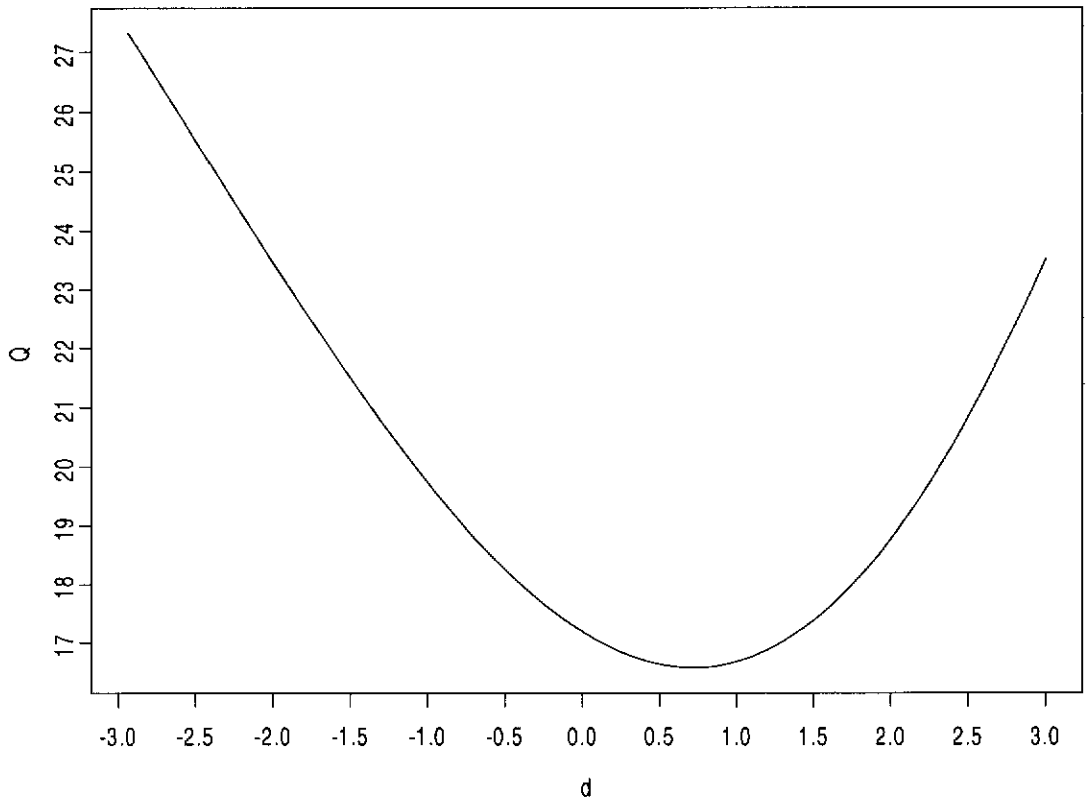


Figure 6.3 plots the absolute value of the normalized diagnostic Q'_t. Both spurious observations located at time points 49 and 50 have exceeded the warning limit of $\frac{1.96}{\sqrt{n}} = 0.196$, suggesting they are locally influential on the test statistic. We next plot $Q(\boldsymbol{\omega})$ against the perturbation scale d for the direction Q' in Figure 6.4, where the range of d is chosen to be $\pm 3\hat{\sigma}$. It can be seen that a small perturbation about the size of $\frac{1}{2}\hat{\sigma}$ in direction Q'_t (which is dominated by its 49th and 50th components) would help reduce $Q(\boldsymbol{\omega}_0)$ towards the value of 16.24 prior to contamination.

Figure 6.4

$Q(\omega)$ in direction Q' for contaminated AR(1) series.

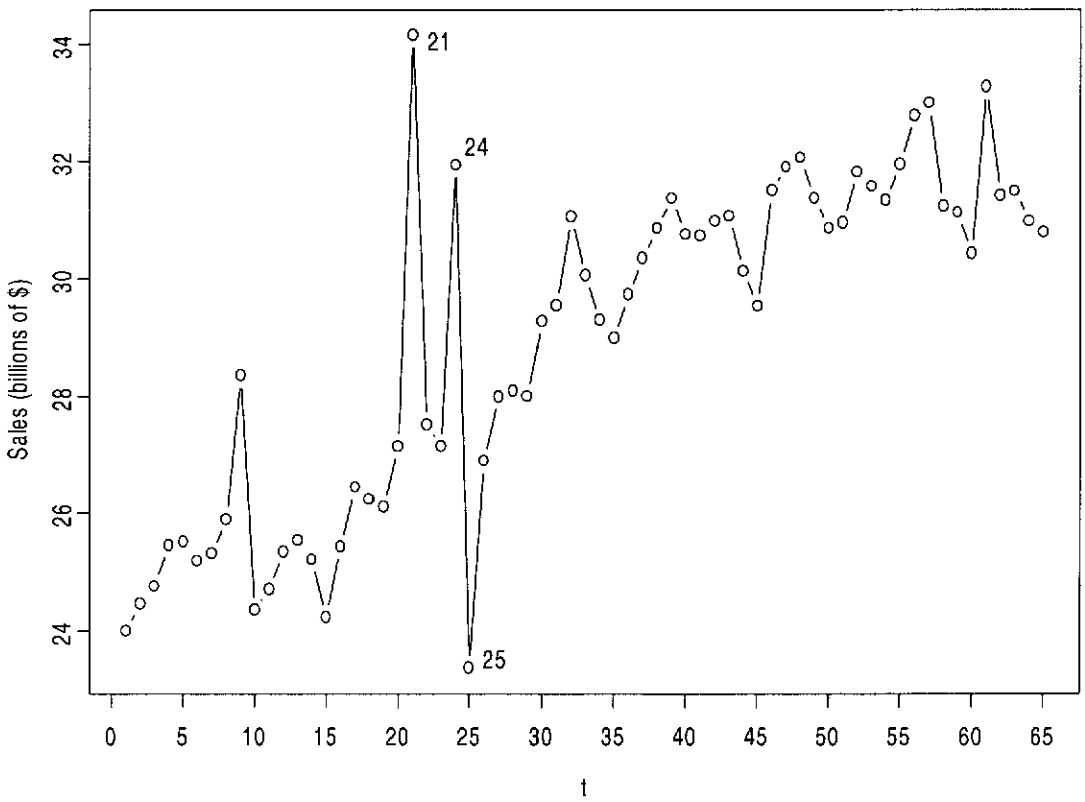


6.2.2 IMA(1,1) with reallocation outliers: Retail sales of automobile dealers

The term *reallocation outliers* is due to Wu, Hosking and Ravishanker (1993), which can be considered as additive outliers whose magnitudes sum to zero. They used 65 monthly observations (January 1985 to May 1990) from the estimated retail sales of automotive dealers (seasonally adjusted) published in the *Survey of Current Business Statistics*, US Department of Commerce.

Figure 6.5

Retail sales of automobile dealers.



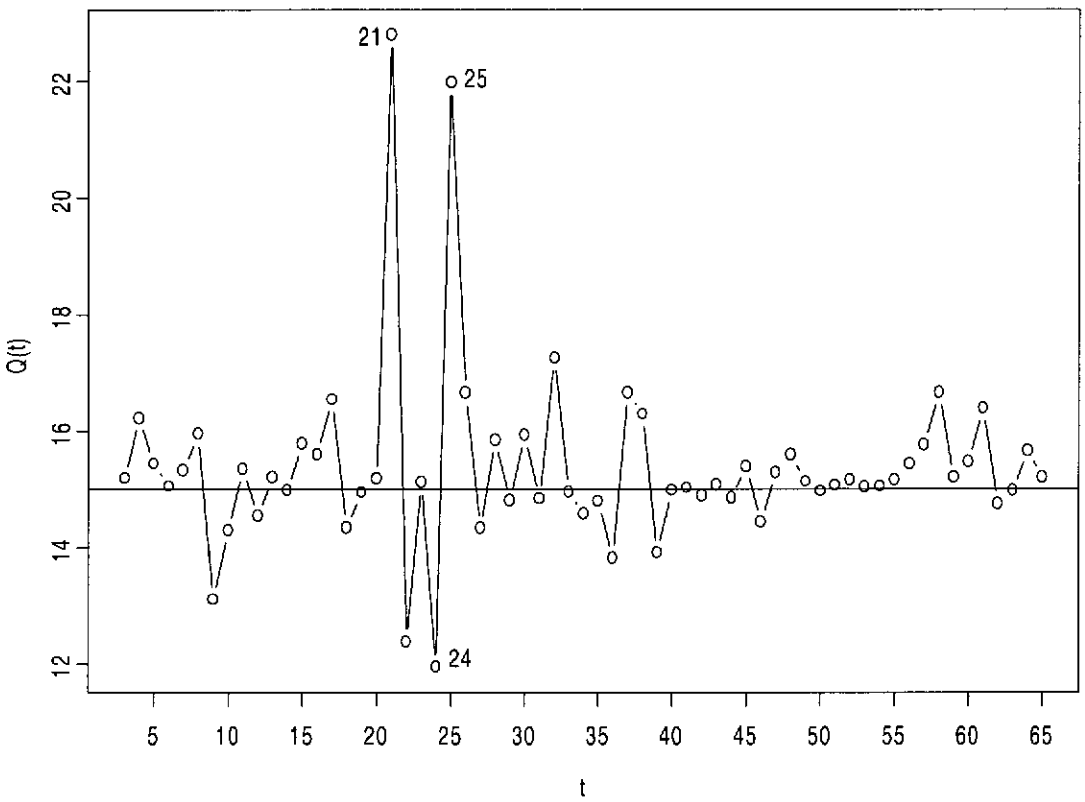
For this series, shown in Figure 5.5, Wu et al. (1993) identified the observation for September 1986 ($t = 21$) as a single additive outlier, whereas observations for December 1986 ($t = 24$) and January 1987 ($t = 25$) constituted a reallocation outlier pair: sales in September 1986 appeared to be unusually high,

sales in the following December and January were merely a reallocation with no overall change in sales volume. The model fitted is IMA(1, 1), with estimated mean 0.118 (0.028), $\hat{\theta} = 0.868$ (0.079), $\hat{\sigma}^2 = 2.405$ and $Q(\omega_0) = 15.01$.

The plot of $Q_{(t)}$ in Figure 6.6 suggests that the observed portmanteau statistic can be substantially distorted by the additive outliers. Indeed, if all three observations are treated as missing, Q increases to 19.51.

Figure 6.6

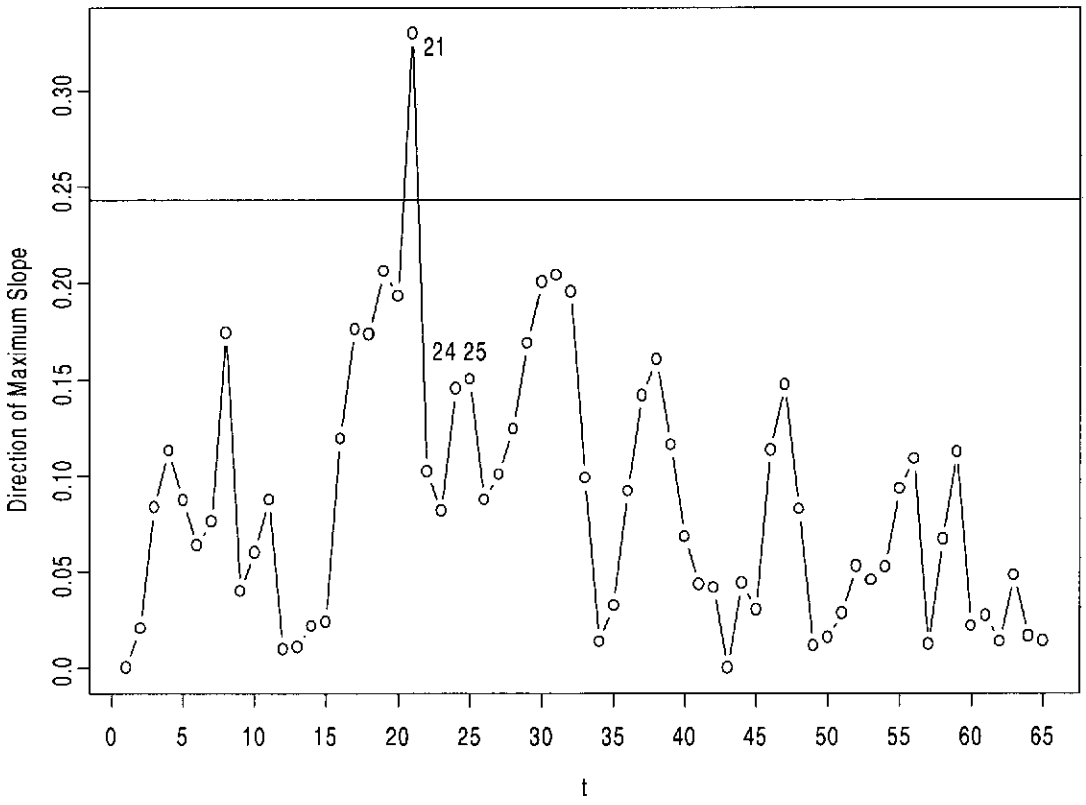
$Q_{(t)}$ for retail sales series.



However, only Q'_{21} in Figure 6.7 is appreciably greater than the warning limit of 0.243, which supports the classification of observation 21 as a single additive outlier. The anomaly in the process occurring at the successive time points 24 and 25 exerts minor impact on Q locally, which is consistent with the reallocation property that their combined net disturbances to the series being zero.

Figure 6.7

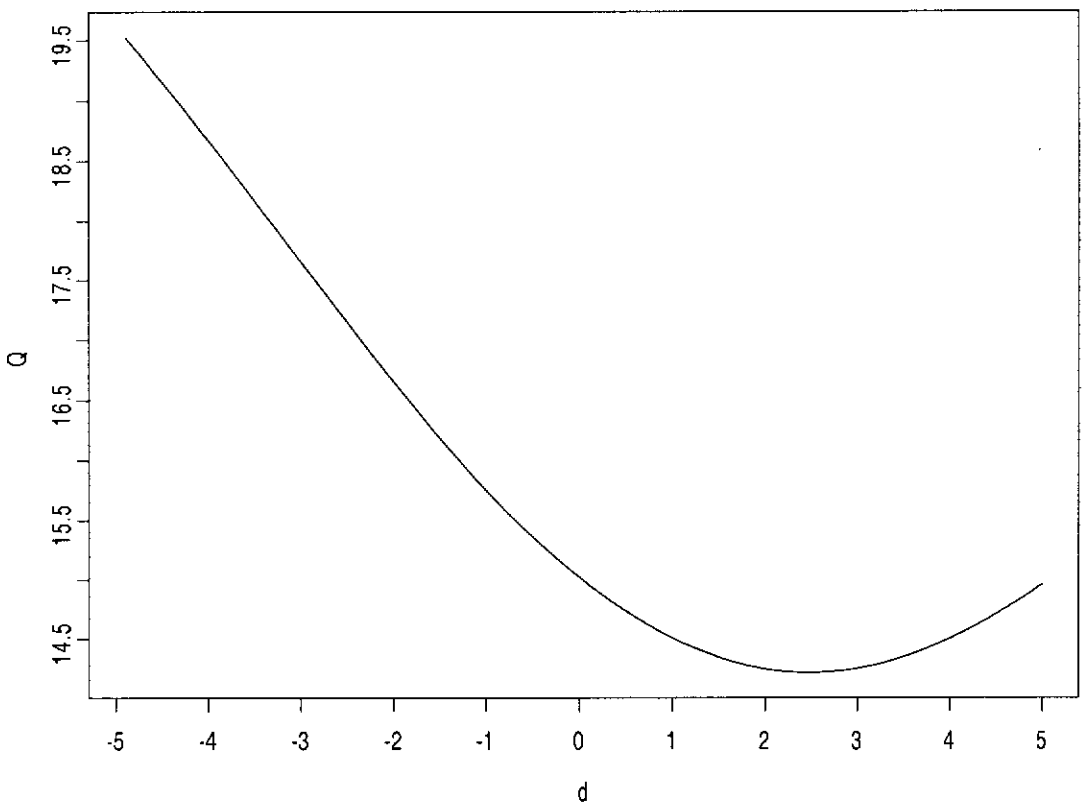
Q'_t for retail sales series.



To further quantify the global effects of the local perturbations, $Q(\omega)$ is plotted against d in Figure 6.8. It is worth noting that an optimal Q can be attained by perturbing the series simultaneously in direction Q' with a magnitude of about $\hat{\sigma}^2$.

Figure 6.8

$Q(\omega)$ in direction Q' for retail sales series.

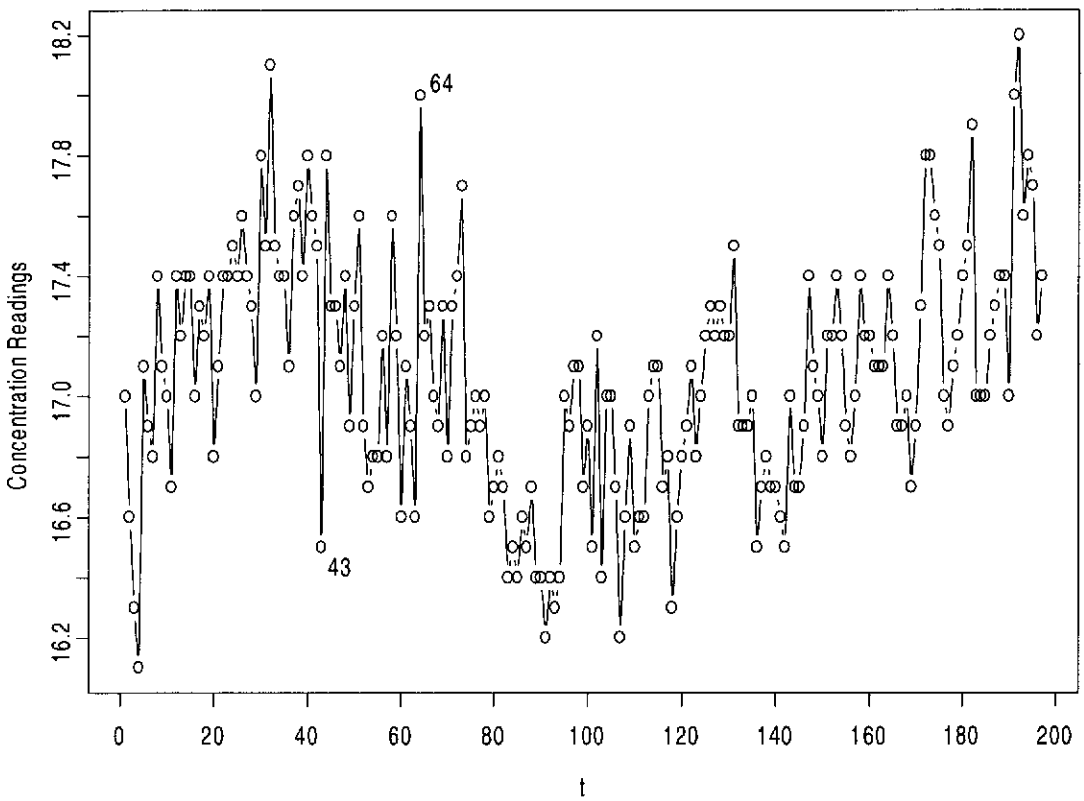


6.2.3 ARMA(1,1) with isolated outliers: Chemical process concentration readings

We next consider the Series A taken from Box, Jenkins and Reinsel (1994) which contains 197 readings of concentration in a chemical process observed every two hours (Figure 6.9). Box et al. (1994, p. 214) fitted an ARMA(1,1) model to this series, with $\hat{\phi} = 0.921$ (0.042), $\hat{\theta} = 0.581$ (0.083), $\hat{\sigma}^2 = 0.098$, and $Q = 25.37$ at $k = 20$ indicates the ARMA(1, 1) model has provided a reasonable fit to the data.

Figure 6.9

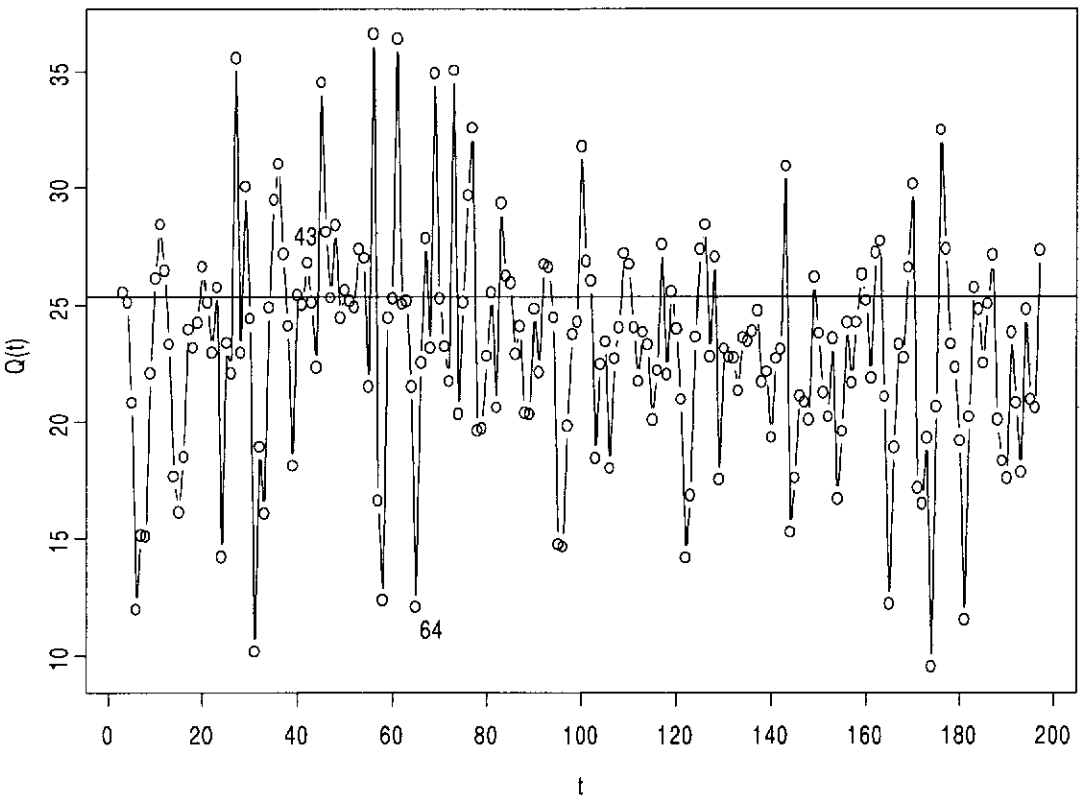
Chemical process concentration readings.



However, using an iterative robust fitting procedure, Luceño (1998) found two potential isolated outliers at times $t = 43$ and $t = 64$. It is therefore of interest to scrutinize the contribution of such isolated outliers to the overall fit in terms of the portmanteau statistic.

Figure 6.10

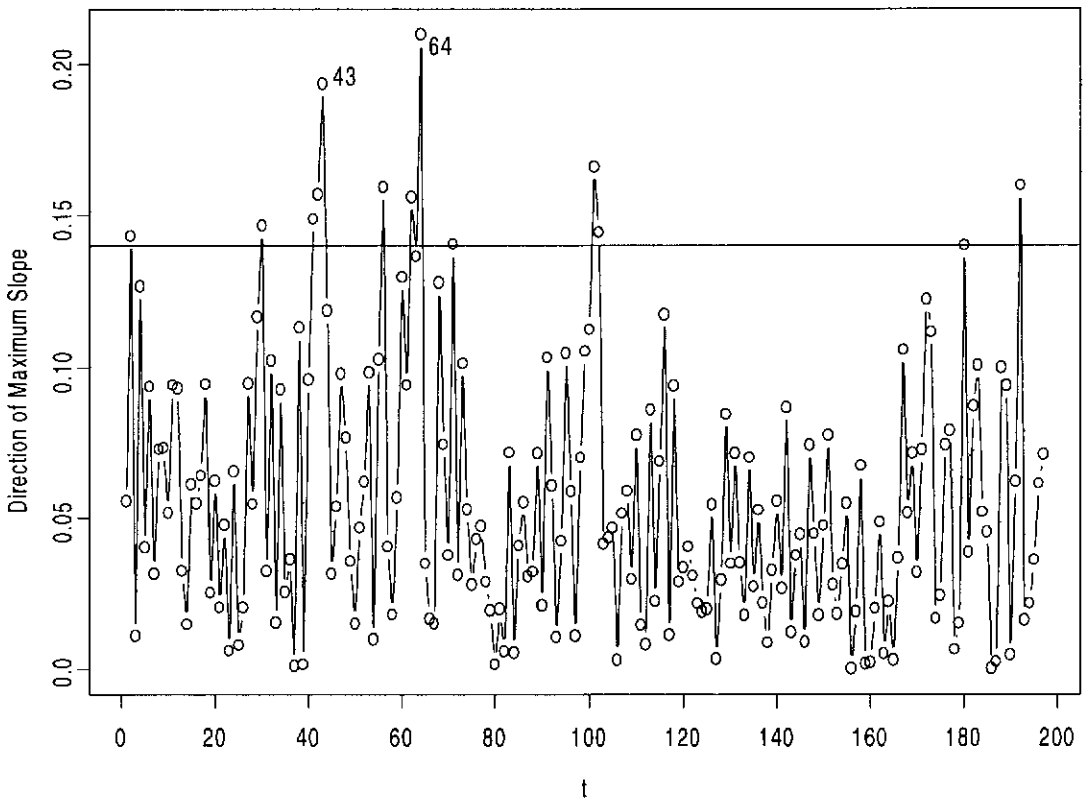
$Q_{(t)}$ for chemical process series.



From Figure 6.11, we found that the greatest local change in Q as measured by Q' depends to a large extent on observations 43 and 64, whose components are well above the warning limit of 0.14. But according to the $Q_{(t)}$ statistic presented in Figure 6.10, both points are not flagged as globally influential.

Figure 6.11

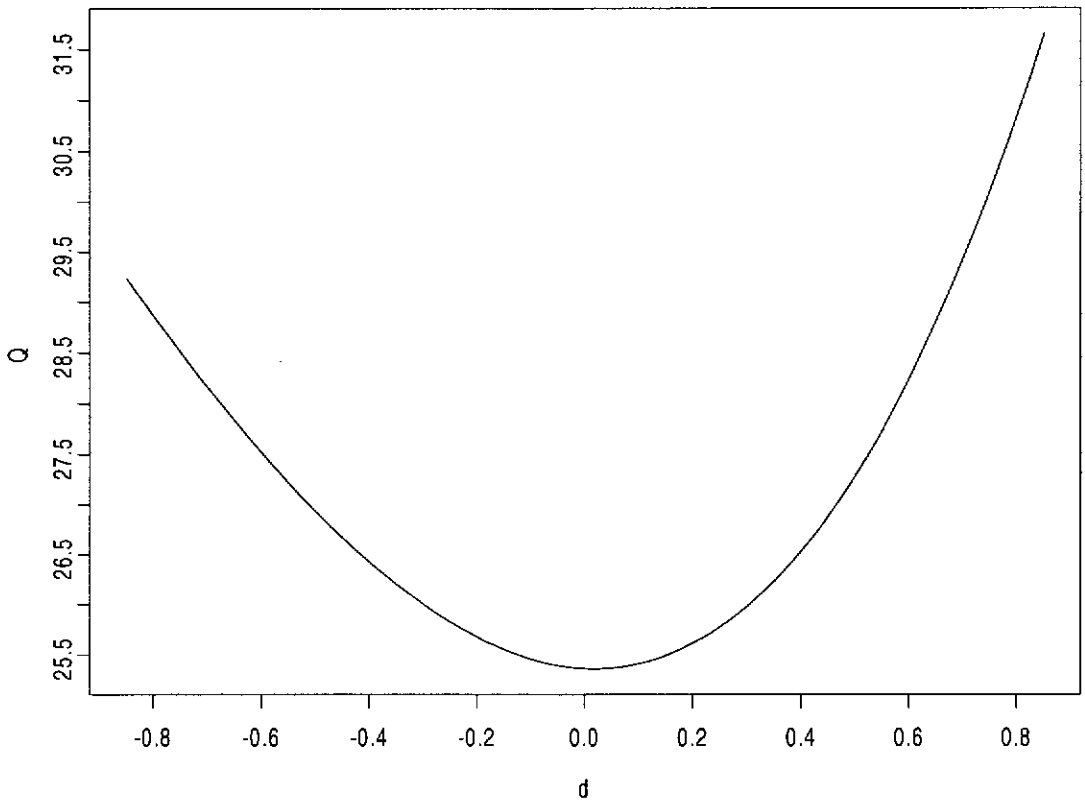
Q'_t for chemical process series.



To confirm the indications of the diagnostics, we examine the actual Q displacements due to the local perturbations. It is evident from Figure 6.12 that no overall improvement in the goodness-of-fit can be achieved by perturbing the series in the neighbourhood of the null point. Apparently the influence exerted by the isolated outliers has been compensated by the rest of the series. This reinforces the implication of the above influence diagnostics results.

Figure 6.12

$Q(\omega)$ in direction Q' for chemical process series.



6.3 Discussion

We have addressed the local sensitivity of the portmanteau statistic through the tangent plane of the perturbation surface $Q(\omega)$ at the null point. An evaluation of the normal curvature of $Q(\omega)$ in the manner of Wu and Luo (1993a) appears logically to be the next step of analysis. However, due to the complexity of $Q^{(2)}$, application of the second order curvature approach is not considered in our investigation.

CHAPTER 7

CONCLUSIONS

7. CONCLUSIONS

7.1 Concluding remarks

In this thesis, we have investigated some applications of the local influence approach to influence assessment on the topics of two-way contingency tables, covariate transformations and parametric link functions in generalized linear models, and Portmanteau statistic in time series. Several diagnostic procedures have been formulated based on the local influence approach. Through practical examples, the proposed diagnostics were all found to be useful and informative.

In two-way contingency tables, we examined the local influence approach for identification of outlying cells by perturbing the Pearson goodness-of-fit and the Likelihood Ratio Goodness-of-fit statistics. It was found that diagnostics derived from the Likelihood Ratio are generally less sensitive. Based on the Pearson Goodness-of-fit statistic, the local influence approach for identification of multiple outliers was preferable to residual-based methods as it is less susceptible to masking and swamping. The simulation study in Chapter 3 further confirmed this. Apart from providing a more reliable set of candidate outliers, an added advantage of the local influence approach is that it does not rely on any explicit formulation of a parametric model which is often required by other diagnostic methods. The local influence outlier detection method was shown to work satisfactorily in conjunction with the confirmatory procedure proposed in Chapter 3.

In Chapter 4, we addressed the application of local influence for assessing the sensitivity of the maximum likelihood estimate of the covariate transformation parameters in generalized linear models. Separate approaches based on analysis of the (i) transformed parameter surface or profile likelihood displacement and (ii) the partial influence, were found to be effective. The two approaches often provide similar diagnostic information but at times offer additional information

that supplement each other, (see snow geese example in Chapter 4). The various perturbation schemes enable us to assess whether sensitivity of the analysis was due to minor modifications in the covariate, response and/or case weight.

For sensitivity of the parametric link functions in generalized linear models, we studied approaches based on analysis of the link parameter surface, profile likelihood displacement and partial influence in Chapter 5. It was shown that such approaches are useful and can identify jointly influential observations on the link even when masking is present. Moreover, the influence diagnostics can assist in the revision of the link function and hence the form of the model.

Application of the local influence approach to assess sensitivity of the Portmanteau statistic in time series was investigated in Chapter 6. Variants of the Portmanteau statistic such as the Ljung-Box-Pierce statistic were not considered since the modifications can be regarded as nuisance parameters in the local influence analysis. The approach we took were through the tangent plane of the perturbation surface at the null point. The three examples presented in Chapter 6 have illustrated that this approach can yield valuable information on the sensitivity of the Portmanteau statistic. An examination of the actual displacement of the Portmanteau statistic due to local perturbations can also provide additional insight into the relationship between global and local influences. The application of the second order curvature approach was not considered due to computational complexity of the resulting diagnostics.

7.2 Topics for future research

In this thesis, we have presented some applications of the local influence approach. In areas where deletion methods suffer from problems such as masking and swamping, the local influence approach often appears to be less susceptible. This is evident from the four topics that we have explored. A brief account of relevant topics requiring further work are outlined below.

In the investigation of contingency tables in Chapter 3, we have restricted the application of local influence diagnostics to goodness-of-fit of statistical independence in two-way contingency tables. It would be useful to extend the approach to multi-way tables and uniform association model for doubly-ordered two-way tables (Goodman (1979), Agresti (1990)). If analytical solutions are not possible for such models, one may consider computational approaches to computing derivatives of the perturbation-formed surface. It may be possible to obtain analytical solutions for a few of the simpler models for three-way tables.

In covariate transformations, we have considered the linear predictor of the following transformation model,

$$\boldsymbol{\eta} = \boldsymbol{x}\boldsymbol{\delta} + G(\boldsymbol{z}, \boldsymbol{\lambda})\boldsymbol{\xi} \ ,$$

where we transform one or more independent covariates $\boldsymbol{z}$ of X.

$$X = \langle \boldsymbol{x}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}_{(1)}, \cdots, \boldsymbol{x}_{(p)}, \boldsymbol{z}_{(1)}, \cdots, \boldsymbol{z}_{(q)} \rangle.$$

The $n \times q$ matrix $G(\boldsymbol{z}, \boldsymbol{\lambda}) = \langle g_1(\boldsymbol{z}_{(1)}, \lambda_1), \cdots, g_q(\boldsymbol{z}_{(q)}, \lambda_q) \rangle$, where $\lambda_1, \cdots, \lambda_q$ are scalars, was structured in a way that each covariate to be transformed has only a single transformation parameter. For example, the covariate transformation model in the tree data example:

$$y_i = \delta + \left(\frac{x_{i1}^{\lambda_1} - 1}{\lambda_1} \right) \xi_1 + \left(\frac{x_{i2}^{\lambda_2} - 1}{\lambda_2} \right) \xi_2 + \varepsilon_i \ .$$

However, it is also possible to have transformations where $\lambda_1, \cdots, \lambda_q$ are vectors. ie.,

$$G(\boldsymbol{z}, \boldsymbol{\lambda}) = \langle g_1(\boldsymbol{z}_{(1)}, \lambda_{11}, \cdots, \lambda_{1k}), \cdots, g_q(\boldsymbol{z}_{(q)}, \lambda_{q1}, \cdots, \lambda_{qk}) \rangle.$$

For example,

$$y_i = \delta + \left(\frac{x_{i1}^{\lambda_{11}} - 1}{\lambda_{12}} \right) \xi_1 + \left(\frac{x_{i2}^{\lambda_{21}} - 1}{\lambda_{22}} \right) \xi_2 + \varepsilon_i \ .$$

The formulation of diagnostics for such models would be complex but feasible for future investigation. Similar generalizations also are applicable to the link parameters in generalized linear models.

The application of second-order local influence was not considered in our analysis of sensitivity of Portmanteau statistics in time series due to the computational complexity of the resulting diagnostics. It may be worthwhile to investigate the feasibility of obtaining the normal curvature for the perturbation-formed Portmanteau statistic surface. The effectiveness of the proposed diagnostics in assessing sensitivity of the Portmanteau statistic was only investigated by way of three numerical examples. The preliminary findings may need to be confirmed by simulations. Possible features to be considered include different types of ARMA models, magnitude and direction of contamination and different additive outlier patterns (consecutive, reallocation, isolated, etc). Sensitivity of the Portmanteau statistic to innovational outliers could also be explored in future research.

APPENDIX

Appendix: List of data sets

A1. Snow geese data- *Weisberg (1985)*

To estimate the number of snow geese in their summer range areas west of Hudson Bay in Canada, small aircraft fly over the range and, when a flock geese is spotted, an experienced person estimates the number of geese in the flock. To investigate the reliability of this method, an experiment was conducted in which an airplane carrying two observers flew over $n = 45$ flocks, and each observer made an independent estimate of the number of birds in each flock. A photograph of the flock was taken so that an exact count of the number of birds in the flock could be made. The resulting data are given in Table A.1.

Table A.1

Snow geese data

Photo	Observer 1	Observer 2	Photo	Observer 1	Observer 2
56	50	40	119	75	200
38	25	30	165	100	200
25	30	40	152	150	150
48	35	45	205	120	200
38	25	30	409	250	300
22	20	20	342	500	500
22	12	20	200	200	300
42	34	35	73	50	40
34	20	30	123	75	80
14	10	12	150	150	120
30	25	30	70	50	60
9	10	10	90	60	100
18	15	18	110	75	120
25	20	30	95	150	150
62	40	50	57	40	40
26	30	20	43	25	35
88	75	120	55	100	110
56	35	60	325	200	400
11	9	10	114	60	120
66	55	80	83	40	40
42	30	35	91	35	60
30	25	30	56	20	40
90	40	120			

A2. Erythrocyte sedimentation rate data - *Collett (1991)*

The set of data from Collett (1991) which relates the chronic disease state (0 = healthy; 1 = unhealthy) of 32 individuals, judged from the erythrocyte sedimentation rate (ESR) reading, to their plasma fibrinogen level (in gm/ℓ). The data are given in Table A.2.

Table A.2

ESR data

Case	WBC	Survival time
1	0	2.52
2	0	2.56
3	0	2.19
4	0	2.18
5	0	3.41
6	0	2.46
7	0	3.22
8	0	2.21
9	0	3.15
10	0	2.60
11	0	2.29
12	0	2.35
13	1	5.06
14	1	3.34
15	1	2.38
16	0	3.15
17	1	3.53
18	0	2.68
19	0	2.60
20	0	2.23
21	0	2.88
22	0	2.65
23	1	2.09
24	0	2.28
25	0	2.67
26	0	2.29
27	0	2.15
28	0	2.54
29	1	3.93
30	0	3.34
31	0	2.99
32	0	3.32

A3. Tree data - *Ryan et al (1976)*

The data, given in Table A.3 consist of measure on the volume, height and diameter at 4.5 feet above ground level for a sample of 31 black cherry trees in the Allegheny National Forest, Pennsylvannia. The Data were collected to provide a basis for determining an easy way of estimating the volume of a tree using its height and diameter.

Table A.3

Tree data

Diameter	Height	Volume	Diameter	Height	Volume
8.3	70	10.3	12.9	85	33.8
8.6	65	10.3	13.3	86	27.4
8.8	63	10.2	13.7	71	25.7
10.5	72	16.4	13.8	64	24.9
10.7	81	18.8	14.0	78	34.5
10.8	83	19.7	14.2	80	31.7
11.0	66	15.6	14.5	74	36.3
11.0	75	18.2	16.0	72	38.3
11.1	80	22.6	16.3	77	42.6
11.2	75	19.9	17.3	81	55.4
11.3	79	24.2	17.5	82	55.7
11.4	76	21.0	17.9	80	58.3
11.4	76	21.4	18.0	80	51.5
11.7	69	21.3	18.0	80	51.0
12.0	75	19.1	20.6	87	77.0
12.9	74	22.2			

A4. Leukemia data - *Cook and Weisberg (1982)*

Leukemia is a type of cancer characterized by an excess of white blood cells. At diagnosis, the count of white blood cells provides a useful measure of the patient's initial condition, more severe conditions being reflected by higher counts. The survival times in weeks and the white blood cell counts (WBC) for a sample of 17 patients classified as AG positive are given in Table A. 4.

Table A.4

Leukemia data

Case	WBC	Survival time
1	2300	65
2	750	156
3	4300	100
4	2600	134
5	6000	16
6	10500	108
7	10000	121
8	17000	4
9	5400	39
10	7000	143
11	9400	56
12	32000	26
13	35000	22
14	100000	1
15	100000	1
16	52000	5
17	100000	65

A5. Simulated series

An artificial AR(1) series of $n = 100$ observations is generated using `S-PLUS` function `arima.sim` with $\phi = 0.5$ and innovation variance $\sigma^2 = 1$. Two consecutive additive outliers are created by adding 3 to the 49th and 50th observations. The contaminated series is given in Table A.5.

Table A.5

Simulated series

Obs		Obs		Obs	
1	0.039	41	1.700	81	-0.832
2	0.344	42	0.610	82	-1.728
3	0.117	43	0.105	83	0.896
4	-1.385	44	-1.841	84	0.899
5	-1.573	45	-0.458	85	-0.477
6	1.104	46	-0.728	86	0.819
7	0.857	47	-0.829	87	0.518
8	-0.062	48	0.295	88	0.238
9	0.842	49	4.520	89	-1.202
10	2.329	50	3.681	90	-1.890
11	2.343	51	0.144	91	-0.585
12	1.059	52	0.735	92	-0.914
13	0.111	53	1.337	93	-1.174
14	0.254	54	-0.466	94	-1.544
15	-0.841	55	-0.152	95	-1.142
16	-0.641	56	-1.938	96	-1.016
17	0.511	57	-2.249	97	0.148
18	0.442	58	-1.307	98	0.679
19	0.489	59	-0.669	99	1.395
20	-1.437	60	-1.150	100	0.140
21	-0.975	61	1.002		
22	0.127	62	-0.198		
23	-0.444	63	1.482		
24	-1.169	64	0.153		
25	-0.312	65	-0.226		
26	0.118	66	0.204		
27	0.534	67	-1.68		
28	1.400	68	-1.452		
29	1.214	69	-0.266		
30	0.099	70	-0.158		
31	-0.946	71	-0.475		
32	-1.961	72	1.420		
33	-2.388	73	-0.449		
34	-0.686	74	-0.539		
35	0.664	75	-0.538		
36	0.449	76	1.569		
37	-0.354	77	1.480		
38	-1.751	78	1.112		
39	-0.737	79	1.432		
40	0.721	80	-0.330		

A6. Retail sales of automobile dealers series - *Wu et al (1993)*

Table A.6 contains 65 monthly observations (January 1985 to May 1990) from the estimated retail sales of automotive dealers (seasonally adjusted) published in the *Survey of Current Business Statistics*, US Department of Commerce.

Table A.6

Retail sales of automobile dealers series

Obs	Sales billions of $	Obs	Sales billions of $	Obs	Sales billions of $
1	24.00	26	26.89	51	30.91
2	24.46	27	27.99	52	31.81
3	24.76	28	28.09	53	31.56
4	25.45	29	28.00	54	31.33
5	25.51	30	29.28	55	31.94
6	25.19	31	29.55	56	32.78
7	25.31	32	31.04	57	32.99
8	25.89	33	30.05	58	31.32
9	28.35	34	29.29	59	31.11
10	24.36	35	28.99	60	30.43
11	24.71	36	29.72	61	33.27
12	25.34	37	30.34	62	31.41
13	25.54	38	30.84	63	31.48
14	25.21	39	31.36	64	30.97
15	24.24	40	30.74	65	30.78
16	25.43	41	30.72		
17	26.44	42	30.97		
18	26.23	43	31.05		
19	26.10	44	30.13		
20	27.14	45	29.53		
21	34.15	46	31.49		
22	27.51	47	31.89		
23	27.14	48	32.05		
24	31.91	49	31.36		
25	23.37	50	30.84		

A7. Chemical process concentration series - *Box et al (1994)*

Table A.7 contains 197 readings of concentration in a chemical process observed every two hours.

Table A.7

Chemical process concentration readings series

Obs	Concentration	Obs	Concentration	Obs	Concentration
1	17.0	71	17.3	141	16.6
2	16.6	72	17.4	142	16.5
3	16.3	73	17.7	143	17.0
4	16.1	74	16.8	144	16.7
5	17.1	75	16.9	145	16.7
6	16.9	76	17.0	146	16.9
7	16.8	77	16.9	147	17.4
8	17.4	78	17.0	148	17.1
9	17.1	79	16.6	149	17.0
10	17.0	80	16.7	150	16.8
11	16.7	81	16.8	151	17.2
12	17.4	82	16.7	152	17.2
13	17.2	83	16.4	153	17.4
14	17.4	84	16.5	154	17.2
15	17.4	85	16.4	155	16.9
16	17.0	86	16.6	156	16.8
17	17.3	87	16.5	157	17.0
18	17.2	88	16.7	158	17.4
19	17.4	89	16.4	159	17.2
20	16.8	90	16.4	160	17.2
21	17.1	91	16.2	161	17.1
22	17.4	92	16.4	162	17.1
23	17.4	93	16.3	163	17.1
24	17.5	94	16.4	164	17.4
25	17.4	95	17.0	165	17.2
26	17.6	96	16.9	166	16.9
27	17.4	97	17.1	167	16.9
28	17.3	98	17.1	168	17.0
29	17.0	99	16.7	169	16.7
30	17.8	100	16.9	170	16.9
31	17.5	101	16.5	171	17.3
32	18.1	102	17.2	172	17.8
33	17.5	103	16.4	173	17.8

Obs	Concentration	Obs	Concentration	Obs	Concentration
34	17.4	104	17.0	174	17.6
35	17.4	105	17.0	175	17.5
36	17.1	106	16.7	176	17.0
37	17.6	107	16.2	177	16.9
38	17.7	108	16.6	178	17.1
39	17.4	109	16.9	179	17.2
40	17.8	110	16.5	180	17.4
41	17.6	111	16.6	181	17.5
42	17.5	112	16.6	182	17.9
43	16.5	113	17.0	183	17.0
44	17.8	114	17.1	184	17.0
45	17.3	115	17.1	185	17.0
46	17.3	116	16.7	186	17.2
47	17.1	117	16.8	187	17.3
48	17.4	118	16.3	188	17.4
49	16.9	119	16.6	189	17.4
50	17.3	120	16.8	190	17.0
51	17.6	121	16.9	191	18.0
52	16.9	122	17.1	192	18.2
53	16.7	123	16.8	193	17.6
54	16.8	124	17.0	194	17.8
55	16.8	125	17.2	195	17.7
56	17.2	126	17.3	196	17.2
57	16.8	127	17.2	197	17.4
58	17.6	128	17.3		
59	17.2	129	17.2		
60	16.6	130	17.2		
61	17.1	131	17.5		
62	16.9	132	16.9		
63	16.6	133	16.9		
64	18.0	134	16.9		
65	17.2	135	17.0		
66	17.3	136	16.5		
67	17.0	137	16.7		
68	16.9	138	16.8		
69	17.3	139	16.7		
70	16.8	140	16.7		

BIBLIOGRAPHY

Bibliography

Agresti, A. (1990). *Categorical data analysis*, New Yoak: Wiley.

Andersen, E. B. (1992). Diagnostics in categorical data analysis, *J. R. Statist. Soc. B*, **54**, 781-791.

Aranda-Ordaz, F. J. (1981). On two families of transformations to additivity for binary response data, *Biometrika*, **68**, 357-363.

Atkinson, A. C. (1986). Diagnostic tests for transformations, *Technometrics*, **28**, 29-38.

Atkinson, A. C. (1988). Transformation unmasked, *Technometrics*, **30**, 311-318.

Barnard, G. A. (1980). Discussion of Professor Box's paper, *J. R. Statist. Soc. A*, **143**, 404-406.

Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data*, 3rd edn. Chichester: Wiley.

Basu, S. and Reinsel, G.C. (1996). Relationship between missing data likelihoods and complete data restricted likelihoods for regression time series models: an application to total ozone data, *Applied Statistics*, **45**, 63-72.

Bell, W., and Hillmer, S. (1987). *Time series analysis, forecasting and control.* Englewood Cliffs: Prentice Hall

Box, G. E. P. and Tidwell, P. W. (1962). Transformation of the independent variables, *Technometrics*, **4**, 531-550.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). *Initializing the Kalman filter in the non-stationary case.* Research report: CENSUS/SRC/RR-87/33, Statistical Research Division, Bureau of the Census, Washington, DC, 20233.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive integrated moving average time series models, *J. Am. Statist. Ass.*, **65**, 1509-1526

Bradu, D. and Hawkins, D. M. (1982). Location of multiple outliers in two-way tables, using tetrads, *Technometrics*, **24**, 103-108.

Brown, M. B. (1974). Identification of the sources of significance in two-way contingency tables, *Appl. Statist.*, **23**, 405-431.

Bruce, A. G., and Martin, R. D. (1989). Leave-k-out diagnostics for time series (with discussion), *J. R. Statist. Soc. B*, **51**, 363-424.

Casjens, L. (1974). *The prehistoric human ecology of Southern Ruby valley, Nevada.* Doctoral dissertation, Harvard University.

Chan, W.S. (1994). On portmanteau goodness-of-fit tests in robust time series modeling, *Computational Statistics*, **9**, 301-310.

Chen, C. and Liu, L. (1993). Forecasting time series with outliers, *J. Forecasting*, **12**, 13-35.

Chatterjee, S. and Hadi, A. S. (1988). *Sensitivity analysis in linear regression*, Wiley, New York.

Collett, D. (1991). *Modeling binary data*, Chapman and Hall, New York.

Cook, R. D. (1986). Assessment of local influence (with discussion), *J. R. Statist. Soc. B*, **48**, 133-169.

Cook, R. D. (1987). Influence assessment, *Journal of Applied Statistics*, **14**, 117-131.

Cook, R. D. and Wang, P. C. (1983). Transformations and influential cases in regression, *Technometrics*, **25**, 337-343.

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

Davies, L. and Gather, U. (1993). The identification of multiple outliers, *J. Am. Statist. Ass.*, **88**, 782-801.

Ezekiel, M. and Fox, F. A. (1959). *Methods of correlation and regression analysis*, Wiley, New York.

Fox, A. J. (1972). Outliers in time series, *J. R. Statist. Soc. B*, **34**, 350-363.

Fuchs, C. and Kenett, R. (1980). A test for detecting outlying cells in the multinomial distribution and two-way contingency tables, *J. Am. Statist. Ass.*, **75**, 395-398.

García-Heras, J., Muñoz-García, J. and Pascual-Acosta, A. (1993). Criteria for the detection of outliers in the multinomial model, *J. Appl. Statist.*, **20**, 137-142.

Goodman, L. A. (1981). Association models and canonical correlation in the analysis of cross-classifications having ordered categories, *J. Am. Statist. Ass.*, **76**, 320-334.

Haberman, S. J. (1973). The analysis of residuals in cross-classified tables, *Biometrics*, **29**, 205-220.

Hinkley, D. V. and Wang, S. (1988). More about transformations and influential cases in regression, *Technometrics*, **30**, 435- 440.

Inclán, C. (1992). Effect of a change in variance on the portmanteau statistic, *Estadistica***44**, 149-170.

Jiang, J., Hui, Y.V. and Zheng, Z. (1999). Robust goodness-of-fit tests for AR(p) models based on L_1-norm fitting, *Science in China, Ser. A*, **42**, (2).

Kaiser, M. S. (1997). Maximum likelihood estimation of link function parameters, *Comput. Stat. Data Anal.*, **24**, 79-87.

Kohn, R. and Ansley, C. F. (1985). Efficient estimation and prediction in time series regression models, *Biometrika.*, **72**, 694-697.

Kohn, R. and Ansley, C. F. (1986). Estimation, prediction, and interpolation for ARIMA models with missing data, *J. Am. Statist. Ass.*, **81**, 751-761.

Kotze, T. J. and Hawkins, D. M. (1984). The identification of outliers in two-way contingency tables using 2×2 subtables, *Appl. Statist.*, **33**, 215-223.

Lawrance, A. J. (1988). Regression transformation diagnostics using local influence, *J. Am. Statist. Ass.*, **83**, 1067-1072.

Lawrance, A. J. (1995). Deletion influence and masking in regression, *J. R. Statist. Soc. B*, **57**, 181-189.

Ledolter, J. (1989). The effect of additive outliers on the forecasts from ARIMA models, *International Journal of Forecasting*, **5**, 231-240.

Ledolter, J. (1990), Outlier diagnostics in time series analysis influence, *Journal of time series analysis*, **11**, 317-324.

Lee, A. H. (1988). Assessing partial influence in generalized linear models, *Biometrics*, **44**, 71-77.

Lee, A. H. and Fung, W. K. (1997). Confirmation of multiple outliers in generalized linear and nonlinear regressions, *Comput. Statist. Data Anal.*, **25**, 55-65.

Li, W.K. (1988). A goodness-of-fit test in robust time series modeling, *Biometrikia*, **75**, 355-361.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models, *Biometrika*, **65**, 297-303.

Ljung, G.M. (1993). On outlier detection in time series, *Journal of the Royal Statistical Society, Ser. B*, **55**, 559-567.

Luceño, A. (1998). Detecting possibly non-consecutive outliers in industrial time series, *Journal of the Royal Statistical Society, Series B*, **60**, 295-310.

McCullagh P. and Nelder J. A. (1989). *Generalized Linear Models.* Chapman and Hall, New York.

Mosteller, F. and Parunak, A. (1985). Identifying extreme cells in a sizable contingency table: probabilistic and exploratory approaches. In *Exploring Data Tables, Trends and Shapes* (eds: D. C. Hoaglin, F. Mosteller, and J. W. Tukey) pp. 189-224. New York: Wiley.

Muñoz-García, J., Moreno-Rebollo, J. L. and Pascual-Acosta, A. (1987). Detecting outliers in s-multinomial distributions through adjusted residuals, *J. Appl. Statist.*, **14**, 171-176.

Perli, H. G., Hommel, G. and Lehmacher, W. (1985). Sequentially rejective test procedures for detecting outlying cells in one- and two-sample multinomial experiments, *Biometrical J.*, **27**, 885-893.

Pregibon, D. (1980). Goodness of link tests for generalized linear models, *Appl*

Statist, **29**, 15-24.

Pregibon, D. (1985). Link test, in *Encyclopedia of Statistical Sciences.* Wiley, **5**, 82-85.

Prentice, R. L. (1976). A generalization of the probit and logit methods for dose-response curves, *Biometrics*, **32**, 761-768.

Ryan, T. A., Joiner, B. L., and Ryan, B. F. (1976). *Minitab Student Handbook*, Duxbury Press, MA.

Scallan, A., Gilchrist, R. and Green, M. (1984). Fitting parametric link functions in generalized linear models, *Comput. Stat. Data Anal.*, **2**, 37-49.

Shih, J. Q. (1993). Regression transformation diagnostics in transform-both-sides model, *Statistics & Probability Letters*, **16**, 411-420.

Shih, J. Q. and Wei, B. C. (1995). Diagnostics for nonlinear regression model with weighting or transformation, *Systems Science and Mathematical Science*, **8**, 240-248.

Simonoff, J. S. (1988). Detecting outlying cells in two-way contingency tables via backwards stepping, *Technometrics*, **30**, 339-345.

Thomas, W. and Cook, R. D. (1989). Assessing influence on regression coefficients in generalized linear models, *Biometrica*, **76**, 741-749.

Thomas, W. and Cook, R. D. (1990). Assessing influence on predictions from generalized models, *Technometrics*, **32**, 59-65.

Tsai, C.-L. and Wu, X. (1990). Diagnostics in transformation and weighted regression, *Technometrics*, **32**, 315-322.

Tsai, C.-L. and Wu, X. (1992). Transformation-model diagnostics, *Technometrics*, **34**, 197-202.

Wang, P. C. (1987). Simultaneous transformations and influence in regression, *Communications in Statistics - Theory & Methods*, **16**, 3417-3425.

Wei, B. C. and Hickernell, F. J. (1996). Regression transformation diagnostics for explanatory variables, *Statistica Sinica*, **6**, 433-454.

Wei, B. C. and Shih, J. Q. (1994a). On statistical models for regression diagnostics, *Annals of the Institute of Statistical Mathematics*, **46**, 267-278.

Wei, B. C. and Shih, J. Q. (1994b). Local influence analysis for regression transformation models, *ACTA Mathematicae Applicatae Sinica*, **17**, 132-143.

Weisberg, S. (1985). *Applied Linear Regression*, 2nd edition, Wiley, New York.

Wu, L. S.-Y., Hosking J. R. M. and Ravishanker, N. (1993). Reallocation outliers in time series, *Applied Statistics*, **42**, 301-313.

Wu, X. and Luo, Z. (1993a). Second-order approach to local influence, *J. R. Statist. Soc. B*, **55**, 929-936.

Wu, X. and Luo, Z. (1993b). Residual sum of squares and multiple potential, diagnostics by a second order local approach, *Statist. Probab. Lett.*, **16**, 289-296.

Wu, X. and Wan, F. (1994). A perturbation scheme for nonlinear models, *Statistics & Probability Letters*, **20**, 197-202.