The Masking and Swamping Effects
Using the Planted Mean-Shift Outliers Models

Jung-Tsung Chiang

Department of Business Administration
Ling Tung University, Taiwan
No. 1 Ling-Tung Rd. Taichung city, Taiwan

Abstract

The masking and swamping effects are still unavoidable in linear models based on the OLS method. Actually, the data structure has provided much information on this topic. The Jackknife residuals and Cook's Distance are good indicators here. In this paper, we have shown that the locations of outliers, their signs of residuals, permutations, and the sum of deviations of all outliers are the main factors for the two effects using the planted mean-shift outliers models. The conclusion has been drawn from the simulation study.

Keywords: OLS method, Outliers, masking effect, swamping effect, Jackknife residuals, Cook's Distance, Mean-shift models, Data Structure

1 Introduction

Identification of outliers or influential points is a crucial part of regression diagnostics. This is because the “aberrant points” could draw the regression function towards them. The outliers in a sample reflect several features. They are from either (1) the errors of measurement or (2) intrinsic variability (mean shift, inflation of variances or others). In case(1), the outliers should be exclusive of the sample or need to be corrected. In case (2), more methods and works will be developed if possible.

A data set with multiple outliers becomes more complicated due to the masking and swamping effects. A number of authors create an Outlier Generating Models to analyze the “aberrant” observations. In general, the multiple outliers are only a small proportion of an entire data set. The good observations are from the target distribution, and the “contaminants” (outliers) are from the different distributions. Davis and Gather (1993) [6] give the definition as follows:
An outlier-generating model allows a small number of observations from a random sample to come from distributions $G_1, ..., G_k$ different from the target distribution $F$. The observations not from $F$ are called “contaminants”, and the task of identifying an unknown number $k$ of outliers is to detect the contaminants.

They suggested that If the target distribution $F$ is $N(\mu, \sigma^2)$, the $\alpha$ outliers region of $F$ is defined by

$$\text{out}(\alpha, \mu, \sigma^2) = \{Y_i : ||Y_i - \mu|| \geq Z_{1-\alpha/2}\}$$ (1)

where $\alpha$ is the level of significance, and $Z_q$ is the $q$ percentile of the $N(0, 1)$ distribution. Thus, the $i$th observation is called an outlier with respect to $N(\mu, \sigma^2)$ if $Y_i \in \text{out}(\alpha, \mu, \sigma^2)$. Note that an outlier here is equivalent to the Type I error when the outliers come from the same population.

If a sample of size $n$ contains no outliers, we call it a clean data set. For a clean data set, we can find a well-fitted function and approximate normal residual distribution to analyze it. Now there exists an entire data set of size $n$. The probability of getting a good observation (not outliers) is $1 - \alpha_n$, where $\alpha_n$ is the significance level for a single case. The probability of a clean entire data set is $(1 - \alpha_n)^n$, since all the $n$ observations are independently and identically distributed on $F$. Next, consider those outliers lie in an outlier region regardless of their different distributions. Then $(1 - \alpha_n)^n = 1 - \alpha$, where $\alpha$ is the type I error for the entire data set. Thus

$$\alpha_n = 1 - (1 - \alpha)^{1/n}$$ (2)

which is approximated by $\frac{\alpha}{n}$. The quantity $\frac{\alpha}{n}$ is just equal to the upper bound of Bonferroni inequality. Furthermore, a data set of size $n$ contains $k$ outliers in the generalized linear model was proposed by Gentleman and Wilk (1975b)[1], and they defined the “$k$ most likely outliers” whose deletion would produce the largest reduction in the sum of squared residuals.

Barnett(1978)[2] and Lewis(1984) [3] extensively investigated the case of the univariate regression model. They presented a method for detecting of single outlier in a linear model based on maximum absolute studentized residual criterion. Ellenberg (1976)[8] showed a different approach which is based on maximum reduction in the residual sum of squares. It is shown to be equivalent to the absolute maximum residual method. Basically, he suggested a test for outliers detection according to the probability of maximum residual to exceed a critical value calculated by a second order Bonferroni inequality. However, a single outlier is readily determined by the critical values. To use Bonferroni inequality in the determination of critical values is time-consuming [9]. Doornbos(1981)[7] demonstrated that the cumbersome computations required for a lower-bound correction can be rectified in a large sample. Several
authors studied multiple outliers detection. Gentleman and Wilk (1975a)[10] proposed the “Deletion Method” to identify a subset of $k$ outliers. Kiani-fard and Swallow (1989)[13] used the recursive residuals with an approximate Student’s t distribution to detect outliers or influential points. The procedure for identifying one or more outliers is investigated via simulation studies. In general, the use of recursive residuals with the adaptively-ordered observations improves power and helps to overcome the masking effect (an outlier is undetected because of the presence of another adjacent outlying observations)(See Hadi and Simonoff(1993)[12]).

Indeed, the word “outlier” is used ambiguously by most researchers, and it is hard to give a clean definition due to complicated masking and swamping effects. The adjusted residuals (Standardized and Studentized residuals) and the residual pattern recognition are widely used in detecting “several” deviant points. In this paper, we try to give an explicit and clean definition for outliers based on a clean data set, combined with a planted mean-shift outliers models. However, the problem arises in an effective factor space established to get the least squared equation in which the interpolation rule is attained.

2 Method Analysis–Studentized and Jackknife Residuals in Linear models

Consider the full rank linear model:

$$Y = X\beta + \epsilon$$

where $X$ is a known $n \times p$ ($n \geq p + 1$) full rank matrix. $\beta$ is an unknown $p \times 1$ vector, and $\epsilon$ is an $n \times 1$ error vector with i.i.d as $N(0, \sigma^2 I_n)$. So the least squares estimated residuals $e$ and its variance are:

$$e = Y - \hat{Y} = (I - H)Y$$

$$Var(e) = (I - H)\sigma^2$$

where the hat matrix $H = X(X'X)^{-1}X'$ is symmetric and idempotent.

For a general linear model (equation(3)), the variance of the residuals, $Var(e) = \sigma^2(I - H)$, are correlated, which would suggest the standardization of the $i$th residual to be $r_i$

$$r_i = \frac{e_i}{s \sqrt{1 - h_{ii}}}, i = 1, 2, ..., n$$

where $s^2 = \frac{e'e}{n-p}$ (Mean squared errors, MSE) is an estimate of $\sigma^2$. The $r_i$ are called Studentized Residuals. Behnken and Draper(1972)[5] proposed the test
statistic $R_n = \max |r_i|$ for a single outlier test in a simple linear regression, and took a simulation to calculate the critical values of $R_n$. For simplicity, the critical values are computed by

$$\left[(n-p)F/(n-p-1+F)\right]^{1/2}$$

It is exclusive of high leverage points of extreme observations.

Next, we sample an entire data of size $n$ and delete the $i$th observation, using the remaining $n-1$ observations to calculate the fitted value of the $i$th case, $\hat{Y}_{i(i)}$. The difference between the observed value $Y_i$ and $\hat{Y}_{i(i)}$ is called the deleted residuals of the $i$th case, denoted by $d_i = Y_i - \hat{Y}_{i(i)}$ (6)

The results related to this topic will be stated without proof below: (See Rencher p226[14])

$$\hat{\beta}_{(i)} = \hat{\beta} - \left(XX^\prime\right)^{-1}X_ie_i$$

$$SSE_{(i)} = SSE - \frac{e_i^2}{1 - h_{ii}}$$

where $\hat{\beta}_{(i)}$ is the estimator of $\beta_{(i)}$ calculated without the $i$th case; $SSE = e'e$ and $SSE_{(i)} = e'_{(i)}e_{(i)}$ are the residual sum of squares with and without $i$th case, respectively.

The deleted residual for the $i$th case $d_i$ and the estimated variance of $d_i$ are:

$$d_i = \frac{e_i}{1 - h_{ii}}$$

$$\text{Var}(d_i) = \frac{MSE_{(i)}}{1 - h_{ii}}$$

where $MSE_{(i)}$ is the mean squared error without the $i$th case. The statistic $\frac{d_i}{\sqrt{\text{Var}(d_i)}}$ is defined as $r_i^*$. Consequently,

$$r_i^* = \frac{e_i}{s_{(i)}\sqrt{1 - h_{ii}}}, \ (1 = 1, 2, ..., n)$$

where $s_{(i)}^2 = \frac{Y_{(i)}'(I-H_{(i)})Y_{(i)}}{n-p-1}$, and $h_{ii} = X_i(X'X)^{-1}X_i'$, the $i$th diagonal element of $H$. The $i$th observation is deleted and indicated by writing the index (i) in brackets. The $r_i^*$ is called Jackknife residuals by Atkinson(1983)[1]. Next, let the $(Y_{(i)}, X_{(i)})$ be a remaining data matrix without $i$th observation. If $\text{rank}(X_{(i)}) = p$ and $\epsilon$ are i.i.d with $N(0, \sigma^2 I_n)$, then the Jackknife residuals $r_i^*$, (i=1, 2, ..., n) are $t_{n-p-1}$ distributed. If $\text{rank}(X_{(i)}) = p-1$, then the residuals
$r^*_i$ is not defined (Beckman and Trussel, 1974 [4]). In addition, the relationship between $r^*_i$ and $r_i$ is

$$r^*_i = r_i \sqrt{\frac{n - p - 1}{n - p - r_i^2}}$$

which is a strictly increasing function of $r_i$. The test statistic $\max_i |r^*_i|$ is equivalent to maximum studentized test.

In linear models mentioned above, the OLS estimators $\hat{\beta}$ and $\hat{\sigma}^2$ are unbiased under the normal assumptions. There are several useful theorems as follows:

**Theorem 2.1** In the linear model (equation (3)), the estimator $\hat{\beta}$ is computed with all of the $n$ observations of the data set. If several observations with “zero” residuals, i.e., $Y_i = \hat{Y}_i$, then the $\hat{\beta}$ is invariant after deleting them.

**Proof:** Suppose that the residual of the $i$th observation is zero. It implies that $Y_i = \hat{Y}_i$. From Cook’s distance $D_i$ is defined by

$$D_i(X'X, ps^2) = \frac{(\hat{\beta}_{(i)} - \hat{\beta})'(X'X)(\hat{\beta}_{(i)} - \hat{\beta})}{ps^2}$$

Alternatively, $D_i$ can be written as the simple form

$$D_i = \frac{r^2_i h_{ii}}{p(1 - h_{ii})}$$

where $r_i$ is the studentized residual of the $i$th case. Thus, we get $\hat{\beta}_{(i)} = \hat{\beta}$ when $e_i$ equals zero. Similarly, the other observations with zero residuals can be deleted in steps. The theorem holds.

**Corollary 2.1** The estimator $\hat{\beta}$ is still invariant if some new observations, say, $(X'_{n+1}, X'_{n+1} \hat{\beta})$, ..., $(X'_{n+k}, X'_{n+k} \hat{\beta})$, are added to the original data set.

**Proof:** Assuming that the new data set contains $n + k$ observations, the last $k$ observations with zero residuals. Thus, the result follows from the Theorem (2.1).

### 3 Masking and Swamping effects

The “masking effect” means that an outlier is undetected because of the presence of another adjacent ones. And “swamping effects” is that a good observation is incorrectly identified as an outlier because of the presence of another clean subset. Both of them are produced from the OLS method together with the planted mean-shift outliers models[12].

A data set $M$ of size $n$ contains $k$ outliers. The entire clean data set $C$ with $n_c$, ($n_c = n - k$), good observations is a subset of $M$. Now suppose that all good observations of the subset $C$ are from the target distribution as (equation (3),
i.e., \( Y_c \sim N(X_c \beta, \sigma^2 I_n) \). The OLS estimator \( \hat{\beta}_c = (X_c'X_c)^{-1}X_c'Y_c \) is invariant if \( k \) new observations \( (X_{n+1}, \hat{Y}_{n+1}), \ldots, (X_{n+k}, \hat{Y}_{n+k}) \) are added to the data set \( C \), (Theorem 2.1). That is,

\[
\hat{\beta}_c = (X'X)^{-1}XY
\]

where \( X \) is an \( n \times P \) matrix , \( \hat{Y} = (Y_1, \ldots, Y_{n_c}, \hat{Y}_{n_c+1}, \ldots, \hat{Y}_{n_k})' \), and \( \hat{Y}_{n_c+j} = X'_{n_c+j}\hat{\beta}_c, j = 1, \ldots, k \). The new data is denoted by \( B \). A predictive confidence interval \( I_p \) derived from subset \( C \) is

\[
X_j'\hat{\beta}_c \pm C_a s_c \sqrt{1 + X_j'(X_c'X_c)^{-1}X_j}
\]

where \( C_a = t(1 - \frac{a}{2n_c}, n_c - p - 1) \) is a critical point, and \( s_c \) is an estimate of \( \sigma \). The data set \( B \) of \( (X, \hat{Y}) \) is associated with the residuals: \( (e_1, e_2, \ldots, e_{n_c}, 0, \ldots, 0) \), and \( Y_i \in I_p \), for \( i = 1, 2, \ldots, n_c \). Next, let the original date set \( M \) be listed below:

\[
\{(X_1, Y_1), \ldots, (X_{n_c}, Y_{n_c}), (X_{n_c+1}, \hat{Y}_{n_c+1} + d_{n_c+1}), \ldots, (X_{n_c+k}, \hat{Y}_{n_c+k} + d_{n_c+k})\}
\]

where \( |d_j| \) is greater than \( 2\sigma Z_{1-\alpha/2} \), \( j = n_c + 1, \ldots, n_c + k \), and the last \( k \) observations of data set \( M \) are outliers. Here, the \( k \) outlier are planted to the data set \( B \) of \( (X, \hat{Y}) \), and the defined mean-shift outliers models will be proposed later if \( |d_j| \) is constant. Obviously, \( Y_M = \hat{Y} + \Delta Y \) if the \( \Delta Y = (0, \ldots, 0, d_{n_c+1}, \ldots, d_{n_c+k}) \). The OLS residuals of the original data set \( M \) is

\[
e' = (I - H)Y_M
\]

\[
= (I - H)(\hat{Y} + \Delta Y)
\]

\[
= (I - H)\hat{Y} + (I - H)\Delta Y
\]

\[
= (e_1, \ldots, e_{n_c}, 0, \ldots, 0)' + \Delta Y - H\Delta Y
\]

\[
= (e_1, \ldots, e_{n_c}, d_{n_c+1}, \ldots, d_{n_c+k})' - H\Delta Y
\]

\[
= e - H\Delta Y
\]

where \( H \) is an \( n \) squared hat matrix. \( e \) is an \( n \times 1 \) vector calculated by the clean data set \( C \). Consequently,

\[
e'_1 = e_1 - \sum_{j=n_c+1}^{n} h_{1j}d_j
\]

\[
e'_2 = e_2 - \sum_{j=n_c+1}^{n} h_{2j}d_j
\]

\[
\vdots
\]

\[
e'_{n_c} = e_{n_c} - \sum_{j=n_c+1}^{n} h_{nj}d_j
\]
\[ e'_{n_c+1} = d_{n_c+1} - \sum_{j=n_c+1}^{n} h_{n_c+1j}d_j \]
\[ \vdots \]
\[ e'_{n_c+k} = d_{n_c+k} - \sum_{j=n_c+1}^{n} h_{n_c+kj}d_j \]

The associated studentized residuals \( r'_i \) corresponding to \( e'_i \) are
\[ r'_i = \frac{e_i - \sum_{j=n_c+1}^{n} h_{ij}d_j}{\sqrt{1 - h_{ii}}} , \quad i f \quad i = 1, 2, \ldots, n_c \] (17)
And
\[ r'_i = \frac{d_i - \sum_{j=n_c+1}^{n} h_{ij}d_j}{\sqrt{1 - h_{ii}}} , \quad i f \quad i = n_c + 1, \ldots, n_c + k \] (18)

which shows that (1) For \( i = 1, 2, \ldots, n_c \), there exist some \( r'_i \)s are greater than critical points, the swamping effect appears on these cases; that is, they are incorrectly regarded as outliers. (2) For \( i = n_c + 1, \ldots, n_c + k \), there exist some \( r'_i \)s are less than the critical points, the masking effect appears on these cases; that is, they are incorrectly regarded as good observations (inliers). Thus, the masking and swamping effects are dependent of the values \( d_j \), their locations \( h_{ii} \) and correlations \( h_{ij} \).

A special case: \( X = (1, 1, \ldots, 1)' \), i.e., \( Y = \beta_0 + \epsilon \), \( \epsilon \sim N(0, \sigma^2 I_n) \), \( h_{ii} = \frac{1}{n} \). Then
\[ r'_i = \frac{ne_i - \sum_{j=n_c+1}^{n} d_j}{\sqrt{n(n-1)}} , \quad i f \quad i = 1, 2, \ldots, n_c \] (19)
And
\[ r'_i = \frac{nd_i - \sum_{j=n_c+1}^{n} d_j}{\sqrt{n(n-1)}} , \quad i f \quad i = n_c + 1, \ldots, n_c + k \] (20)

(i) If \( e_i > 0, \sum_{j=n_c+1}^{n} d_j < 0 \) such that \( e'_i \gg e_i \), then the swamping effect appears on these observations. Similarly, so does \( e_i > 0, \sum_{j=n_c+1}^{n} d_j < 0 \), for \( i = 1, 2, \ldots, n_c \)
(ii) If \( d_i > 0, \sum_{j=n_c+1}^{n} d_j > 0 \) such that \( |e'_i| \ll |e_i| \), then the masking effect appears on these observations. Similarly, so does \( d_i < 0, \sum_{j=n_c+1}^{n} d_j < 0 \), for \( i = 1, 2, \ldots, n_c \)

As argued above, the masking and swamping effects depend on the locations, signs of \( d_i \), and the sum of \( d_i \) of these outliers. An arranged form of \( k \) outliers may be one of them below:

\[ (+, +, \ldots, +) , \quad for \ all \ d_i > 0 \]
\[ (+, +, \ldots, -) , \quad only \ d_n < 0 \]
\[ (-, -, \ldots, -), \text{for all } d_i < 0 \]

There are \(2^k\) ways of permutations above. Basically, the \(k\) outliers can be regarded as a “perturbation effects” on a clean data set \(C\).

## 4 Simulation Study–The Locations of Outliers

### 4.1 \(Y = X + \epsilon, X_i \sim U(0, 10), \epsilon \sim N(0, 1)\)

1000 samples of size 25, 35, ..., 70 are generated from software \(R\). Two or three outliers are created by the adding the quantity of 4 to the observations of 1, 2 or 1, 2, 3. We want to examine the absolute Jackknife residuals of these outliers if they are ranking at top 20 % of \(|r_i^*|\) in a sample. The results of simulation study depend on the sample of size and the locations of outliers. Two tables are completed below:

**Table 1: The frequency of two outliers identified at top 20 % of \(|r_i^*|\) in 1000 samples, (\(p=2\))**

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Case I</th>
<th>Case II</th>
<th>Sample Size</th>
<th>Case I</th>
<th>Case II</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>983</td>
<td>956</td>
<td>50</td>
<td>992</td>
<td>989</td>
</tr>
<tr>
<td>35</td>
<td>992</td>
<td>985</td>
<td>60</td>
<td>989</td>
<td>990</td>
</tr>
<tr>
<td>45</td>
<td>995</td>
<td>980</td>
<td>70</td>
<td>995</td>
<td>992</td>
</tr>
</tbody>
</table>

Case I is the two outliers at opposite sides. That is, \(Y_1 = X_1 + 4 + \epsilon, \) and \(Y_2 = X_2 - 4 + \epsilon.\) Case II is the two outliers at the same side. That is, \(Y_1 = X_1 + 4 + \epsilon, \) and \(Y_2 = X_2 + 4 + \epsilon.\) Obviously, the masking effect apparently appears at Case II and small samples.

**Table 2: Three Outliers**

Case I shows that two outliers are at the same side, and the third one at opposite one. That is, \(Y_1 = X_1 + 4 + \epsilon, Y_2 = X_2 + 4 + \epsilon\) and \(Y_3 = X_3 - 4 + \epsilon.\) Case II shows that three outliers are at the same side. That is, \(Y_1 = X_1 + 4 + \epsilon, \) \(Y_2 = X_2 + 4 + \epsilon,\) and \(Y_3 = X_3 + 4 + \epsilon.\) Obviously, the masking effect apparently appears at Case II and small samples.
Table 2: The frequency of three outliers identified at top 20% of $\{|r^*_i|\}$ in 1000 samples, (p=2)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Case I</th>
<th>Case II</th>
<th>Sample Size</th>
<th>Case I</th>
<th>Case II</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>960</td>
<td>878</td>
<td>50</td>
<td>979</td>
<td>973</td>
</tr>
<tr>
<td>35</td>
<td>974</td>
<td>951</td>
<td>60</td>
<td>979</td>
<td>986</td>
</tr>
<tr>
<td>45</td>
<td>978</td>
<td>967</td>
<td>70</td>
<td>987</td>
<td>977</td>
</tr>
</tbody>
</table>

4.2 $Y = X1 + X2 + \epsilon, \epsilon \sim N(0, 1)$

The data sets are generated from the linear model $Y = X1 + X2 + \epsilon$, where $X_{1i} \sim U(0,10), X_{2i} \sim U(0,15)$ and $\epsilon \sim N(0,1)$. Three outliers are planted at the first three observations. Case I and II are mentioned as before. Table 3 is completed through 1000 samples based on different sizes of 25, 35, ..., 70.

Table 3: The frequency of three outliers identified at top 20% of $\{|r^*_i|\}$ in 1000 samples, (p=3)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Case I</th>
<th>Case II</th>
<th>Sample Size</th>
<th>Case I</th>
<th>Case II</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>939</td>
<td>825</td>
<td>50</td>
<td>986</td>
<td>972</td>
</tr>
<tr>
<td>35</td>
<td>967</td>
<td>926</td>
<td>60</td>
<td>979</td>
<td>970</td>
</tr>
<tr>
<td>45</td>
<td>980</td>
<td>958</td>
<td>70</td>
<td>986</td>
<td>976</td>
</tr>
</tbody>
</table>

Still, the masking effect is in the presence of Case II and small sample of size 25.

4.3 $Y = X1 + X2 + X3 + \epsilon, \epsilon \sim N(0, 1)$

This study is limited for “four outliers” in a sample of size 60 or 100. The data sets are generated from the linear model $Y = X1 + X2 + X3 + \epsilon$, where $X_{1i} \sim U(0,10), X_{2i} \sim U(0,10), X_{3i} \sim U(0,15)$ and $\epsilon \sim N(0,1)$. Four outliers are planted at the first four observations in a sample, and there are three cases I, II, III in this study:

(I) $Y_i = X_{1i} + X_{2i} + X_{3i} + 4 + \epsilon_i$, if $i = 1, 2, 3, 4$

(II) $Y_i = X_{1i} + X_{2i} + X_{3i} + 4 + \epsilon$, if $i = 1, 2, 3, 4$

(III) $Y_i = X_{1i} + X_{2i} + X_{3i} + 4 + \epsilon$, if $i = 1, 2, 3, 4$

The frequency in the table is computed by 1000 samples based on the three
cases, and the size of 60 and 100, respectively.

Table 4: The frequency of four outliers identified at top 20% of $\{|r^*_i|\}$ in 1000 samples, ($p=4$)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Case I</th>
<th>Case II</th>
<th>Case III</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>976</td>
<td>956</td>
<td>925</td>
</tr>
<tr>
<td>100</td>
<td>982</td>
<td>980</td>
<td>969</td>
</tr>
</tbody>
</table>

Still, the masking effect apparently appears at Case III; that is, four outliers are at the same side. In general, the masking and swamping effects will disappear in large samples. However, the absolute Jackknife residuals of outliers are almost ranked at top 20% of $\{|r^*_i|\}$. Next, the swamping effect also could be detected based on the simulations, although we never show it here.

5 Conclusions

The OLS method is widely used in linear regression models in this article. The $k$ outliers of confirmation are completed by $C \cup \{i\}, \forall i \in \{\text{outliers}\}$, ($C$: the best construction data set), and the masking and swamping effects are still unavoidable if the data set contains a few high leverage points. Moreover, the data structure has provided much information on this topic. The Jackknife residuals and Cook’s Distance are good indicators here. On the other hand, the locations of outliers, their signs of residuals, their permutations and the total sum of deviations of all outliers are the main factors for masking and swamping effects. The conclusion has been drawn from the simulation study.

ACKNOWLEDGEMENTS

The author would like to thank his advisor Dr. Kenny Ye, Prof. Nancy Mendell, and Prof. Hongshik Ahn for their assistances in modifying the paper. This paper is a part of his dissertation in August 2002, AMS Dept. of SUNY-Stony Brook, USA.

References


Received: July 10, 2006