The Korean Communications
in Statistics Vol. 10 No. 2, 2003
pp. 519-524

Detection of Outliers in Constrained Regression

Myung Geun Kim[1)]

Abstract

We suggest a method of identifying outliers, using local influence, in regression when linear constraints are imposed on the regression coefficients. An example is given for illustration.

Keywords: Linear constraints, local influence, regression, outliers.

1. Introduction

Tests of linear hypotheses are used for uncovering some relationships among regression coefficients. When these tests reveal possible linear relationships among regression coefficients, methods of detecting outliers appropriate for this situation are necessary. However, few methods of detecting outliers in regression with linear constraints seem to be avaliable while many outlier detection methods in unconstrained regression have been known and some of them can be found in Cook and Weisberg (1982), Chatterjee and Hadi (1988), Barnett and Lewis (1994).

In this work we will suggest a method of detecting outliers, using local influence, in linear regression when there are some linear relationships among regression coefficients. The local influence method introduced by Cook (1986) is a general method of assessing the influence of local departures from assumptions for the underlying model based on the likelihood displacement. It measures the sensitivity of the analysis to a change in the model caused by a minor perturbation and has been known as a method of detecting outliers that avoids masking and swamping effects. A recent review paper by Rancel and Sierra (2001) provides some literature about the local influence method in regression. We review some results about regression in Section 2. In Section 3 computations necessary for the local influence measure are provided. An illustrative example is provided in Section 4.

1) Professor, Department of Applied Statistics, Seowon University, Cheongju, 361-742.
E-mail: mgkim@seowon.ac.kr.

2. Preliminaries

We consider the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}$ is an n by 1 vector of response variables, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$ is an n by p matrix of fixed independent variables, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{p-1})^T$ is a p by 1 vector of unknown regression coefficients and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$ is an n by 1 vector of unobservable errors. We assume that the ε_r (r = 1, ..., n) are independent and identically distributed as a normal distribution $N(0, \sigma^2)$ with mean zero and unknown variance σ^2.

First, we consider the case where there is no constraint on $\boldsymbol{\beta}$. In this case the maximum likelihood estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}_* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ and the residual vector is given by $\boldsymbol{e}_* = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_*$. The maximum likelihood estimator of σ^2 is then given by $\hat{\sigma}^2_* = \boldsymbol{e}_*^T\boldsymbol{e}_*/n$.

Next, the linear relationships among $\boldsymbol{\beta}$ can be expressed as

$$\boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{c} \tag{1}$$

where $\boldsymbol{A}$ is a specified q by p ($q \le p$) matrix of rank q and $\boldsymbol{c}$ is a known q by 1 vector. Under the linear constraints (1), the maximum likelihood estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_* - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T[\boldsymbol{A}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T]^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}}_* - \boldsymbol{c})$$

and the maximum likelihood estimator of σ^2 is $\hat{\sigma}^2 = \boldsymbol{e}^T\boldsymbol{e}/n$, where $\boldsymbol{e} = (e_1, \ldots, e_n)^T = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$. The existence of the linear relationships among $\boldsymbol{\beta}$ can be checked by using the usual F-statistic given by $F = [(n-p)/q](\hat{\sigma}^2/\hat{\sigma}^2_* - 1)$ which is distributed as the F-distribution with q and n-p degrees of freedom when the linear relationships (1) hold. More details can be found in Chap. 4 of Seber (1977).

3. Local Influence Procedure

In this section we will derive local influence measures for the constrained regression model with linear constraints (1).

Let $\boldsymbol{w} = (w_1, \ldots, w_n)^T$ be an n by 1 vector of perturbations. We consider the perturbed model in which the r-th error ε_r is perturbed according to a normal distribution $N(0, \sigma^2/w_r)$ for r = 1,...,n. This perturbation scheme simultaneously perturbs all the cases. When the ε_r

are set equal to one, the perturbed model reduces to the unperturbed model.

Let $\boldsymbol{\theta}$ be the p+1 by 1 vector of parameters formed by stacking $\boldsymbol{\beta}$ and σ^2. We denote the log-likelihoods for the unperturbed and perturbed models by $L(\boldsymbol{\theta})$ and $L(\boldsymbol{\theta}|\boldsymbol{w})$, respectively. The likelihood displacement $LD(\boldsymbol{w})$ is defined by $2[L(\hat{\boldsymbol{\theta}})-L(\hat{\boldsymbol{\theta}}_{\boldsymbol{w}})]$, where $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}$ are the maximum likelihood estimators of $\boldsymbol{\theta}$ under the unperturbed and perturbed models, respectively. The surface of interest is formed by the n+1 by 1 vector of the values $\boldsymbol{w}$ and $LD(\boldsymbol{w})$ as $\boldsymbol{w}$ varies along a direction $\boldsymbol{d}$ with $\boldsymbol{w}=\boldsymbol{w}_0+a\,\boldsymbol{d}$, where $\boldsymbol{w}_0$ is the null point at which $L(\boldsymbol{\theta})=L(\boldsymbol{\theta}|\boldsymbol{w}_0)$, $\boldsymbol{d}$ is a direction vector of unit length and the scalar a measures the magnitude of the perturbation. The direction vector $\boldsymbol{d}_{\max}$ associated with the largest curvature of the curve at $a=0$ provides information about outliers that cause a great change in the likelihood displacement. Observations corresponding to the component of $\boldsymbol{d}_{\max}$ that has substantially larger magnitude than the others are potential outliers.

Let $\dot{LD}=\partial LD(\boldsymbol{w})/\partial \boldsymbol{w}\,|_{a=0}$ and $\ddot{LD}=\partial^2 LD(\boldsymbol{w})/\partial \boldsymbol{w}\,\partial \boldsymbol{w}^T|_{a=0}$. For the constrained regression model, $dL(\boldsymbol{\theta})/d\boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}$ is not in general zero. However, it will be shown later that the first order derivative $\dot{LD}$ of the likelihood displacement evaluated at $a=0$ is zero. Thus the curvature is given by

$$C = |\,\boldsymbol{d}^T \ddot{LD}\,\boldsymbol{d}\,|.$$

Then $\boldsymbol{d}_{\max}$ is the eigenvector corresponding to the largest absolute eigenvalue of $\ddot{LD}$.

In what follows we will derive $\dot{LD}$ and $\ddot{LD}$. First we note that

$$\dot{LD} = -2\left(\frac{\partial \hat{\boldsymbol{\theta}}_{\boldsymbol{w}}}{\partial \boldsymbol{w}}\bigg|_{a=0}\right)\dot{\boldsymbol{L}} \tag{2}$$

$$\ddot{LD} = -2\left(\frac{\partial \hat{\boldsymbol{\theta}}_{\boldsymbol{w}}}{\partial \boldsymbol{w}}^T\bigg|_{a=0}\right)\ddot{\boldsymbol{L}}\left(\frac{\partial \hat{\boldsymbol{\theta}}_{\boldsymbol{w}}}{\partial \boldsymbol{w}^T}\bigg|_{a=0}\right)-2\sum_{i=1}^{p+1}\dot{L}_i\left(\frac{\partial^2 \hat{\theta}_{iw}}{\partial \boldsymbol{w}\partial \boldsymbol{w}^T}\bigg|_{a=0}\right), \tag{3}$$

where $\dot{\boldsymbol{L}}=\partial L(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}$, $\ddot{\boldsymbol{L}}=\partial^2 L(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T$ evaluated at $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}$, $\dot{L}_i$ is the i-th element of $\dot{\boldsymbol{L}}$ and $\hat{\theta}_{iw}$ is the i-th element of $\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}$.

Let $\hat{\boldsymbol{\beta}}_{*,\boldsymbol{w}} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y}$, where $\boldsymbol{W}$ is a diagonal matrix whose diagonal elements are those of $\boldsymbol{w}$. Under the constraints (1), the maximum likelihood estimator of $\boldsymbol{\beta}$ for the perturbed model is

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{w}} = \hat{\boldsymbol{\beta}}_{*,\boldsymbol{w}}-(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{A}^T[\boldsymbol{A}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{A}^T]^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}}_{*,\boldsymbol{w}}-\boldsymbol{c}).$$

Using the identity $\partial \boldsymbol{S}(\boldsymbol{w})^{-1}/\partial w_r=-\boldsymbol{S}(\boldsymbol{w})^{-1}[\partial \boldsymbol{S}(\boldsymbol{w})/\partial w_r]\boldsymbol{S}(\boldsymbol{w})^{-1}$ for a matrix

$S(w)$ being a function of w, we easily get

$$\frac{\partial \hat{\beta}_{*,w}}{\partial w_r}\Bigg|_{a=0} = (X^T X)^{-1} X^T D_r e_* \tag{4}$$

$$\frac{\partial^2 \hat{\beta}_{*,w}}{\partial w_r \partial w_s}\Bigg|_{a=0} = -(X^T X)^{-1} X^T D_r H D_s e_* - (X^T X)^{-1} X^T D_s H D_r e_* \tag{5}$$

where D_r is a diagonal matrix in which the i-th element is one and the other diagcnal elements are all zero, and $H = X(X^T X)^{-1} X^T$.

A little complicated algebra together with (4) and (5) shows that

$$\frac{\partial \hat{\beta}_w}{\partial w_r}\Bigg|_{a=0} = (I - Q)(X^T X)^{-1} X^T D_r e \tag{6}$$

$$\frac{\partial^2 \hat{\beta}_w}{\partial w_r \partial w_s}\Bigg|_{a=0} = -(I-Q)(X^T X)^{-1} X^T D_r (H-K) D_s e - (I-Q)(X^T X)^{-1} X^T D_s (H-K) D_r e \tag{7}$$

where $Q = (X^T X)^{-1} A^T [A(X^T X)^{-1} A^T]^{-1} A$ and $K = X(X^T X)^{-1} A^T [A(X^T X)^{-1} A^T]^{-1} A (X^T X)^{-1} X^T$. For the perturbed model, the maximum likelihood estimator of σ^2 is

$$\begin{aligned}\widehat{\sigma_w}^2 &= \frac{1}{n}(y - X\hat{\beta}_w)^T W (y - X\hat{\beta}_w) \\ &= \frac{1}{n}(y - X\hat{\beta}_{*,w})^T W (y - X\hat{\beta}_{*,w}) \\ &\quad + \frac{1}{n}(A\hat{\beta}_{*,w} - c)^T [A(X^T W X)^{-1} A^T]^{-1} (A\hat{\beta}_{*,w} - c)\end{aligned}$$

and using the previous results (6) and (7) with a little further computation we get

$$\frac{\partial \widehat{\sigma_w}^2}{\partial w_r}\Bigg|_{a=0} = \frac{1}{n} e^T D_r e \tag{8}$$

$$\frac{\partial^2 \widehat{\sigma_w}^2}{\partial w_r \partial w_s}\Bigg|_{a=0} = \frac{1}{n} e^T [D_r (K-H) D_s + D_s (K-H) D_r] e \tag{9}$$

The first order derivatives of the log-likelihood evaluated at $\theta = \hat{\theta}$ are easily obtained as

$$\frac{\partial L(\theta)}{\partial \beta}\Bigg|_{\theta=\hat{\theta}} = \frac{1}{\hat{\sigma}^2} A^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta}_* - c) \tag{10}$$

$$\frac{\partial L(\theta)}{\partial \sigma^2}\Bigg|_{\theta=\hat{\theta}} = 0. \tag{11}$$

Using (6), (8), (10) and (11) it follows from (2) that $\dot{LD} = 0$. In order to get $\ddot{LD}$ we further need to compute the second order derivatives of the log-likelihood eavluated at $\theta = \hat{\theta}$ that are given by

$$\left.\frac{\partial^2 L(\theta)}{\partial \beta \partial \beta^T}\right|_{\theta=\hat{\theta}} = -\frac{1}{\hat{\sigma}^2} X^T X$$

$$\left.\frac{\partial^2 L(\theta)}{\partial \beta \partial \sigma^2}\right|_{\theta=\hat{\theta}} = -\frac{1}{\hat{\sigma}^4} A^T[A(X^T X)^{-1} A^T]^{-1}(A\hat{\beta}_* - c)$$

$$\left.\frac{\partial^2 L(\theta)}{\partial (\sigma^2)^2}\right|_{\theta=\hat{\theta}} = -\frac{n}{2\hat{\sigma}^4}.$$

Using (7), (9), (10) and (11) the second term in (3) turns out to be zero and thus the (r,s)-th element of $\ddot{LD}$ is computed as

$$\ddot{LD}_{rs} = \frac{2e_r e_s}{\hat{\sigma}^2} x_r^T (X^T X)^{-1}(I-Q)^T X^T X(I-Q)(X^T X)^{-1} x_s + \frac{e_r^2 e_s^2}{n\hat{\sigma}^4}.$$

4. Example

We consider the chemical shipment data (Neter et al.. 1996, p. 253), taken on 20 incoming shipments of chemicals in drums arriving at a warehouse, in which the response variable y is the number of minutes required to handle shipment, the first independent variable x_1 is the number of drums in shipment and the second independent variable x_2 is the total weight of shipment in hundred pounds. When the unconstrained regression model $y = \beta_0 + \beta_1 x_1 + \beta x_2 + \varepsilon$ is fitted to the chemical shipment data, the p-value for the F-test of the linear hypothesis that $4\beta_1 - \beta_2 = 10$ becomes 0.998 and thus we can conclude that there exists a linear relationship $4\beta_1 - \beta_2 = 10$ at any reasonable significance level.

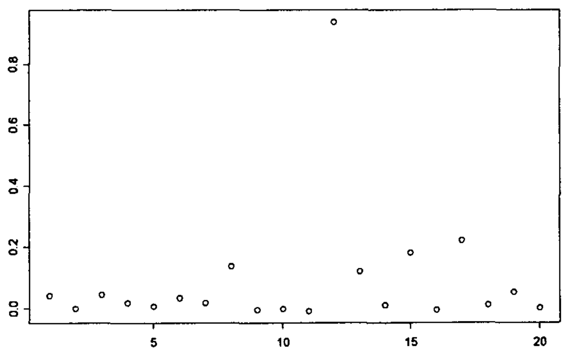

Figure 1. Index plot of the direction cosines in $\boldsymbol{d}_{\max}$

An index plot of the direction cosines in $\boldsymbol{d}_{\max}$ for the regression with $4\beta_1 - \beta_2 = 10$ is provided in Figure 1 from which we can see that case 12 is remarkably distinct from the other cases. The direction cosines for the other 19 cases except for case 12 are very small. Hence case 12 is an outlier for the regression with $4\beta_1 - \beta_2 = 10$. This result can be supported by investigating the e_i : case 12 has the largest absolute residual 15.2 and the other cases have small residuals relative to case 12.

5. Conclusions

A method of detecting outliers in constrained regressions has been suggested and it was applied to a numerical example. As compared with the results from residual analysis, the method detects outliers well and it will be a useful diagnostic tool.

References

[1] Barnett, V. and Lewis, T. (1994) *Outliers in Statistical Data,* 3rd ed., Wiley, New York.
[2] Chatterji, S. and Hadi, A.S. (1988) *Sensitivity Analysis in Linear Regression*, Wiley, New York.
[3] Cook, R.D. (1986) Assessment of local influence (with discussions), *Journal of the Royal Statistical Society, Series B,* **48**, 133-169.
[4] Cook, R.D. and Weisberg, S. (1982) *Residuals and Influence in Regression*, Chapman and Hall.
[5] Neter, J., Kutner, M.H., Nachtsheim, C.J. and Wasserman, W. (1996) *Applied Linear regression Models,* 3rd ed., Irwin.
[6] Rancel, M.M.S. and Sierra, M.A.G. (2001) Regression diagnostics using local influence : a review, *Communications in Statistics: Theory and Methods*, **30**, 799-813.
[7] Seber, G.A.F. (1977). *Linear Regression Analysis*, Wiley, New York.

[Received February 2003, Accepted May 2003]