Effect of Correlations and Unequal Variances in Testing for Outliers in Linear Regression

Scand J Statist 19: 91–95, 1992

Effect of Correlations and Unequal Variances in Testing for Outliers in Linear Regression

JERZY K. BAKSALARY

Tadeusz Kotarbiński Pedagogical University, Zielona Góra, Poland, and University of Tampere, Finland

MARKKU NURHONEN and SIMO PUNTANEN

University of Tampere, Finland

ABSTRACT. It is well known that testing for outliers in linear regression can be based on the externally Studentized residuals. In this paper we study the labeled version of the mean-shift outlier model, and consider the multiple analogue of the externally Studentized residual. We characterize the robustness of this test against the effect of correlations and unequal variances in the dispersion matrix of the model. The general result is illustrated by applying it to the simple sample, by discussing the class of matrices V having the so-called Baldessari's structure, and by considering a specific analysis-of-covariance model.

Key words: mean-shift model, robustness, dispersion matrix, *F*-test, Studentized residual, Baldessari's structure

1. Introduction

Let the symbol

$$\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V}\} \tag{1}$$

denote the linear regression model, where $\mathbf{y}$ is an $n \times 1$ observable random vector following multinormal distribution with expectation vector $\mathbf{X}\boldsymbol{\beta}$ and dispersion matrix $\sigma^2\mathbf{V}$. We assume that an $n \times p$ matrix $\mathbf{X}$ and an $n \times n$ positive definite matrix $\mathbf{V}$ are known, while a $p \times 1$ vector $\boldsymbol{\beta}$ and a positive scalar σ^2 are unknown parameters.

For the problem of testing for outliers in model (1), we consider the *labeled version of the mean-shift model* (cf. Beckman & Cook, 1983, p. 132; Cook & Weisberg, 1982, p. 28), which is a modification of (1) to the form

$$\{\mathbf{y}, \mathbf{Z}\gamma, \sigma^2\mathbf{V}\}, \tag{2}$$

with

$$\mathbf{Z} = (\mathbf{X} : \mathbf{D}) \quad \text{and} \quad \gamma = (\boldsymbol{\beta}' : \boldsymbol{\delta}')'. \tag{3}$$

Here $\mathbf{X} = (\mathbf{X}_1' : \mathbf{X}_2')'$ and $\mathbf{D} = (\mathbf{0} : \mathbf{I}_m)'$, where $\mathbf{X}_1'$ and $\mathbf{X}_2'$ are the transposes of $(n-m) \times p$ and $m \times p$ submatrices of $\mathbf{X}$ and $\mathbf{I}_m$ stands for the identity matrix of order m. Thus if $\mathbf{y}$ is partitioned as $\mathbf{y} = (\mathbf{y}_1' : \mathbf{y}_2')'$, then the components of the $m \times 1$ subvector $\mathbf{y}_2$ are suspected to be outliers in the sense that their expectations are shifted according to the unknown vector $\boldsymbol{\delta}$.

Under the labeled version of the mean-shift model, the problem of detecting alleged outliers in the subvector $\mathbf{y}_2$ reduces to applying the usual F-test for $H_0 : \boldsymbol{\delta} = \mathbf{0}$ vs. $H_a : \boldsymbol{\delta} \neq \mathbf{0}$. In this paper, we establish a robustness property of such a procedure by characterizing all those matrices $\mathbf{V}$ for which the test statistic and the corresponding F-test in the model (2) are the same as in the particular case of (2) when $\mathbf{V} = \mathbf{I}_n$. We illustrate the general result by applying it to the simple sample, by discussing the class of matrices $\mathbf{V}$ having the so-called Baldessari's structure, and by considering a specific analysis-of-covariance model.

Abstract

It is well known that testing for outliers in linear regression can be based on the externally Studentized residuals. In this paper we study the labeled version of the mean-shift outlier model, and consider the multiple analogue of the externally Studentized residual. We characterize the robustness of this test against the effect of correlations and unequal variances in the dispersion matrix of the model. The general result is illustrated by applying it to the simple sample, by discussing the class of matrices V having the so-called Baldessari's structure, and by considering a specific analysis-of-covariance model.