비선형 모델에 있어서의 다양한 종류의 잔차들에 관한 연구
—The Different Types of Residuals in Nonlinear Regression Models—

姜 昶 旭*

요 지

본 논문은 비선형 회귀분석 모델(Nonlinear Regression Models)에서의 추산잔차(Recursive Residuals)를 정의하기 위한 것을 목적으로 한다.

선형 회귀분석 모델(Linear Regression Models)에서는 추산잔차가 우리가 측정할 수 없는 진짜 오차(True Error)와 같은 확률 분포를 갖는데 이의 평균은 0이고 분산은 σ^2이다.

그러나 비선형 회귀분석에서는 이와같은 정확한 분포를 알수가 없기 때문에, 여러 종류의 잔차들을 연구·검토하고 나아가서 시뮬레이션(Simulation)을 통하여 분석·비교한 뒤 추산잔차를 정의하기로 한다.

1. Introduction

We often use statistical models to do the data analysis. The main interests are how to construct the relation ship between the responses and the predictors for prediction and how to check the departures from model.

Regression is the best known statistical method to describe the response by the given predictors. There are unobservable errors in measurements of responses or predictors or both. We need assumptions for these errors when we build the regression models. Thus we assume that errors are independent, identically distributed normal variates with mean o and σ^2.

There are many tools to check the departures from model. Ordinary residuals have been the most common building blocks for checking models in linear regression models. Every statistical books and computer packages use ordinary residuals as diagnostics. But they have deficiencies as diagnostics, so recursive residuals are suggested(Brown, Durbin, and Evans(1975); Galpin and Hawkins (1984); Hawkins(1987)).

There will be further discussions in the following section. In section 3, we will discuss nonlinear regression models. In nonlinear regression models, unlikely as linear models, the exact behavior of residuals are intractable. We require various approximations to express different types of residuals. We will compare the distributional behaviors of the different types of residuals through the simulation study.

2. Linear Regression Models

We consider a standard linear regression model

$$Y=X\beta+\varepsilon \qquad (2.1)$$

Where Y is a n×1 vector of observed responses;

X is a n×p matrix of known constants;

β is a p×1 vector of unknown parameter to be estimated; ε is a n×1 vector of true errors which we assume they are independent, identically normally distributed with mean 0 and variance σ^2.

Let $\hat{\beta}$ be the maximum likelihood estimate of β. Then n×1 vector of ordinary residuals is

$$e=Y-X\hat{\beta} \qquad (2.2)$$

where $\hat{\beta}=(X^TX)^{-1}X^TY$.

By using (2.1), we can rewrite e as

$$e=X\beta+\varepsilon-X(X^TX)^{-1}X^TY$$

*Univesity of Minnesota 통계학과 박사과정

접수 : 1989. 4. 17.

$$=X\beta+\varepsilon-X(X^TX)^{-1}X^T(X\beta+\varepsilon)$$
$$=X\beta+\varepsilon-X\beta-X(X^TX)^{-1}X^T\varepsilon$$
$$=(I-P)\varepsilon \qquad (2.3)$$

where I is n×n identity matrix and $P=X(X^TX)^{-1}X^T$ is the projection operator onto the column space of X.

Under the model(2.1), $\varepsilon\sim N(0, \sigma^2 I)$, the ordinary residuals are distributed as

$$e\sim N(0, (I-P)\sigma^2) \qquad (2.4)$$

From(2.4), we can see the deficiencies of ordinary residuals as diagnostics : first, they are mutually correlated but ε are uncorrelated; second, they are heteroscedastic; third, they are dependent on P, that is, they are dependent on the structure of X; Fourth, their structure can be affected by one outlying observation.

Suppose we partition the data set into the first k cases and n−k cases. Then we can write X and Y as, for $k=p+1, \cdots\cdots, n$,

$$X^T=(X_k^T \quad X_{n-k}^T),\ Y^T=(Y_k^T \quad Y_{n-k}^T)$$

And $\hat{\beta}_k=(X_k^2\times X_k)^{-1}X_k^TY_k$ is the maximum likelihood estimate of β based on the first k cases; projection, operator $H_k=X_k(X_k^TX_k)^{-1}X_k^T$

The recursive residuals are defined by (Brown, Durbin, and Evans (1975)),

$$rr_K=\frac{y_K-X_K^T\hat{\beta}_K}{\sqrt{1-X_K^T(X_K^TX_K)^{-1}X_K}}$$
$$=\frac{y_K-X_K^T\hat{\beta}_{K-1}}{\sqrt{1-X_K^T(X_K^TX_{K-1})^{-1}X_K}} \qquad (2.5)$$
$$\text{for } K=p+1, \cdots\cdots, n$$

where X_K^T is K^{th} row of X.

The recursive residuals can be obtained by the iterative processes as above. These processes can be understood as :

1. For n cases (full data), fit the model and standardize the n^{th} residual so that $r_n\sim N(0, \sigma^2)$ then delete it.
2. For n−1 cases (without n^{th} case), fit the model and standardize the $(n-1)^{th}$ residual so that $r_{n-1}\sim N(0, \sigma^2)$ then delete it.
3. Continue the process until K=p+1.

Since we need at least p cases to fit the model, we can get only n−p recursive residuals. We can show that these recursive residuals are uncorrelated thus they are independent under normality. By the definition, with the independence, the recursive residuals are i.i.d N $(0, \sigma^2)$.

The recursive residuals are more effective under the situations when the model holds at one of the ends of data set. If the model confines to the first K cases, then, after deleting the last $n-k$ cases, the recursive residuals will not be affected by those outlying or influential cases.

3. Nonlinear Regression Models

A stadard nonlinear regression model is given by

$$y_i=f(x_i, \theta)+\varepsilon_i \qquad (3.1)$$
$$\text{for } i=1, \cdots, n$$

where x_i is Known constant associated with the i^{th} response value y_i; θ is a p×1 vector of unknown parameters; f is a response function which is assumed to be known, continuous, and twice differentiable with respect to θ; ε is n×1 vector of random variates which are i.i.d N $(0, \sigma^2)$. Let $f_i=f(x_i, \theta)$ for simplicity. We denote V is a n×p matrix of the first derivatives of f with respect to θ, then its element $V_{ij}=\frac{\partial f_i}{\partial \theta_j}$ for i=1, ……, n and j=1, ……, p. Let denote W be the n×p×p array of the second derivatives of f with respect to θ, then its element $wijk=\frac{\partial^2 f_i}{\partial \theta_j \partial \theta_k}$ for i=1, ……, n; j=1, ……, P; k=1, ……, P. The array can be shown as the stack of matrices. $\hat{\theta}$ is the maximum likelihood estimate of θ. V and W are evaluated at the true value of θ.

As mentioned earlier, we require various approximations to experss the different types of residuals. The most common approximation has been used is a linearization technique. The linear approximation (LA) assumes that the lifted line which are mapped on the solution locus corresponding to a line through the θ value of parameter space is linear. The solution locus in linear models is linear. If the lifted line is close to be linear, then almost of all diagnostics are similar to those of linear models. We simply replace X in linear models by V and use all inferences for diagnostics with the same manners. If the intrinsic curvature which measures the curvature of the lifted line at $f(\theta)$ is large, however, LA is no longer accurate. The interested reader may refer to Bates and Watts (1980). The refined approximation is called the quadratic approximation (QA). Since QA adopts the second derivatives of f, it is more accurate even when the intrinsic curvature is nonnegalizable. The calculation if much more complicated. If we employ the further approximations which are tedious, then we may expect the more accurate results. But QA is sufficient for this research.

3.1 Ordinary Residual

The ordinary residuals are

$$e=Y-f(\hat{\theta}) \qquad (3.1.1)$$

Where $f(\hat{\theta})$ is the fitted value.

With the differentiability assumption of f, we can obtain the second order Taylor's expansion of $f(\hat{\theta})$:

$$f(\hat{\theta}) \approx f(\theta)+V\cdot(\hat{\theta}-\theta)+\frac{1}{2}(\hat{\theta}-\theta)^{T}W(\hat{\theta}-\theta)$$
$$=f(\theta)+V\cdot\phi+\frac{1}{2}\phi^{T}W\phi$$

where $\phi=\hat{\theta}-\theta$

e can be rewritten as

$$e\approx Y-f(\theta)-V\phi-\frac{1}{2}\phi^{T}W\theta$$
$$=\varepsilon-V\phi-\frac{1}{2}\phi^{T}W\phi \qquad (3.1.2)$$

For (3.1.2) to be useful, we require the analogous expression for ϕ.

We can obtain the expression for ϕ by expanding the likelihood equation about θ and ignoring all terms that involve third and higher derivatives of f. The loglikelihood function is

$$L(\theta)=-\frac{1}{2\sigma^{2}}\sum_{i=1}^{n}(y_i-f_i(\theta))^{2} \qquad (3.1.3)$$

assuming $V^{T}V$ nonsingular, we can write ϕ as

$$\phi\approx(V^{T}V)^{-1}V^{T}\varepsilon+(V^{T}V)^{-1}[e^{T}][W]\phi-\frac{1}{2}(V^{T}V)^{-1}V^{T}(\phi^{T}W\phi) \qquad (3.1.4)$$

the first term in (3.1.4) is the standard LA of ϕ.

We substitute (3.1.4) into the second term of (3.1.2) then e can be rewritten

$$e\approx(I-H)\varepsilon-V(V^{T}V)^{-1}[e^{T}][W]\phi-\frac{1}{2}(I-H)(\phi^{T}W\phi) \qquad (3.1.5)$$

where $H=V(V^{T}V)^{-1}V^{T}$ is the projection operator onto the space spanned by the columns of V.

The first in (3.1.5) is the standard LA of e.

To obtain the quadratic approximation of e, we substitute $(I-H)\varepsilon$ and $(V^{T}V)^{-1}V^{T}\varepsilon$ for e and ϕ in (3.1.5) respectively. The QA of e is

$$e\approx(I-H)\varepsilon-V(V^{T}V)^{-1}[\varepsilon^{T}][I-H][W](V^{T}V)^{-1}V^{T}\varepsilon-\frac{1}{2}\varepsilon^{T}V(V^{T}V)^{-1}[I-H][W](V^{T}V)^{-1}V^{T}\varepsilon$$
$$=(I-H)\varepsilon-V(V^{T}V)^{-1}\varepsilon^{T}C\varepsilon-\frac{1}{2}\varepsilon^{T}B\varepsilon \qquad (3.1.6)$$

where $C=\{[I-H][W](V^{T}V)^{-1}V^{T}\}^{V}$ is a $p\times n\times n$ array composed of vertical faces of $[I-H][W](V^{T}\times V)^{-1}V^{T}$ and $B=V(V^{T}V)^{-1}[I-H][W](V^{T}V)^{-1}V^{T}$ is a $n\times n\times n$ array.

Here are theorems and facts, which are already known, to be useful for finding moments of different types of residuals:

suppose A, B be any $n\times n$ symmetric matrices

Theorem 1 : $E(\varepsilon^{T}A\varepsilon)=\sigma^{2}tr(A)$

Theorem 2 : $Var(\varepsilon^{T}A\varepsilon)=2\sigma^{4}t_r(A^{2})$

Theorem 3 : $Cov(\varepsilon^{T}A\varepsilon,\ \varepsilon^{T}B\varepsilon)=2\sigma^{4}t_r(AB)$

Fact 1 : $V^T(I-H)=(I-H)V=0$

Fact 2 : $E(\epsilon)=E(\epsilon^3)=0$

From (3.1.6), we can get the mean of e easily

$$E(e)\approx -\frac{1}{2}\sigma^2 t_r(B) \qquad (3.1.7)$$

where $t_r(B)$ is $n\times 1$ vector with its element $t_r(B_i)$, trace of i^{th} face of B.

We can easily show that the three terms in (3.1.6) are uncorrelated. Therefore the variance of e is sum of variances of each term.

$$Var(e)\approx \sigma^2(I-H)+2\sigma^4 V(V^TV)^{-1}t_r(C^2)(V^TV)^{-1}V^T+\frac{1}{2}\sigma^4 t_r(B^2) \qquad (3.1.8)$$

we know that e and the fitted values are uncorrelated in the linear regression models. In nonlinear regression models,

$$cov(e,\ f(\hat{\theta}))\approx \sigma^2(I-H)-var(e) \qquad (3.1.9)$$

thus they have negative correlation. It shows that e may strongly depend upon the θ value. This gives the motivation to the projected residual which will be discussed in section 3.1.3.

3.2 Predicted Residual

Suppose we fit the model based on $n-1$ cases. Then we can obtain $n-1$ ordinary residuals, that is,

$$e_i=Y_i-f(x_1,\bar{\theta}) \qquad (3.2.1)$$

for $i=1,\cdots\cdots,n-1$

where $\bar{\theta}$ is the maximum likelihood estimate of θ from the regression on $n-1$ cases. And we can get.

$$r_n=y_n-f(x_n,\hat{\theta}) \qquad (3.2.2)$$

We call r_n as a predicted residual.

In this section, we introduce some other notations to avoid the confusion, but they are corresponding to notations in the previous section.

1. $\overline{V}$ is $(n-1)\times p$ matrix of the first $n-1$ rows of V
2. $\overline{W}$ is $(n-1)\times p\times p$ array of the first $n-1$ faces of W
3. $\bar{\theta}$ is the maximum likelihood estimate of θ based on $n-1$ cases.

Let $r=(\bar{e}\ \ r_n)$ be $a_n\times 1$ vector of residuals from the regression which use only the first $n-1$ cases to estimate the parameters, where $\bar{e}$ is the $(n-1)\times 1$ vector with its element as (3.2.1). Then we follow the same processes to obtain the QA of r.

Sine $r=Y-f(\bar{\theta})$, we can rewrite r in the expanded form.

$$\begin{aligned} r&\approx Y-f(\theta)-V\overline{\phi}-\frac{1}{2}\overline{\phi}^T W\overline{\phi} \\ &=\epsilon-V\overline{\phi}-\frac{1}{2}\overline{\phi}^T W\overline{\phi} \qquad (3.2.3) \end{aligned}$$

where $\overline{\phi}=\bar{\theta}-\theta$

To obtain the expression for $\overline{\phi}$, we also expand the likelihood equation and ignore all the terms with third and hiqher derivatives of f.

Assumin $\overline{V}^T\overline{V}$ to be nonsingular,

$$\overline{\phi}\approx (\overline{V}^T\overline{V})^{-1}\overline{V}^T\bar{\epsilon}+(\overline{V}^T\overline{V})^{-1}[\bar{e}^T][\overline{W}]\overline{\phi}-\frac{1}{2}(\overline{V}^T\overline{V})^{-1}\overline{V}^T(\overline{\phi}^T\overline{W}\overline{\phi}) \qquad (3.2.4)$$

The first term in (3.2.4) is LA of $\overline{\phi}$.

We replace $\overline{\phi}$ of the second term in (3.2.3) by (3.2.4) then r can be rewritten as

$$r\approx \epsilon-V\{(\overline{V}^T\overline{V})^{-1}\overline{V}^T\bar{\epsilon}+(\overline{V}^T\overline{V})^{-1}[\bar{e}^T][\overline{W}]\overline{\phi}-\frac{1}{2}(\overline{V}^T(\overline{V})^{-1}\overline{V}^T(\overline{\phi}^T\overline{W}\overline{\phi})\}-\frac{1}{2}\overline{\phi}^T W\overline{\phi} \qquad (3.2.5)$$

From (3.2.5), $(\bar{I}-\overline{H})\bar{\epsilon}$ is LA of $\bar{e}$ where $\bar{I}$ is the $(n-1)\times(n-1)$ identity matrix and $\overline{H}=\overline{V}(\overline{V}^T\overline{V})^{-1}\overline{V}^T$.

We substitute $(\bar{I}-\overline{H})\bar{\epsilon}$, $(\overline{V}^T\overline{V})^{-1}\overline{V}^T\bar{\epsilon}$ for $\bar{e}$ and $\overline{\phi}$ in (3.2.5) respectively to obtain the QA of r.

We are interested in r_n only so we show the QA of r_n and moments of r_n in the rest of the present section. The QA of r_n is

$$r_n\approx a\cdot\epsilon-r_n^T(\overline{V}^T\overline{V})^{-1}\bar{\epsilon}^T dE\bar{\epsilon}-\frac{1}{2}\bar{\epsilon}^T G\bar{\epsilon} \qquad (3.2.6)$$

where $a=(-r_n^T(\overline{V}^T\overline{V})^{-1}\overline{V}^T\ \ 1)$ a $n\times 1$ vector;

$E=\{[\bar{I}-\overline{H}][\overline{W}](\overline{V}^T\overline{V})^{-1}\overline{V}^T\}^V$ is a $p\times(n-1)\times(n-1)$ array;

and $G=\overline{V}(\overline{V}^T\overline{V})^{-1}[a][w](\overline{V}^T\overline{V})^{-1}\overline{V}^T$ is a $(n-1)\times(n-1)$ matrix.

The mean of r_n is,

$$E(r_n) \approx -\frac{1}{2}\sigma^2 t_r(G) \qquad (3.2.7)$$

The variance of r_n is again the sum of variances of each term in (3.2.6).

$$Var(r_n) \approx \sigma^2 a \cdot a^T + 2\sigma^4 V_n^T (\bar{V}^T \bar{V})^{-1} t_r(E^2)(\bar{V}^T \bar{V})^{-1} V_n + \frac{1}{2}\sigma^4 t_r(G^2) \qquad (3.2.8)$$

3.3 Projected Residual

As seen in section 3.1, the distribution of e may strongly dependent on true θ. It makes hard to use e as the representation of ε, since the mean of ordinary residuals may subnstautially depart from 0. In (3.1.6), it is clear that the second term is in the column space of V and the third term is in the column space of [I−H][W]. These two terms contribute to the potential problems when the intrinsic curvature is large. Cook and Tsai (1985) proposed to project e onto the orthogonal space against the space spanned by the combination of V and W. This projection removes the dependence of e on θ substantially. They call this type as the projected residuals.

They define the projected residual, Pe, as

$$P^*_e \approx (I-H)\varepsilon - P'\varepsilon \qquad (3.3.1)$$

where P^* is the projector onto the orthogonal space against the space spanned by $P+\frac{P(P+1)}{2}$ vectors of V and W.; P' is the projector onto the orthogonal space against the space spanned by $\frac{P(P+1)}{2}$ vector of W; and $P^*=(I-H)+P'$. Therefore the moments of P^*_e are

$$E(P^*_e) \approx 0 \qquad (3.3.2)$$

$$Var(P^*_e) \approx P^*\sigma^2 \qquad (3.3.3)$$

3.4 Projected Ordinary Residual

In linear regression models, we know that e_n is uncorrelated with $\bar{e}$, $(n-1)\times 1$ vector of vesidual from the regression based on $n-1$ cases. The independence of recursive residuals as defined in section 2 could be obtained in part. However, we can not obtain the uncorrelatedness between e_n and $\bar{e}$ in nonlinear regression models.

We project e_n onto the space of $\bar{e}$, $(n-1)\times 1$ vector of residual from the regression based on n cases.

This projection will reduce the dependence of e_n on $\bar{e}$. So we regress e_n on $\bar{e}$ so that its residual is uncorrelated with $\bar{e}$. We call this residual as the projected ordinary residual, Pen. Then Pen is defined as

$$Pe_n \approx e_n - r\bar{e} \qquad (3.4.1)$$

where $r=cov(e_n,\bar{e})[cov(\bar{e},\bar{e})]^{-1}$

Therefore we obtain that $cov(Pe_n,\bar{e})=0$

We use the covariance and the variance terms to express the moments of Pe_n only for convenience.

The mean of Pe_n is

$$E(Pe_n) \approx E(e_n) - cov(e_n,\bar{e})[cov(\bar{e},\bar{e})]^{-1}E(\bar{e}) \qquad (3.4.2)$$

The variance of Pe_n is

$$Var(Pe_n) \approx Var(e_n) - 2Cov(e_n,\bar{e})[cov(\bar{e},\bar{e})]^{-1}cov(e_n,\bar{e})^T + cov(e_n,\bar{e})[cov(\bar{e},\bar{e})]^{-1} \cdot var(\bar{e}) \cdot [cov(\bar{e},\bar{e})]^{-1} cov(\bar{e},e_n) \qquad (3.4.3)$$

3.5 Projected Predicted Residual

The predicted residual r_n is uncorrelated with $\bar{e}$ in linear regression models. But they are not uncorrelated each other in nonlinear regression models. We try to reduce this dependency by using another projection method. We project rn onto the space of $\bar{e}$ so that its residual can be uncorrelated with $\bar{e}$. We call this residual as the projected predicted residual, Pr_n. Then Pr_n is defined as

$$Pr_n \approx r_n - \lambda\bar{e}$$

such that $cov(Pr_n,\bar{e})=0$, where

$\lambda = cov(r_n,\bar{e})[var(\bar{e})]^{-1}$.

The mean of Pr_n is

$$E(Pr_n) \approx E(r_n) - cov(r_n, \bar{e})[var(\bar{e})]^{-1}E(\bar{e}) \qquad (3.5.2)$$

The variance of Pr_n is

$$var(Pr_n) \approx var(r_n) - cov(r_n, \bar{e})[var(\bar{e})]^{-1}cov(\bar{e}, r_n) \qquad (3.5.3)$$

4. Simulation Study

In the previous section, we investigated the distri butional behaviors of five different residuals. Since the goal of this paper is to define the recursive residual in nonlinear regression models, we must check the residuals whether they have the same properties of recursive residuals as in linear models. As we have seen in section 3, it is very hard to compare them algebraically. thus we employ the simulation method to compare them in various aspects.

First, we use three different response functions which are

ⅰ) $f(\theta) = \theta_1 \cdot \sqrt{|x - \theta_2| + \delta}$

ⅱ) $f(\theta) = \theta_1 \cdot (1 + e^{\theta_2 \cdot x})$

ⅲ) $f(\theta) = \theta_1 \cdot e^{\frac{1}{|x - \theta_2| + n}}$

The magnitude of intrinsic curvatures of each model is in increasing order of (ⅱ), (ⅰ), (ⅲ)

Second, we use random values for x with different ranges which are

ⅰ) $X \sim U(0.4)$

ⅱ) $X \sim U(1.5, 2.5)$

ⅲ) $X \sim U(-2, 6)$

By changing the range of x, we tried to control the intrinsic curvature of each response function as above, in increasing order of (ⅲ), (ⅰ), (ⅱ).

Third, we use normal random number generator for ε, where $\varepsilon \sim N(0, 1)$ in the study.

Fourth, we generated data n=20 based on the above schemes with the true $\theta_1=2$, $\theta_2=2$ values, and 100 sample data set are used.

Fifth, we standardized each type of residual so that its mean is 0 and constant variance σ^2

Finally, we deleted case after case for each sample data set.

Now we summarize the results of the simulation study.

Table. 1.

f / x, K	Ⅰ			Ⅱ			Ⅲ		
	ⅰ	ⅱ	ⅲ	ⅰ	ⅱ	ⅲ	ⅰ	ⅱ	ⅲ
5	none	none	②	②③⑤	②⑤	②③⑤	none	none	②
6	①②⑤	②	①②④⑤	①②④⑤	①②⑤	same	②	none	②⑤
7	②④⑤	②④	②③④⑤	②④	②	②③④⑤	①②⑤	②	①②⑤
10	③④	②④	①②④⑤	same	same	same	②⑤	②⑤	①②⑤
20	same	same	same	same	same	same	same	same	same

remarks : none—no candidate, same-all equivalent, ① or ②… —good candidates

We used the pnormal probability plot to check the normality of each candidata at different K. All five candidates follow the normal distribution. After checking correlation matrix of each candidate, we cau say that they are all equivalently uncorrelated. There fore we conclude that the predicted residual is the most stable one among those residuals.

5. Summary

The recursive residuals are obtained by the iterative processes as descrbed in section 2. They may

require more efforts and time to compute and may face difficultie in ordering of data. But we can investigate each case to be deleted and gather more informations on each case. The recursive residuals are much more effective with conjecture of cusum technique.

We suggest to use the predicted residual for the construction of recursive residuals in nonlinear regression models. The assessment of influence and leverage by the connection with recursive residuals will be necessary.

참 고 문 헌

1. Bates, D. and Watts, D(1980) Relative curvature measures of nonlinearity(with discussion) J. Roy. Statist. Soc., Ser. B, 22, 41-88.
2. Brown, R. L., Durbin, J., and Evans, J. M.(1975) Techniques for testing the constancy of regression relationships (with discussion). J. Roy. Statist. Soc., Ser. B, 37, 149-63.
3. Cook, R. D.(1979) Influential observations in linear regression. J. Amer. Statist. Assoc., 74, 169-74.
4. Cook, R. D., and C. L. Tsai(1985) Residuals in nonlinear regression, Biometrika, 72(1), 23-29.
5. Cook, R. D., and Weisberg, S. (1982) Residuals and Influence in regression London : Chapman and Hall.
6. Galpin, J., and Hawkins, D. (1984) Use of recursive residuals in linear regression models, American statistician, 40, 35-48.
7. Gallant, A. R. (1987) Nonlinear Statistical Models. New York : Wiley.
8. Hawkins, D. M. (1980) Identification of Outliers. London : Chapman and Hall.
9. Hawkins, D. M. (1988) Diagnostic method with recursive residuals (submitted).
10. Plackett, R. L. (1950) Some theorems in least sqares. Biometrika, 37, 149-157.
11. Ratkowsky, D. A. (1983) Nonlinear Regression Modelling : A unified Practical Approach, New York : Marcel Dekker.
12. Scheffe, H. (1959) The awalysis of Variance. New York : Wiley.
13. Seber, G.A.F. (1977) Linear Regression Analysis. New York : Wiley.
14. Weisberg, S. (1985) Applied Linear Regression. New York : Wiley.