Open Access Library Journal
2020, Volume 7, e6619
ISSN Online: 2333-9721
ISSN Print: 2333-9705

Outlier Detection and Effects on Modeling

Christopher O. Arimie[1], Emmanuel O. Biu[2], Maxwell A. Ijomah[2]

[1]Department of Radiology, University of Port Harcourt Teaching Hospital, Port Harcourt, Nigeria
[2]Department of Mathematics and Statistics, University of Port Harcourt, Port Harcourt, Nigeria
Email: codarimie@yahoo.com, biu.emmanuel@uniport.edu.ng, Maxwell.ijomah@uniport.edu.ng

How to cite this paper: Arimie, C.O., Biu, E.O. and Ijomah, M.A. (2020) Outlier Detection and Effects on Modeling. *Open Access Library Journal*, **7**: e6619.
https://doi.org/10.4236/oalib.1106619

Received: July 16, 2020
Accepted: September 13, 2020
Published: September 16, 2020

Abstract

In this work, a comprehensive framework for traditional outlier detection techniques based on simple and multiple linear regression models was studied. Two data sets were used for the illustration and evaluation of each class of outlier detection techniques (analytical and graphical methods). Outlier detection aims at identifying such outlier in order to improve the analytic of data and suitable model built. Furthermore, comparisons of the different methods were done to highlight the advantages, disadvantages and performance issues of each class of outlier detection techniques. The results show that by removing the influential points (or outliers), the model adequacy increased (from $R^2 = 0.72$ to $R^2 = 0.97$). It was observed that Jackknife residuals and Atkinson's measure methods are very useful in detecting outliers; hence, both methods were recommended for outliers' detection.

Subject Areas

Applied Statistical Mathematics

Keywords

Outliers' Detection, Classification and Comparisons, Simple and Multiple Linear Regression Models

1. Background of the Study

Outliers are observations that appear inconsistent with the remainder of the dataset. Bollen and Jackman [1] defined outliers as observations that are distinct from most of the data points in the sample. Hawkins [2] described outlier as an observation that deviates so much from other observations as to arouse suspicions that it was generated by different mechanisms. Dixon [3] sees outlier as values that are dubious in the eyes of the researcher and contaminants. Outliers are often present in real data but may go unnoticed because nowadays much data are produced by computer without careful inspection or screening. Outliers may be mistakes in data entry or otherwise, accurate but unexpected observations which could shed new light on the phenomenon under study.

In general, outliers can be classified into two types: man-made one and random one. Man-made outliers may arise because of typographical error(s), misreporting of information, incorrect distribution assumption and sampling error. Random outliers, on the other hand, may arise because of random chance for drawing sample from a population. Presence of man-made or random outliers or both would seriously influence the result of statistical analysis including point and interval estimates and type 1 and type 11 errors. Outlier can cause us to misinterpret patterns in plots; it can affect visual resolution of remaining data in plot (forces observations into clusters) and may indicate that our model fails to capture important characteristics of the data. Unusual cases can substantially influence the fit of the ordinary least square model thus, leading to faulty conclusion. Some man-made outliers can be avoided by a strict data entry and rechecking process before conducting a statistical analysis. Data transformation is another way to reduce the influence of outliers (see Barnett and Lewis [4]; Montgomery [5]).

Quite a number of authors have proposed different methods of detecting outliers.

Abuzaid *et al.* [6], proposed a number of diagrammatical plots and hypothesis testing to detect outliers in the simple circular regression model. The work focused on detecting outlier in the down and Mardia's circular-circular regression model. Zhang *et al.* [7] presented a method which applies signal processing techniques to solve important problems in data mining. They introduced an outlier detection approach termed to find out, based on wavelet transform. The main idea in the method is to remove the clusters from the original data and then identify the outliers. Rousseeuw [8] proposed a depth based method to detect outlier. The data points are organized in layers in the data space according to the value of the point depth. Aggarwal and Yu [9] proposed a method which entails studying the projection from the rest of the data in a sparse data with high dimensionality. Arning *et al.* [10] introduced a method which relies on the observations such that after seeing a series of similar data, an element disturbing the series is considered an outlier.

Hodge and Austin [11] identified an efficient method for on-line classification and outlier detection in multivariate sensor data. This involved a comparative review to identify and distinguish their advantages and disadvantages and introducing a survey of contemporary outlier detection techniques. Sebert *et al.* [12] and Montgomery [5] asserts that to identify the existence of outliers the standardized residuals are computed and a large standardized residual of ($d > 3$) indicates outlier. Worden *et al.* [13] used the concept of discordances to signal deviance from the norm.

Kitagawa [14] and Wing-Kam Fung and Bacon-Shone [15] applied Akaike information criterion (AIC) in detection of outliers by using (quasi) Bayesian approach with predictive likelihood. Belsey *et al.* [16] suggested standardizing each residual with an estimate of its standard deviation that is independent of the residual. This is accomplished by using as the estimate of variance for the ith residual, the residual mean square from an analysis where that observation has been omitted. The result is a jackknife residual also called a fully studentized residual.

In this study, concentration is on the effect of outlier as well as detection methods on linear regression model. Specifically, we are concerned with observations that differ from the regression plane defined by the data. It is important to identify these types of outliers in regression modeling because the observations when undetected can lead to erroneous parameter estimates and inference from the model. Deleting outliers from a regression model can sometimes give completely different results as bias or distortion of estimates are removed. Identifying outliers in the real world data-base is important for improving the quality of original data and for reducing the impact of outliers. Identifying outliers and high-leverage points, is a fundamental step in the least-squares regression model building process. On this note, it is the purpose of this research to examine the effect of outliers on simple and multiple linear regression, compare different methods of detecting outlier and show its effect on modeling.

2. Methods of Outlier Detection and Effects on Regression Model

There are many methods for the detection of outliers in linear regression. They may be classified into graphical and analytical methods. The graphical methods include Scatter graph, Boxplot, Williams graph, Rankit graph (or Q-Q Plot) and graph of predicted residuals. The analytical methods are predicted residuals, standardized residuals, studentized residuals, Jack-knife residuals, Cook's distance, Different-in-fits (DFFITS) and Atkinson's measure.

The general linear model is given as

$$y_{n\times 1} = X_{n\times k}\beta_{k\times 1} + \varepsilon_{n\times 1} \tag{2.1}$$

where,

$y_{n\times 1}$ is the observation (dependent variables);

$X_{n\times k}$ is the design matrix including a constant;

$\varepsilon \sim N\left(0,\sigma^2 I\right)$ is the error term;

$\beta = \left(\beta_0,\beta_1,\beta_2,\cdots,\beta_{k-1}\right)'$ is the coefficient and k is the number of covariates or predictors for each observation.

Then, the least square estimator is given as

$$\beta = \left(X'X\right)^{-1}\left(X'Y\right) \tag{2.2}$$

where the fitted (predicted) values for mean of Y are

$$\hat{Y} = \mathrm{X}\beta = \mathrm{X}\left(\mathrm{X}'\mathrm{X}\right)^{-1}\left(\mathrm{X}'Y\right) = \mathrm{H}Y \tag{2.3}$$

where $H = X\left(X'X\right)^{-1}X'$ is the projection matrix (or hat matrix).

The i^{th} diagonal element of H (given by $h_i = x_i'(\mathrm{X}'\mathrm{X})^{-1} x_i$ is known as the leverage of the i^{th} observation).

Similarly, the i^{th} element of the residual vector

$$\varepsilon_i = Y_i - \hat{Y}_i = (I - H) Y_i \tag{2.4}$$

where H is as defined in Equation (2.3), I is the identity matrix and Y_i is corresponding fitted value ($i = 1, 2, \cdots, n$).

In this study, Outliers were detected using the following methods:

2.1. Studentized and Standardized Residuals

The Studentized residual is obtained by

$$\varepsilon_{S.i} = \frac{\hat{\varepsilon}_i}{\hat{\sigma}\sqrt{1-h_i}} = \frac{\hat{\varepsilon}_i}{\sqrt{\hat{\sigma}^2(1-h_i)}} \tag{2.5}$$

where $\hat{\sigma}$ is an appropriate estimate of σ, and the estimate of $\hat{\sigma}^2$ (Mean Residual Sum of Squares) is the internally studentized residuals. h_i is leverage points as already defined.

$$\hat{\sigma}^2 = \frac{1}{n-k}\sum_{j=1}^{n} \hat{\varepsilon}_j \tag{2.6}$$

where k is the number of parameters in the model (Equation (2.1)).

Studentized residuals with large absolute values are considered large. If the regression model is appropriate, with no outlying observations, each Studentized residual follows a t distribution with $n - k - 1$ degrees of freedom.

CRITICAL: Each deleted residual has a student's t-distribution with $n - k - 1$ degrees of freedom.

If the Studentized residual is divided by the estimates of its standard error so that the outcome is a residual with zero mean and standard deviation one, it becomes standardized residual denoted by

$$\varepsilon_{ST.i} = \frac{\hat{\varepsilon}_i}{sd(\sigma)} \tag{2.7}$$

where $sd(\sigma)$ is the standard deviation of $\hat{\sigma}^2$ in Equation (2.6).

CRITICAL: The standardized residuals, $d_i > 3$ potentially indicate outlier.

2.2. Jackknife Residuals

The jackknife residuals are residuals which with an assumption of normality of errors have a student distribution with $(n - k - 1)$ degrees of freedom. The formula is:

$$\varepsilon_{J.i} = \hat{\varepsilon}_{S.i}\sqrt{\frac{(n-k-1)}{\left(n-k-\hat{\varepsilon}_{S.i}^2\right)}} \tag{2.8}$$

where $\varepsilon_{S.i}$ is the Studentized residuals Equation (2.5).

The jackknife residual examines the influence of individual point on the mean quadratic error of prediction.

CRITICAL: Each deleted residual has a student's t-distribution with $n-k-1$ degrees of freedom.

2.3. Predicted Residuals

The predicted residual or Cross-validated residuals for observation i is defined as the residual for the i^{th} observation that results from dropping the i^{th} observation from the parameter estimates. The sum of squares of predicted residual errors is called the PRESS statistic:

$$\varepsilon_{P,i} = \frac{\varepsilon_i}{1-h_i} \tag{2.9}$$

PRESS is called Prediction sum of squares; an assessment of your model's predictive ability. PRESS, similar to the residual sum of squares, is the sum of squares of the prediction error. In least squares regression, PRESS is calculated with the following formula:

$$\text{PRESS} = \sum_{i=1}^{n}\left(\frac{\varepsilon_i}{1-h_i}\right)^2 \tag{2.10}$$

where ε_i = residual and h_i = leverage value for the i^{th} observation. In general, the smaller the PRESS value, the better the model's predictive ability.

2.4. Cook's Distance

Cook's distance D_i of observation, i is defined as the sum of all the changes in the regression model when observation i is removed from it. Cook [17] proposed a statistic for detection of outlier, given as:

$$D_i = \frac{\sum_{j=1}^{n}\left(y_j - \hat{y}_{j(i)}\right)^2}{kS^2} \tag{2.11}$$

and $S^2 = (n-k)^{-1}\varepsilon'\varepsilon$ is the mean squared error of the regression model. Equivalently, it can be expressed using the leverage

$$D_i = \frac{\varepsilon_i^2}{kS^2}\left[\frac{h_i}{(1-h_i)^2}\right] \tag{2.12}$$

Here, D_i measures the sum of squared changes in the predictions when observation "i" is not used in estimating β. D_i approximately follows $F(p, n-p)$ distribution.

CRITICAL: The cut off value of Cook-Statistic is $4/n$.

2.5. Difference-in-Fit (DFFIT)

It is the difference between the predicted responses from the model constructed using complete data and the predicted responses from the model constructed by setting the i^{th} observation aside. It is similar to cook's distance. Unlike cook's distance, it does not look at all of the predicted values with the i^{th} observation set aside. It looks only at the predicted values for the i^{th} observation. It combines leverage and studentized residual (deleted t residuals) values into one overall measure of how unusual an observation is. DFFIT is computed as follows:

$$\text{DFFIT} = \varepsilon_i \left[\sqrt{\frac{n-k-1}{\sigma^2 (1-h_i) - \varepsilon_i^2}} \right] \sqrt{\frac{h_i}{1-h_i}} \tag{2.13}$$

where ε_i = residual, n = sample size, k = the number of parameters in the model, σ^2 = variance and h_i = leverage value for the i^{th} observation.

CRITICAL: The cut off value of DFFIT is $2\sqrt{\frac{k}{n}}$.

2.6. Atkinson's Measure (A_i)

It enhances the sensitivity of distance measures to high-leverage point. This modified version of cook's measure D_i suggested by Atkinson is even more closely related to Belsey *et al.* (1980) [16] DFFITS and has the form

$$A_i = |\varepsilon_{J.i}| \left[\sqrt{\frac{n-k}{k} \times \frac{h_i}{1-h_i}} \right] \tag{2.14}$$

where n, k, h_i are as defined in Equations (2.13) and $\varepsilon_{J.i}$ is the absolute value of Jackknife residuals.

This measure is also convenient for graphical interpretation.

2.7. Scatter Graph and Box Plot

Scatter plot is a line of best fit (alternatively called "trendline") drawn in order to study the relationship between the variables measured. For a set of data variables (dimensions) $X_1, X_2, \cdots, X_k$, the scatter plot matrix shows all the pairwise scatter plots of the variables on the dependent variable.

A box plot is a method for graphically depicting groups of numerical data through their quartiles (*i.e.* Mean, Median Mode, quartiles). Box plots may also have lines extending vertically from the boxes (whiskers) indicating variability outside the upper and lower quartiles. It is also called box-and-whisker plot and box-and-whisker diagram. Outliers may be plotted as individual points and it can be used for outlier detection in regression model, where the primary aim here is not to fit a regression model but find out outliers using regression and to improve a regression model by removing the outliers.

2.8. Willams Graph, Rankit Graph (or Q-Q Plot) and Predicted Residuals Graph

The Williams graph (Williams [18]) has the diagonal elements H_{ii} on the x-axis and the jackknife residuals $\hat{\varepsilon}_{Ji}$ on the y-axis. Two boundary lines are drawn, the first for outlier $y = t_{0.95}(n - k - i)$, and the second for high leverages, $x = 2k/n$. Note: $t_{0.95}(n - k - 1)$ is the 95% quantile of the student distribution with $(n - k - 1)$ degrees of freedom.

The Q-Q plot (or Rankit Graph) has the quantile of the standardized normal distribution μp_i for $P_i = i/(n+1)$ on the x-axis and the ordered residuals

$\hat{\varepsilon}_{S.i}, \hat{\varepsilon}_{P.i}, \hat{\varepsilon}_{J.i}$ *i.e.* increasingly ordered values of various types of residuals on the y-axis.

The graph of predicted residuals (or Predicted Residuals Graph) has the predicted residuals $\hat{\varepsilon}_{Pi}$ on the x-axis and the ordinary residuals $\hat{\varepsilon}_i$ (Equation (3.4)) on the y-axis. The outlier can easily be detected by their location, as they lie outside the line $y = x$ far from its central pattern (Meloun and Militky [19]).

3. Methodology

Two sets of data (see **Appendix**), were collected and used to build the regression models. The first was data of rainfall (in Millimetres) and yield of Wheat (in kg) and the second was data of agricultural products (Crop production, Livestock, Forestry and fishing) and Nigeria gross domestic product (GDP). The seven analytical and five graphical methods listed above were then applied to detect outliers in the simple and multiple linear regression models.

4. Results and Discussion

4.1. Simple Linear Regression

Simple linear regression of rainfall (Millimetres) on yield of Wheat (kg) was done, using Minitab 17 software to obtain the residuals of the model for outlier detection Methods to be applied.

Regression Analysis: Wheat (kg) versus Rain (Millimetres)

```
The regression equation is
Wheat (kg) = 252 + 4.50 Rain (Millimetres)

Predictor              Coef  SE Coef      T      P
Constant             252.38    14.51  17.39  0.000
Rain (Millimetres)   4.4962   0.5676   7.92  0.000

S = 21.9546   R-Sq = 72.3%   R-Sq(adj) = 71.2%
PRESS = 26511.9   R-Sq(pred) = 36.60%

Analysis of Variance
Source          DF     SS     MS      F      P
Regression       1  30250  30250  62.76  0.000
Residual Error  24  11568    482
Total           25  41818

Unusual Observations
              Rain
Obs    Millimetres)  Wheat (kg)      Fit    SE Fit   Residual    St Resid
  1            12.0      260.00   306.34      8.26     -46.34      -2.28R
 26            50.0      406.00   477.19     15.14     -71.19      -4.48RX

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large influence.
```

4.1.1. Outlier Detection Methods

The seven analytical and five graphical methods for outlier detection discussed in Section 2 were used to detect outliers when the regression model was built using the rainfall and yield of wheat data. The results of the computation using the analytical methods are shown in **Table 1** while **Table 2** gives a summary of the number of outliers detected by each analytical method. **Figures 1(a)-(f)** show the scatter plot of outliers detected by the methods indicated. **Figures 2(a)-(f)** show the box plot of outliers detected by the methods indicated while **Figures 3(a)-(d)** show the Rankit Graph (or Q-Q Plot) of outliers detected.

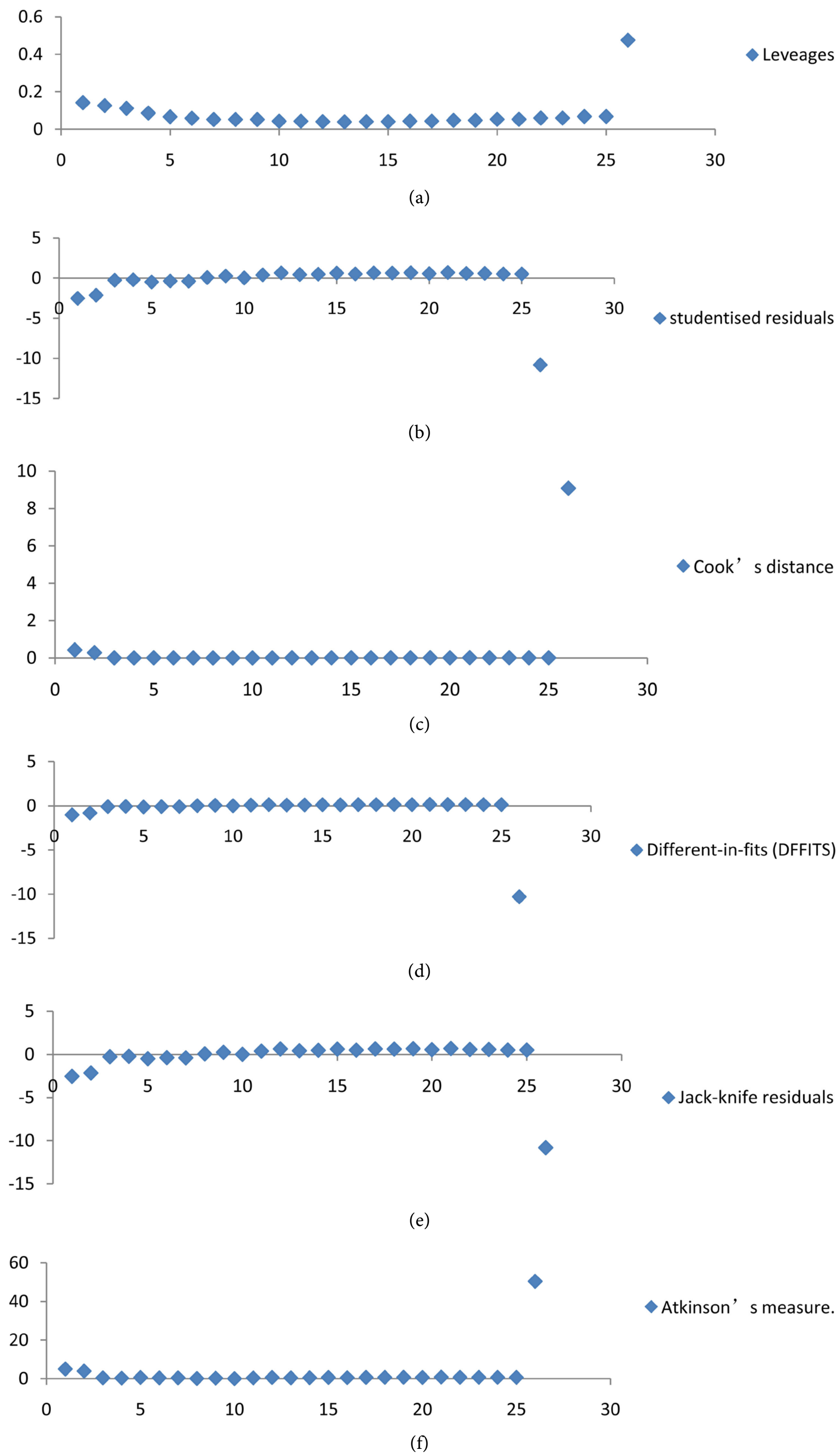
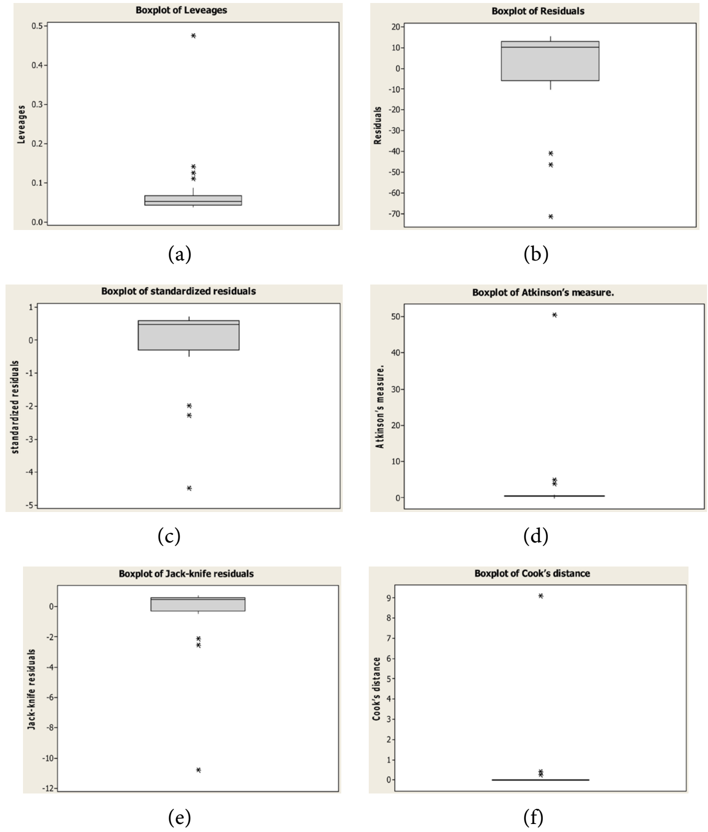


Figure 1. Scatter plots showing outliers detected by each of the analytical methods indicated. (a) Leveages; (b) Studentised residuals; (c) Cook's distance; (d) Different-in-fits (DFFITS); (e) Jack-knife residuals; (f) Atkinson's measure.

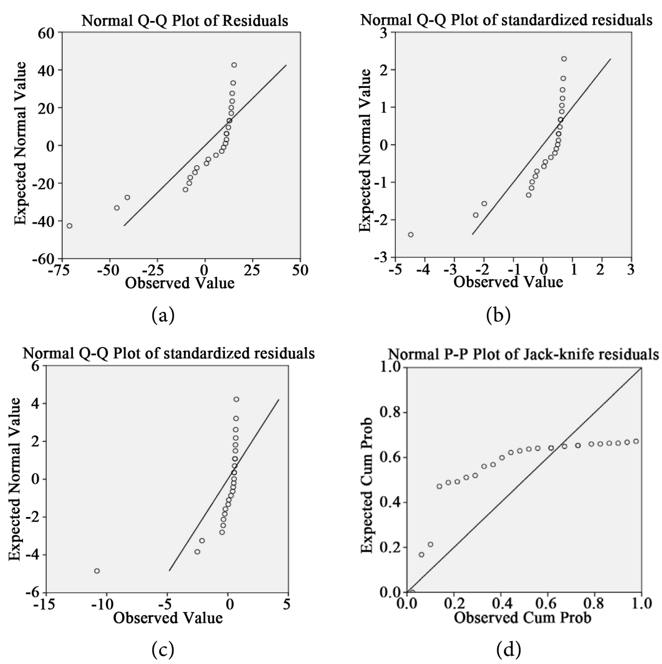


(a) (b)

(c) (d)

(e) (f)

Figure 2. Boxplots showing outliers detected by each of the analytical methods indicated. (a) Leverages; (b) Residuals; (c) Standardized residuals; (d) Atkinson's measure; (e) Jack-knife residual; (f) Cook's distance.



(a) (b)

(c) (d)

Figure 3. Rankit graph (or Q-Q Plot) of outliers detected by each of the analytical methods indicated. (a) Q-Q plot of Residuals; (b) Standardized Residuals; (c) Studentized Residuals; (d) Jack-knife Residuals.

Table 1. Outliers detected using the analytical methods.

Leverages h_i	Residuals ε_i	Standardized residuals $\varepsilon_{ST.i}$	Studentised residuals $\varepsilon_{S.i}$	Predicted residuals $\varepsilon_{P.i}$	Cook's distance D_i	Different-in-fits (DFFITS) DFFit	Jack-knife residuals $\varepsilon_{J.i}$	Atkinson's measure A_i
0.141601	−46.3356	−2.27796	**−2.5189**	−53.979094	**0.42799**	**−1.023**	**−2.518867**	**5.0118657**
0.125665	−40.8318	−1.989	**−2.1306**	−46.700407	**0.2843**	**−0.8077**	**−2.130629**	**3.9571423**
0.111065	−5.328	−0.2574	−0.2523	−5.9936891	0.00414	−0.0892	−0.252329	0.4369446
0.085876	−4.3204	−0.20582	−0.2017	−4.7262735	0.00199	−0.0618	−0.201665	0.3028088
0.066033	−10.3128	−0.48606	−0.4782	−11.041932	0.00835	−0.1271	−0.478185	0.6228977
0.058116	−7.809	−0.3665	−0.3598	−8.2908299	0.00414	−0.0894	−0.359792	0.4378303
0.051536	−8.3052	−0.38843	−0.3815	−8.7564736	0.0041	−0.0889	−0.381453	0.4356032
0.051536	1.6948	0.07926	0.0776	1.78688912	0.00017	0.0181	0.077601	0.0886176
0.051536	5.6948	0.26634	0.2611	6.00423421	0.00193	0.0609	0.261118	0.2981865
0.042385	0.7024	0.03269	0.032	0.73348893	0.00002	0.0067	0.032002	0.0329837
0.042385	8.7024	0.40506	0.3979	9.08757695	0.00363	0.0837	0.397894	0.4100941
0.039815	14.2061	0.66035	0.6524	14.7951697	0.00904	0.1328	0.6524	0.6508272
0.038581	9.7099	0.45106	0.4434	10.0995508	0.00408	0.0888	0.443447	0.4351884
0.040123	10.7175	0.49827	0.4903	11.1654931	0.00519	0.1002	0.490322	0.4911064
0.040123	13.7175	0.63774	0.6297	14.2908935	0.0085	0.1287	0.62967	0.6306783
0.042899	11.2213	0.52244	0.5144	11.724259	0.00612	0.1089	0.514373	0.533493
0.042899	14.2213	0.66212	0.6542	14.8587244	0.00983	0.1385	0.654182	0.678498
0.047012	13.7251	0.64039	0.6323	14.4021751	0.01012	0.1404	0.632332	0.6880368
0.047012	14.7251	0.68705	0.6793	15.4515062	0.01164	0.1509	0.679298	0.7391394
0.052461	12.2289	0.57222	0.564	12.9059595	0.00906	0.1327	0.564033	0.6501741
0.052461	15.2289	0.7126	0.7051	16.0720561	0.01406	0.1659	0.705095	0.8127804
0.059247	12.7327	0.59794	0.5898	13.5345835	0.01126	0.148	0.58976	0.7250636
0.059247	12.7327	0.59794	0.5898	13.5345835	0.01126	0.148	0.58976	0.7250636
0.067369	11.2365	0.52997	0.5219	12.0481734	0.01014	0.1403	0.521874	0.6871421
0.067369	11.2365	0.52997	0.5219	12.0481734	0.01014	0.1403	0.521874	0.6871421
0.475646	−71.1915	**−4.47807**	**−10.8101**	−135.76992	**9.09517**	**−10.2957**	**−10.81004**	**50.438549**

Footnote: $d_i > 3$; $t_{0.05}(22) = 2.074$; $4/n = 4/25 = 0.160$; $2\sqrt{\frac{k}{n}} = 0.566$; $t_{0.05}(22) = 2.074 A_i > 3$.

Table 2. Summary of the results in **Table 1**.

Outlier detection methods	Standardized residuals $\varepsilon_{ST.i}$	Studentised residuals $\varepsilon_{S.i}$	Cook's distance D_i	Different-in-fits (DFFITS) DFFit	Jack-knife residuals $\varepsilon_{J.i}$	Atkinson's measure A_i
Numbers of outliers	1	**3**	**3**	**3**	**3**	**3**

4.1.2. Simple Linear Regression without Outliers (Using Mean Imputation Methods)

Similarly, Simple linear regression of rainfall (Millimetres) on Wheat (kg) was done without the outliers identified (Mean imputation methods). The results of the computation are shown in Table 3 and summary of the number of outliers detected is shown in Table 4.

Regression Analysis: Wheat (kg) versus Rain (Millimetres)

```
The regression equation is
Wheat (kg) = 314 + 2.19 Rain  (Millimetres)


Predictor               Coef  SE Coef      T      P
Constant              314.35    15.32  20.52  0.000
Rain  (Millimetres)   2.1950   0.5989   3.66  0.001


S = 23.1677     R-Sq = 35.9%  R-Sq(adj) = 33.2%
PRESS = 24576.2   R-Sq(pred) = 0.00%


Analysis of Variance

Source          DF       SS      MS      F      P
Regression       1   7209.2  7209.2  13.43  0.001
Residual Error  24  12881.8   536.7
Total           25  20091.0


Unusual Observations
              Rain
Obs  (Millimetres)   Wheat (kg)     Fit  SE Fit  Residual    St Resid
 26           50.0       362.00  424.10   15.98    -62.10     -3.70RX

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large influence.
```

4.1.3. Simple Linear Regression without Outliers (Using Remove Methods)

Also, Simple linear regression of rainfall (Millimetres) on Wheat (kg) done without outliers identified (remove Methods), has results of the computation displayed in Table 5 and summary of the number of outliers detected is shown in Table 6.

Regression Analysis: Wheat (kg) versus Rain (Millimetres)

```
The regression equation is
Wheat (kg) = 225 + 5.92 Rain  (Millimetres)

Predictor               Coef  SE Coef      T      P
Constant             224.715    4.995  44.98  0.000
Rain  (Millimetres)   5.9153   0.2011  29.41  0.000

S = 4.74701     R-Sq = 97.6%   R-Sq(adj) = 97.5%
PRESS = 563.964    R-Sq(pred) = 97.18%

Analysis of Variance
Source          DF     SS     MS       F      P
Regression       1  19497  19497  865.23  0.000
Residual Error  21    473     23
Total           22  19970
```

Table 3. Outliers detected using the analytical methods when the regression model was considered without outlier (Mean imputation method).

Leverages h_i	Residuals ε_i	Standardized residuals $\varepsilon_{ST.i}$	Studentised residuals $\varepsilon_{S.i}$	Predicted residuals $\varepsilon_{P.i}$	Cook's distance D_i	Different-in-fits (DFFITS) DFFit	Jack-knife residuals $\varepsilon_{J.i}$	Atkinson's measure A_i
0.141601	21.3066	0.99263	0.99232	24.82132	0.08127	0.40303	0.992313	1.974436
0.125665	19.1117	0.88222	0.878	21.85856	0.05593	0.33286	0.877999	1.630676
0.111065	−35.0833	−1.60614	−1.66431	−39.4667	0.16116	−0.58829	−1.66431	2.881998
0.085876	−29.4732	−1.33058	−1.35345	−32.242	0.08316	−0.41483	−1.35344	2.032256
0.066033	−30.8631	−1.37845	−1.40624	−33.0452	0.06717	−0.37392	−1.40624	1.83181
0.058116	−26.0581	−1.15894	−1.16768	−27.6659	0.04144	−0.29005	−1.16768	1.420953
0.051536	−24.2531	−1.07491	−1.07856	−25.5709	0.03139	−0.25141	−1.07856	1.231672
0.051536	−14.2531	−0.63171	−0.62361	−15.0276	0.01084	−0.14536	−0.62362	0.712144
0.051536	−10.2531	−0.45442	−0.44678	−10.8102	0.00561	−0.10415	−0.44678	0.510203
0.042385	−10.643	−0.46945	−0.46169	−11.1141	0.00488	−0.09713	−0.46169	0.475847
0.042385	−2.643	−0.11658	−0.11416	−2.75998	0.0003	−0.02402	−0.11416	0.117658
0.039815	5.1621	0.22739	0.22284	5.376151	0.00107	0.04538	0.222842	0.222305
0.038581	2.9671	0.13062	0.12791	3.086167	0.00034	0.02562	0.127915	0.125533
0.040123	8.5772	0.37788	0.37103	8.935728	0.00298	0.07586	0.371029	0.371623
0.040123	11.5772	0.51005	0.50204	12.06113	0.00544	0.10264	0.502039	0.502843
0.042899	11.3822	0.50219	0.49422	11.89237	0.00565	0.10463	0.49422	0.512591
0.042899	14.3822	0.63455	0.62647	15.02684	0.00902	0.13263	0.626467	0.649753
0.047012	16.1873	0.71573	0.70826	16.98584	0.01264	0.15731	0.70826	0.770653
0.047012	17.1873	0.75994	0.75306	18.03517	0.01424	0.16726	0.753055	0.819394
0.052461	16.9923	0.75348	0.7465	17.93309	0.01572	0.17565	0.746498	0.860506
0.052461	19.9923	0.88651	0.88241	21.09918	0.02176	0.20763	0.882412	1.017178
0.059247	19.7973	0.88102	0.87677	21.0441	0.02444	0.22003	0.876765	1.077914
0.059247	19.7973	0.88102	0.87677	21.0441	0.02444	0.22003	0.876765	1.077914
0.067369	20.6024	0.92083	0.9178	22.09062	0.03063	0.24667	0.917801	1.208451
0.067369	20.6024	0.92083	0.9178	22.09062	0.03063	0.24667	0.917801	1.208451
0.475646	−62.1019	**−3.70178**	**−5.53251**	−118.435	**6.21512**	**−5.26929**	**−5.53252**	**25.81417**

Footnote: $d_i > 3$; $t_{0.05}(22) = 2.074$; $4/n = 4/25 = 0.160$; $2\sqrt{\frac{k}{n}} = 0.566$; $t_{0.05}(22) = 2.074 A_i > 3$.

Table 4. Summary of the results in **Table 3**.

Outlier detection methods	Standardized residuals $\varepsilon_{ST.i}$	Studentised residuals $\varepsilon_{S.i}$	Cook's distance D_i	Different-in-fits (DFFITS) DFFit	Jack-knife residuals $\varepsilon_{J.i}$	Atkinson's measure A_i
Numbers of outliers	1	1	1	1	1	1

Table 5. Outliers detected using the analytical methods when the regression model was considered without outlier (Remove method).

Leverages h_i	Residuals ε_i	Standardized residuals $\varepsilon_{ST.i}$	Studentised residuals $\varepsilon_{S.i}$	Predicted residuals $\varepsilon_{P.i}$	Cook's distance D_i	Different-in-fits (DFFITS) DFFit	Jack-knife residuals $\varepsilon_{J.i}$	Atkinson's measure A_i
0.235643	2.47097	0.59539	0.58601	3.232743	0.054642	0.325373	0.586445	1.527277
0.168539	0.64045	0.14796	0.14447	0.770271	0.002219	0.065044	0.14463	0.305421
0.115793	−8.19007	−1.83481	−1.95405	−9.26262	0.220434	−0.70713	−1.94784	**3.306202**
0.094803	−7.10534	−1.57323	−1.63467	−7.8495	0.12961	−0.52902	−1.63157	2.476611
0.077403	−9.0206	−1.97838	−2.14044	−9.7774	0.164186	−0.61998	−2.13181	2.896224
0.077403	0.9794	0.2148	0.20985	1.061569	0.001935	0.060784	0.210082	0.285412
0.077403	4.9794	1.09207	1.09737	5.397156	0.050029	0.317853	1.097113	1.490512
0.053371	−2.85112	−0.61731	−0.60798	−3.01187	0.010743	−0.14436	−0.60841	0.677595
0.053371	5.14888	1.11481	1.12164	5.439174	0.035035	0.266328	1.121311	1.248821
0.046738	9.23361	1.99226	2.15894	9.68633	0.097302	0.478049	**2.150042**	2.232992
0.043695	3.31835	0.71483	0.70625	3.46997	0.011674	0.150966	0.70665	0.708489
0.048377	1.48783	0.32129	0.31432	1.563466	0.002624	0.07087	0.314642	0.332748
0.048377	4.48783	0.96913	0.96766	4.715975	0.023873	0.218178	0.96773	1.023417
0.056102	0.57257	0.12415	0.1212	0.606602	0.000458	0.029548	0.121338	0.138751
0.056102	3.57257	0.77464	0.767	3.784911	0.017833	0.186992	0.767367	0.877488
0.067416	1.6573	0.36152	0.35391	1.777105	0.004724	0.095156	0.354262	0.446759
0.067416	2.6573	0.57966	0.57028	2.849395	0.012145	0.153328	0.570708	0.719718
0.082319	−1.25796	−0.27663	−0.27046	−1.3708	0.003432	−0.081	−0.27074	0.380338
0.082319	1.74204	0.38308	0.37516	1.898307	0.006582	0.112363	0.375527	0.527541
0.100811	−2.17322	−0.48279	−0.47379	−2.41687	0.013066	−0.15864	−0.47421	0.744748
0.100811	−2.17322	−0.48279	−0.47379	−2.41687	0.013066	−0.15864	−0.47421	0.744748
0.122893	−5.08848	−1.14457	−1.15354	−5.80144	0.091776	−0.43179	−1.15311	2.024514
0.122893	−5.08848	−1.14457	−1.15354	−5.80144	0.091776	−0.43179	−1.15311	2.024514
0.067369	11.2365	0.52997	0.5219	12.04817	0.01014	0.1403	0.521122	0.656941

Footnote: $d_i > 3$; $t_{0.05}(22) = 2.074$; $4/n = 4/25 = 0.160$; $2\sqrt{\frac{k}{n}} = 0.566$; $t_{0.05}(22) = 2.074 A_i > 3$.

Table 6. Summary of results in **Table 5**.

Outlier detection methods	Standardized residuals $\varepsilon_{ST.i}$	Studentised residuals $\varepsilon_{S.i}$	Cook's distance D_i	Different-in-fits (DFFITS) DFFit	Jack-knife residuals $\varepsilon_{J.i}$	Atkinson's measure A_i
Numbers of outliers	0	0	0	0	**1**	**1**

4.2. Multiple Linear Regression

Multiple linear regression of agricultural products (Crop production, Livestock, Forestry and fishing) on Nigeria gross domestic product (GDP) was done. The results of the computation are shown in Table 7 while the summary of the number of outliers detected by each of the analytical methods is shown in Table 8.

Regression Analysis: GDP versus Crop production, Livestock, ...

```
The regression equation is
GDP = - 43929 + 0.995 Crop production + 7.58 Livestock + 85.7 Forestry
      + 17.8 Fishing

Predictor           Coef  SE Coef      T      P
Constant          -43929    11705  -3.75  0.000
Crop production   0.9949   0.1088   9.14  0.000
Livestock          7.576    2.345   3.23  0.001
Forestry           85.75    12.41   6.91  0.000
Fishing           17.805    4.279   4.16  0.000


S = 142464   R-Sq = 99.0%   R-Sq(adj) = 99.0%
PRESS = 4.658142E+12   R-Sq(pred) = 98.80%


Analysis of Variance
Source          DF           SS           MS        F      P
Regression       4  3.83524E+14  9.58811E+13  4724.11  0.000
Residual Error 191  3.87656E+12  20296116024
Total          195  3.87401E+14

Source          DF       Seq SS
Crop production  1  3.73368E+14
Livestock        1  7.76702E+12
Forestry         1  2.03799E+12
Fishing          1  3.51362E+11

Unusual Observations
          Crop
Obs  production      GDP      Fit  SE Fit  Residual   St Resid
155      360372   685016   974349   22449   -289334     -2.06R
158      240550   799246  1142086   37531   -342840     -2.49R
165      266058  1164239  1530877   31332   -366638     -2.64R
166      339385  1182576  1593447   24925   -410870     -2.93R
167      390466  1181000  1470805   19006   -289805     -2.05R
169      606640  1625546  2134003   15393   -508456     -3.59R
170      773833  1735803  2287721   21061   -551919     -3.92R
171      890303  1792349  2194706   41125   -402356     -2.95RX
172      779468  1758883  2086168   28505   -327286     -2.34R
173      651425  2039516  2463602   22632   -424086     -3.02R
174      830961  2127893  2345948   98614   -218055     -2.12RX
175      956030  2171579  2501750   38900   -330171     -2.41R
176      837013  2148243  2462496   25306   -314253     -2.24R
177      618328  2831256  2624812   40130    206444      1.51 X
178      811556  2592273  2986711   34694   -394438     -2.85R
179     1062752  2985542  2860760   99670    124781      1.23 X
185      931325  3986280  4010379   52870    -24100     -0.18 X
187     1619235  4986489  4854292   47008    132198      0.98 X
188     1506111  5165742  4775696   61521    390046      3.04RX
189     1087908  4740807  4610574   59946    130233      1.01 X
190     1440102  4853839  5036364   44850   -182524     -1.35 X
191     1889913  5524364  5529255   56499     -4891     -0.04 X
192     1606558  5538295  5293220   58708    245074      1.89 X
193     1148460  5421317  4970468   64858    450848      3.55RX
194     1554121  5704400  5475048   50750    229352      1.72 X
195     2074683  6340199  6288836   58280     51363      0.40 X
196     1767306  6376225  5585943   44522    790283      5.84RX

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large influence.
```

Table 7. Outliers detected using the analytical methods.

Leverages h_i	Residuals ε_i	Standardized residuals $\varepsilon_{ST.i}$	Studentised residuals $\varepsilon_{S.i}$	Predicted residuals $\varepsilon_{P.i}$	Cook's distance D_i	Different-in-fits (DFFITS) DFFit	Jack-knife residuals $\varepsilon_{J.i}$	Atkinson's measure A_i
0.006745	40,628	0.28615	0.28546	40,903.9	0.000111	0.02352	0.864667	0.997557
0.006745	40,792	0.2873	0.28661	41,069.01	0.000112	0.02362	0.868156	1.001582
0.006745	41,167	0.28994	0.28924	41,446.56	0.000114	0.02384	0.876166	1.010822
0.006745	41,305	0.29092	0.29022	41,585.49	0.000115	0.02392	0.879139	1.014253
0.006742	40,791	0.2873	0.2866	41,067.88	0.000112	0.02361	0.868156	1.001358
0.006743	40,883	0.28794	0.28725	41,160.55	0.000113	0.02367	0.870098	1.003672
0.006742	41,312	0.29096	0.29027	41,592.42	0.000115	0.02392	0.87926	1.014166
0.006743	41,445	0.2919	0.2912	41,726.36	0.000116	0.02399	0.882113	1.017531
0.006742	40,801	0.28737	0.28667	41,077.95	0.000112	0.02362	0.868369	1.001603
0.006742	40,797	0.28734	0.28664	41,073.92	0.000112	0.02362	0.868278	1.001498
0.006742	41,325	0.29106	0.29036	41,605.5	0.000115	0.02392	0.879564	1.014516
0.006742	41,464	0.29204	0.29134	41,745.45	0.000116	0.024	0.882537	1.017945
0.006746	40,290	0.28377	0.28308	40,563.64	0.000109	0.02333	0.857448	0.989302
0.006742	40,337	0.28409	0.28341	40,610.8	0.00011	0.02335	0.858419	0.990126
0.006746	40,900	0.28806	0.28737	41,177.79	0.000113	0.02368	0.870462	1.004317
0.006746	41,062	0.28921	0.28851	41,340.89	0.000114	0.02378	0.873951	1.008342
0.006744	40,368	0.28431	0.28363	40,642.09	0.00011	0.02337	0.859086	0.991044
0.006745	40,469	0.28503	0.28434	40,743.82	0.00011	0.02343	0.86127	0.993637
0.006744	40,978	0.28861	0.28792	41,256.23	0.000113	0.02372	0.87213	1.006092
0.006744	41,113	0.28956	0.28887	41,392.15	0.000114	0.0238	0.875013	1.009417
0.006745	40,131	0.28264	0.28196	40,403.52	0.000108	0.02323	0.85402	0.985274
0.006745	40,250	0.28348	0.2828	40,523.33	0.000109	0.0233	0.856568	0.988213
0.006744	40,771	0.28715	0.28646	41,047.83	0.000112	0.0236	0.867701	1.000982
0.006744	40,947	0.28839	0.2877	41,225.02	0.000113	0.02371	0.871463	1.005322
0.006745	40,289	0.28376	0.28307	40,562.59	0.000109	0.02333	0.857418	0.989193
0.006746	40,398	0.28453	0.28384	40,672.38	0.00011	0.02339	0.859753	0.991962
0.006745	40,913	0.28815	0.28746	41,190.83	0.000113	0.02369	0.870735	1.004557
0.006746	41,070	0.28926	0.28857	41,348.94	0.000114	0.02378	0.874102	1.008517
0.006743	41,763	0.29414	0.29344	42,046.52	0.000117	0.02418	0.88891	1.025372
0.006744	41,819	0.29454	0.29383	42,102.94	0.000118	0.02421	0.890123	1.026849
0.006744	42,137	0.29677	0.29606	42,423.1	0.00012	0.0244	0.896891	1.034656
0.006745	42,216	0.29733	0.29662	42,502.68	0.00012	0.02444	0.898591	1.036694

Continued

0.006744	41,738	0.29396	0.29326	42,021.39	0.000117	0.02416	0.888363	1.024818
0.006745	41,796	0.29437	0.29367	42,079.83	0.000118	0.0242	0.889608	1.02633
0.006744	42,122	0.29667	0.29596	42,408	0.00012	0.02439	0.896587	1.034306
0.006745	42,201	0.29722	0.29651	42,487.58	0.00012	0.02443	0.898257	1.036309
0.006744	41,673	0.29351	0.2928	41,955.95	0.000117	0.02413	0.886998	1.023243
0.006745	41,737	0.29395	0.29325	42,020.43	0.000117	0.02417	0.888333	1.02486
0.006745	42,080	0.29638	0.29567	42,365.76	0.000119	0.02436	0.895707	1.033367
0.006745	42,187	0.29713	0.29642	42,473.48	0.00012	0.02443	0.897984	1.035993
0.006755	39,592	0.27885	0.27817	39,861.26	0.000106	0.02294	0.842526	0.972738
0.006756	39,684	0.2795	0.27882	39,953.93	0.000106	0.023	0.844497	0.975086
0.006753	40,354	0.28422	0.28353	40,628.36	0.00011	0.02338	0.858813	0.991394
0.006755	40,433	0.28477	0.28409	40,707.98	0.00011	0.02343	0.860481	0.993468
0.006759	39,000	0.27468	0.27402	39,265.39	0.000103	0.0226	0.829881	0.958424
0.00676	39,122	0.27554	0.27487	39,388.26	0.000103	0.02268	0.832488	0.961507
0.006757	40,008	0.28178	0.2811	40,280.17	0.000108	0.02318	0.851412	0.983144
0.006759	40,081	0.2823	0.28161	40,353.75	0.000108	0.02323	0.852989	0.985112
0.006762	38,009	0.2677	0.26705	38,267.77	0.000098	0.02203	0.808719	0.934193
0.006763	37,945	0.26725	0.2666	38,203.37	0.000097	0.022	0.807355	0.932687
0.006757	39,036	0.27494	0.27427	39,301.56	0.000103	0.02262	0.830669	0.959191
0.00676	39,092	0.27533	0.27466	39,358.06	0.000103	0.02266	0.831852	0.960772
0.006754	37,078	0.26114	0.2605	37,330.13	0.000093	0.02148	0.788836	0.910682
0.006755	37,257	0.26241	0.26177	37,510.38	0.000094	0.02159	0.792685	0.915194
0.006748	38,477	0.271	0.27034	38,738.41	0.0001	0.02228	0.818723	0.944763
0.006752	38,514	0.27126	0.2706	38,775.81	0.0001	0.02231	0.819511	0.945955
0.006725	36,796	0.25915	0.25852	37,045.13	0.000091	0.02127	0.782805	0.901765
0.006729	39,052	0.27504	0.27438	39,316.56	0.000102	0.02258	0.830972	0.957538
0.006718	40,338	0.2841	0.28341	40,610.82	0.000109	0.02331	0.858449	0.988385
0.006722	40,434	0.28478	0.28409	40,707.64	0.00011	0.02337	0.860512	0.991057
0.006756	33,988	0.23938	0.23879	34,219.18	0.000078	0.01969	0.722917	0.834705
0.00676	34,268	0.24135	0.24076	34,501.23	0.000079	0.01986	0.728882	0.841844
0.006739	36,267	0.25543	0.2548	36,513.06	0.000089	0.02099	0.771533	0.88971
0.006755	36,047	0.25389	0.25326	36,292.15	0.000088	0.02089	0.766867	0.885386
0.006687	36,451	0.25672	0.25609	36,696.39	0.000089	0.02101	0.775442	0.890738
0.006688	36,579	0.25762	0.25699	36,825.29	0.000089	0.02109	0.778169	0.893938

Continued

0.006682	38,170	0.26883	0.26818	38,426.77	0.000097	0.022	0.812145	0.932546
0.00669	38,131	0.26855	0.2679	38,387.81	0.000097	0.02199	0.811296	0.932133
0.006661	36,637	0.25803	0.2574	36,882.68	0.000089	0.02108	0.779411	0.893544
0.006663	37,559	0.26452	0.26387	37,810.93	0.000094	0.02161	0.79908	0.916231
0.00666	39,104	0.2754	0.27473	39,366.18	0.000102	0.0225	0.832064	0.953834
0.006671	38,900	0.27397	0.2733	39,161.24	0.000101	0.0224	0.827728	0.949652
0.006726	34,804	0.24512	0.24452	35,039.68	0.000081	0.02012	0.7403	0.852864
0.006733	34,986	0.24641	0.2458	35,223.16	0.000082	0.02024	0.744208	0.857815
0.006704	37,329	0.26291	0.26227	37,580.94	0.000093	0.02155	0.7942	0.913452
0.006738	36,706	0.25852	0.25789	36,955	0.000091	0.02124	0.780896	0.90044
0.006792	32,899	0.23171	0.23114	33,123.98	0.000073	0.01911	0.699693	0.810055
0.006802	33,840	0.23835	0.23776	34,071.76	0.000078	0.01968	0.719798	0.833948
0.006744	27,684	0.19498	0.19449	27,871.97	0.000052	0.01603	0.58856	0.678964
0.006793	36,093	0.25421	0.25359	36,339.86	0.000088	0.02097	0.767837	0.889013
0.006629	36,810	0.25924	0.2586	37,055.64	0.00009	0.02112	0.783078	0.895574
0.006634	37,542	0.2644	0.26376	37,792.72	0.000093	0.02155	0.798716	0.913805
0.006622	39,906	0.28105	0.28037	40,172.02	0.000105	0.02289	0.849198	0.970676
0.006645	39,487	0.2781	0.27742	39,751.15	0.000103	0.02269	0.840251	0.962127
0.006953	23,171	0.16321	0.16279	23,333.24	0.000037	0.01362	0.492526	0.576977
0.006927	21,289	0.14995	0.14957	21,437.5	0.000031	0.01249	0.452466	0.529049
0.006845	25,843	0.18203	0.18156	26,021.11	0.000046	0.01507	0.549405	0.638556
0.006827	26,589	0.18728	0.1868	26,771.77	0.000048	0.01549	0.565277	0.656133
0.006765	19,580	0.13791	0.13755	19,713.36	0.000026	0.01135	0.416102	0.480768
0.006743	18,337	0.12915	0.12882	18,461.49	0.000023	0.01061	0.389649	0.449467
0.006684	23,331	0.16431	0.16389	23,487.99	0.000036	0.01344	0.49585	0.569446
0.006671	24,251	0.17079	0.17036	24,413.86	0.000039	0.01396	0.515432	0.591355
0.00667	13,968	0.09838	0.09812	14,061.79	0.000013	0.00804	0.296766	0.340454
0.006646	12,997	0.09153	0.0913	13,083.96	0.000011	0.00747	0.276094	0.316165
0.006593	18,912	0.13319	0.13285	19,037.51	0.000024	0.01082	0.401848	0.458319
0.006586	19,856	0.13983	0.13947	19,987.64	0.000026	0.01136	0.4219	0.480932
0.006619	10,696	0.07533	0.07513	10,767.27	0.000008	0.00613	0.227213	0.259657
0.00659	10,074	0.07094	0.07076	10,140.83	0.000007	0.00576	0.213969	0.243981
0.006538	16,118	0.11351	0.11321	16,224.07	0.000017	0.00918	0.342432	0.38891
0.006522	17,562	0.12367	0.12335	17,677.29	0.00002	0.00999	0.373104	0.423223

Continued

0.006703	11,014	0.07757	0.07737	11,088.33	0.000008	0.00636	0.233972	0.269083
0.006669	10,657	0.07506	0.07486	10,728.55	0.000008	0.00613	0.226399	0.259708
0.006593	16,785	0.11821	0.1179	16,896.4	0.000019	0.00961	0.35662	0.406735
0.006567	19,595	0.13799	0.13764	19,724.53	0.000025	0.01119	0.416343	0.473908
0.006717	7039	0.04958	0.04945	7086.601	0.000003	0.00407	0.149534	0.172154
0.006675	8335	0.0587	0.05855	8391.01	0.000005	0.0048	0.177044	0.203184
0.006593	13,533	0.09531	0.09506	13,622.82	0.000012	0.00774	0.287501	0.327903
0.006564	16,353	0.11517	0.11487	16,461.05	0.000018	0.00934	0.347443	0.39539
0.006679	13,040	0.09184	0.0916	13,127.68	0.000011	0.00751	0.27703	0.318028
0.00665	12,586	0.08864	0.08841	12,670.26	0.000011	0.00723	0.267373	0.306271
0.006719	13,070	0.09205	0.09181	13,158.41	0.000011	0.00755	0.277663	0.319715
0.006545	21,679	0.15267	0.15228	21,821.82	0.000031	0.01236	0.460683	0.523493
0.006759	7056	0.0497	0.04957	7104.016	0.000003	0.00409	0.149896	0.173113
0.006719	8117	0.05717	0.05702	8171.907	0.000004	0.00469	0.172428	0.198542
0.006647	14,072	0.09911	0.09885	14,166.16	0.000013	0.00809	0.298969	0.342385
0.006587	18,224	0.12835	0.12801	18,344.84	0.000022	0.01042	0.387234	0.441449
0.006523	8069	0.05682	0.05668	8121.98	0.000004	0.00459	0.171373	0.194408
0.006631	12,589	0.08866	0.08843	12,673.03	0.00001	0.00723	0.267434	0.305899
0.006401	14,479	0.10196	0.10169	14,572.28	0.000013	0.00816	0.30757	0.345614
0.016748	−110,431	−0.78172	−0.78092	−112,312	0.002082	−0.10192	−2.3926	**4.371676**
0.006547	18,806	0.13244	0.1321	18,929.93	0.000023	0.01072	0.399584	0.454133
0.006306	5064	0.03566	0.03556	5096.136	0.000002	0.00283	0.107548	0.119944
0.00625	13,374	0.09417	0.09393	13,458.11	0.000011	0.00745	0.284061	0.315385
0.00625	19,343	0.1362	0.13585	19,464.65	0.000023	0.01077	0.410938	0.456253
0.006993	15,273	0.10758	0.10731	15,380.56	0.000016	0.00901	0.324533	0.381278
0.006223	4129	0.02907	0.029	4154.856	0.000001	0.00229	0.087672	0.097128
0.006183	15,727	0.11074	0.11045	15,824.85	0.000015	0.00871	0.334071	0.368904
0.006192	21,843	0.1538	0.15341	21,979.09	0.000029	0.01211	0.464096	0.512862
0.006049	29,116	0.205	0.20448	29,293.19	0.000051	0.01595	0.618865	0.675901
0.005951	23,060	0.16235	0.16194	23,198.05	0.000032	0.01253	0.489928	0.530703
0.005997	31,983	0.22518	0.22462	32,175.96	0.000061	0.01745	0.679926	0.739372
0.005974	40,856	0.28764	0.28695	41,101.54	0.000099	0.02225	0.869188	0.943355
0.006387	−5515	−0.03884	−0.03873	−5550.45	0.000002	−0.00311	−0.11714	0.131483
0.006213	−2138	−0.01505	−0.01502	−2151.37	0	−0.00119	−0.04539	0.050243

Continued

0.006577	10,690	0.07529	0.07509	10,760.77	0.000008	0.00611	0.227093	0.258689
0.006209	23,834	0.16782	0.16739	23,982.91	0.000035	0.01323	0.506457	0.560446
0.007934	−52,714	−0.37149	−0.37065	−53,135.6	0.000221	−0.03315	−1.12405	1.40731
0.007447	−52,923	−0.37287	−0.37203	−53,320.1	0.000209	−0.03222	−1.12825	1.368197
0.008159	−31,286	−0.22051	−0.21996	−31,543.4	0.00008	−0.01995	−0.66579	0.845404
0.007124	−9704	−0.06836	−0.06818	−9773.63	0.000007	−0.00578	−0.20619	0.244512
0.016214	16,153	0.11431	0.11401	16,419.22	0.000043	0.01464	0.344847	0.619797
0.015191	−1697	−0.012	−0.01197	−1723.18	0	−0.00149	−0.03619	0.062927
0.029804	176,446	1.25741	1.25933	181,866.3	0.009714	0.22072	**3.943493**	**9.676463**
0.014671	59,006	0.41725	0.41635	59,884.57	0.000518	0.0508	1.263604	2.158631
0.027189	33,348	0.23733	0.23674	34,280.04	0.000315	0.03958	0.716709	1.677464
0.025367	1771	0.01259	0.01256	1817.094	0.000001	0.00203	0.037969	0.085758
0.027198	37,345	0.26577	0.26513	38,389.11	0.000395	0.04433	0.802869	1.879442
0.024538	79,062	0.5619	0.56089	81,050.83	0.001588	0.08896	**1.70748**	**3.791387**
0.028664	−65,898	−0.46933	−0.46837	−67,842.6	0.0013	−0.08046	−1.4229	**3.422053**
0.025857	−92,206	−0.65575	−0.65477	−94,653.5	0.002283	−0.10668	−1.9982	**4.557678**
0.027779	−39,633	−0.28214	−0.28146	−40,765.4	0.000455	−0.04758	−0.8525	2.017438
0.025337	1024	0.00728	0.00726	1050.62	0	0.00117	0.021955	0.049558
0.026983	−247,429	−1.7607	**−1.77051**	−254,291	0.017194	**−0.29484**	**−5.75144**	**13.40878**
0.020807	−257,930	−1.82962	**−1.84103**	−263,411	0.014227	−0.26837	**−6.01831**	**12.28212**
0.02483	−289,334	−2.05661	**−2.07432**	−296,701	**0.021539**	**−0.33099**	**−6.94061**	**15.50507**
0.021941	53,796	0.38182	0.38097	55,002.82	0.000654	0.05706	1.155523	2.422993
0.024957	−243,028	−1.72758	−1.73668	−249,248	0.015278	−0.27784	**−5.62512**	**12.59924**
0.069402	−342,840	−2.49462	**−2.52963**	−368,408	**0.092822**	**−0.69082**	**−8.96871**	**34.28963**
0.014566	−185,919	−1.31463	−1.31716	−188,667	0.005109	−0.16014	**−4.13866**	**7.044401**
0.017882	−129,235	−0.91536	−0.91497	−131,588	0.003051	−0.12346	**−2.81735**	**5.322244**
0.031035	−5145	−0.03669	−0.03659	−5309.79	0.000009	−0.00655	−0.11065	0.277248
0.01901	−71,030	−0.50339	−0.5024	−72,406.4	0.000982	−0.06994	−1.52738	2.976697
0.01257	5073	0.03584	0.03574	5137.579	0.000003	0.00403	0.108091	0.170738
0.011602	66,459	0.46923	0.46827	67,239.11	0.000517	0.05073	1.422596	2.157794
0.048367	−366,638	−2.63813	**−2.6805**	−385,272	**0.070746**	**−0.60431**	**−9.73028**	**30.71093**
0.030608	−410,870	−2.9292	**−2.98944**	−423,843	**0.054184**	**−0.5312**	**−11.4871**	**28.5762**
0.017798	−289,805	−2.05258	**−2.07015**	−295,056	0.015269	−0.27867	**−6.92359**	**13.04802**
0.02339	−226,547	−1.60913	−1.6159	−231,973	0.012403	−0.25007	**−5.1829**	**11.22937**

Continued

0.011674	−508,456	**−3.59002**	**−3.7079**	−514,462	**0.030446**	**−0.40298**	**−17.4203**	**26.50598**
0.021854	−551,919	**−3.91712**	**−4.07391**	−564,250	**0.068564**	**−0.60895**	**−22.7626**	**47.63366**
0.083329	−402,356	−2.94983	**−3.0115**	−438,932	0.1582	**−0.90797**	**−11.6249**	**49.06897**
0.040035	−327,286	−2.34473	**−2.37298**	−340,935	**0.045856**	**−0.4846**	**−8.23024**	**23.53057**
0.025237	−424,086	**−3.01508**	**−3.0814**	−435,066	**0.047072**	**−0.49581**	**−12.0746**	**27.20006**
0.479139	−218,055	−2.12079	**−2.14059**	−418,643	**0.827497**	**−2.05307**	**−7.21511**	**96.88145**
0.074555	−330,171	−2.40911	**−2.44015**	−356,770	**0.093512**	**−0.6926**	**−8.54095**	**33.93885**
0.031553	−314,253	−2.24148	**−2.2656**	−324,492	**0.032738**	**−0.40894**	**−7.75034**	**19.58534**
0.079348	206,444	1.51025	**1.51536**	224,236.7	**0.039316**	**0.44487**	**4.824169**	**19.82763**
0.059306	−394,438	−2.85462	**−2.90989**	−419,305	0.102749	**−0.73064**	**−11.0051**	**38.68541**
0.489461	124,781	1.22583	1.22745	244,410.3	**0.288122**	**1.20185**	**3.836712**	**52.59347**
0.059191	−169,428	−1.2261	−1.22773	−180,088	0.018916	**−0.30795**	**−3.83762**	**13.4762**
0.072177	−64,078	−0.46695	−0.466	−69,062.7	0.003392	−0.12997	−1.41561	**5.527627**
0.032944	−172,907	−1.23419	−1.23589	−178,797	0.010378	−0.22811	**−3.86492**	**9.986897**
0.06561	−87,542	−0.63569	−0.6347	−93,688.9	0.005675	−0.16819	**−1.93585**	**7.181591**
0.036653	173,358	1.23979	1.24154	179,953.8	0.011696	0.24217	**3.883835**	**10.60601**
0.137725	−24,100	−0.18217	−0.18171	−27,949.3	0.00106	−0.07262	−0.54983	**3.076372**
0.052288	134,008	0.96624	0.96607	141,401.6	0.010302	0.22692	**2.981036**	**9.802978**
0.108874	132,198	0.98299	0.9829	148,349.4	**0.023611**	**0.34356**	**3.035183**	**14.8527**
0.186478	390,046	**3.03546**	**3.10328**	479,453.5	**0.422413**	**1.48577**	**12.21964**	**81.90596**
0.177056	130,233	1.00769	1.00774	158,252.6	**0.043695**	**0.46743**	**3.115275**	**20.22994**
0.09911	−182,524	−1.34983	−1.35276	−202,604	0.040089	−0.44869	**−4.25985**	**19.78087**
0.15728	−4891	−0.0374	−0.0373	−5803.83	0.000052	−0.01612	−0.1128	0.682207
0.169819	245,074	1.88801	**1.90089**	295,205.5	**0.145832**	**0.85973**	**6.248917**	**39.56759**
0.20726	450,848	3.55434	**3.66842**	568,721.1	**0.66059**	**1.87573**	**16.98241**	**121.5681**
0.126898	229,352	1.72292	**1.73191**	262,686.4	**0.086287**	**0.66027**	**5.607449**	**29.92873**
0.167351	51,363	0.39511	0.39423	61,686.26	0.006275	0.17674	1.196039	**7.506829**
0.097666	790,283	**5.83972**	**6.4263**	875,820.9	**0.738231**	**2.11422**	**#NUM!**	**#NUM!**

Footnote: $d_i > 3$; $t_{0.05}(193) = 1.646$; $4/n = 4/193 = 0.0204$; $2\sqrt{\frac{k}{n}} = 0.286$; $t_{0.05}(193) = 1.646$; $A_i > 3$.

Table 8. Summary of results in Table 7.

Outlier detection methods	Standardized residuals $\varepsilon_{ST.i}$	Studentised residuals $\varepsilon_{S.i}$	Cook's distance D_i	Different-in-fits (DFFITS) DFFit	Jack-knife residuals $\varepsilon_{J.i}$	Atkinson's measure A_i
Numbers of outliers	5	20	21	25	39	42

Graphical Methods

Figures 4-10 show the results of graphical methods.

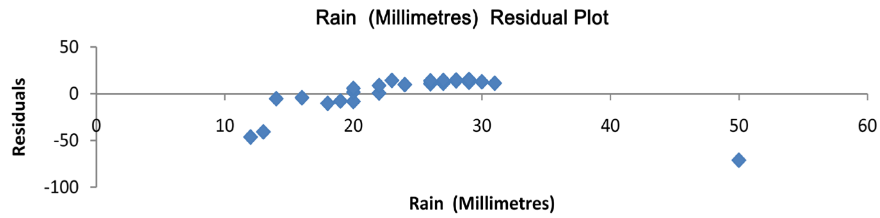


Figure 4. Graph of the predicted residuals.

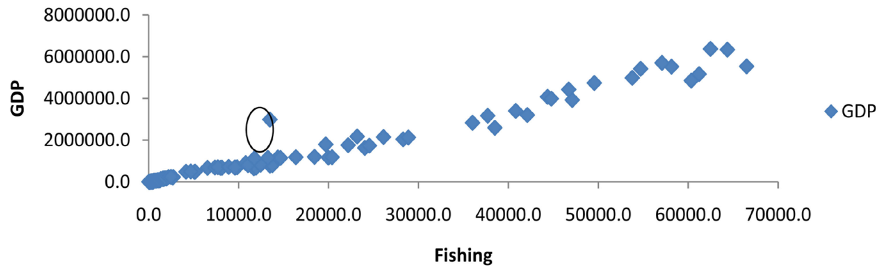


Figure 5. Scatter plot of Fishing against GDP.

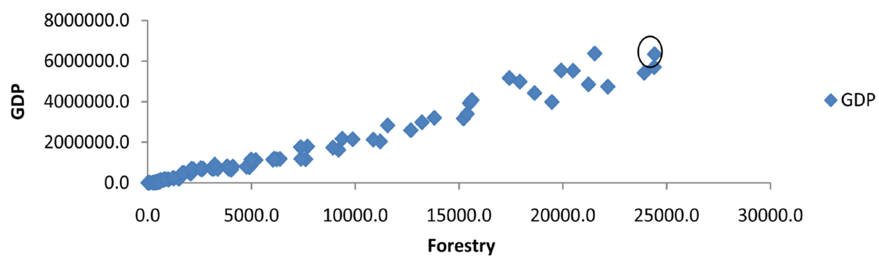


Figure 6. Forestry against GDP.

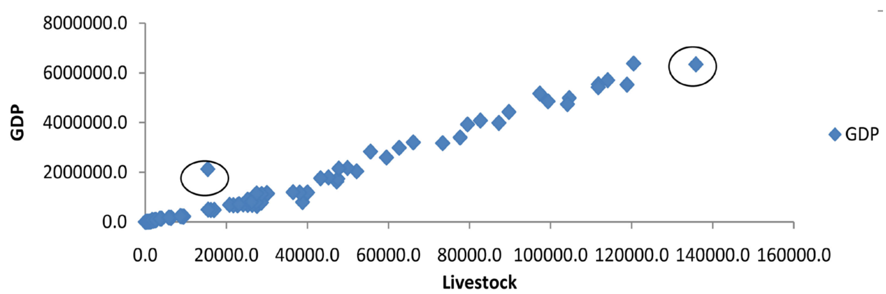


Figure 7. Livestock against GDP.

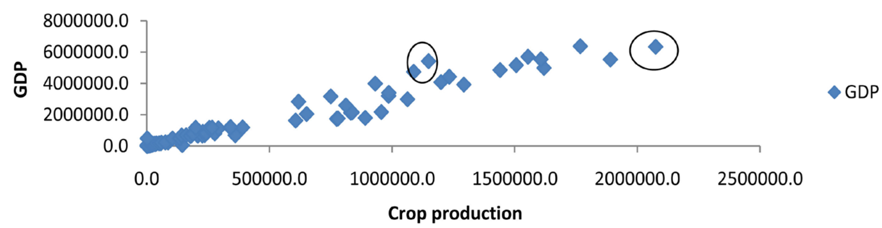


Figure 8. Crop production against GDP.

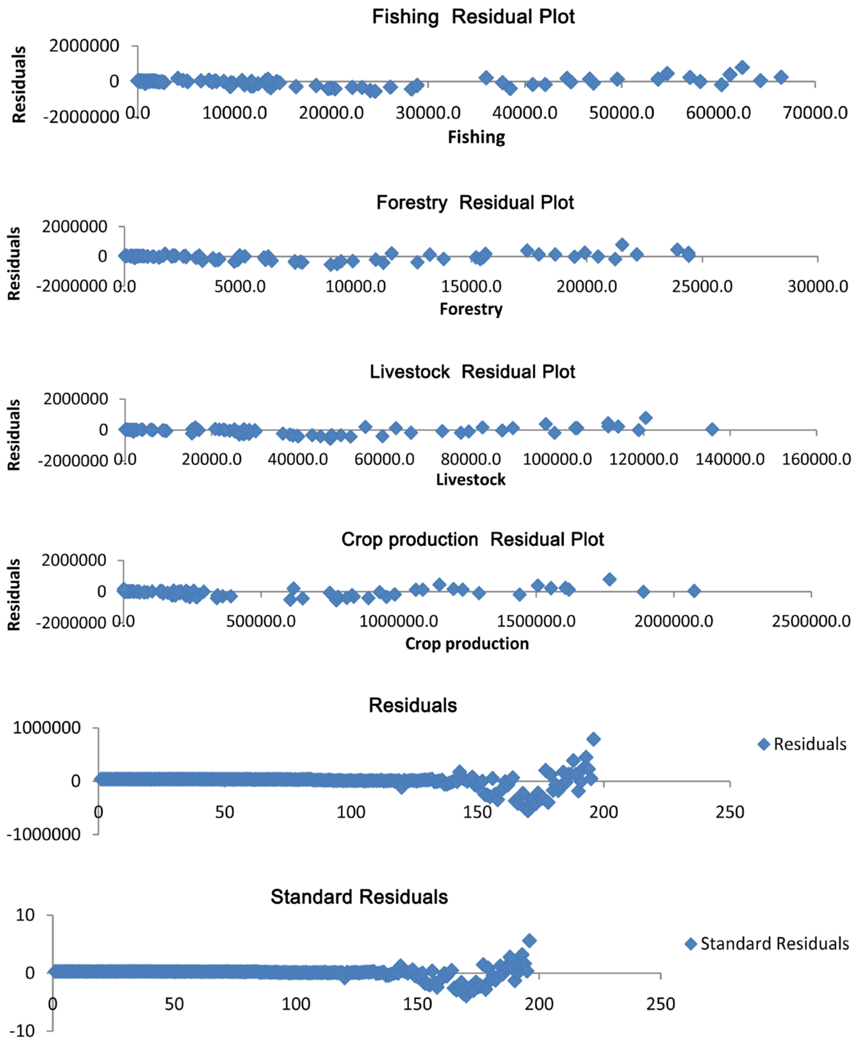


Figure 9. Graphs of predicted residuals.

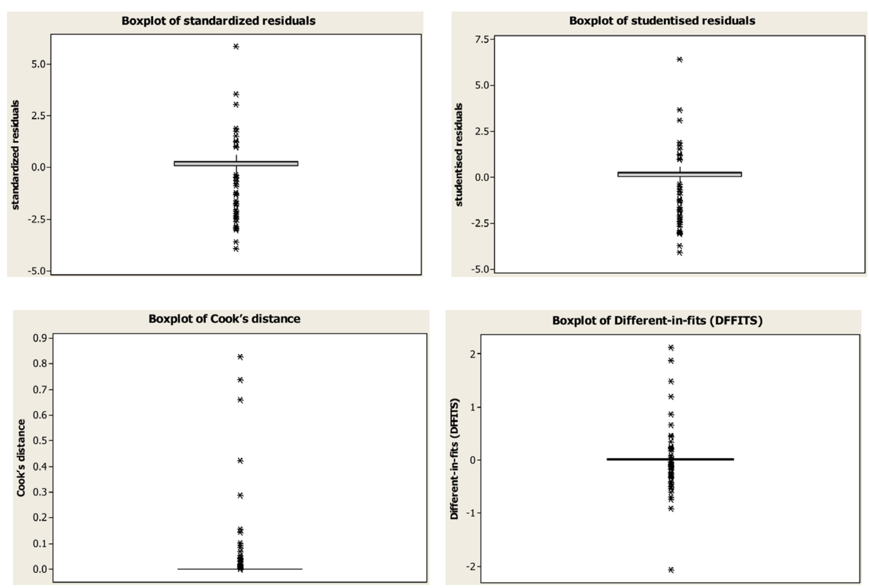


Figure 10. Boxplots of outliers detected by the analytical methods indicated.

5. Conclusion

Outliers detection and effects on simple and multiple linear regression modeling were studied using the above listed analytical and graphical methods. Two data sets were used for the illustration. From the results obtained, we concluded that by removing the influential point (or Outliers), the model adequacy increased (from $R^2 = 0.72$ to $R^2 = 0.97$). Also, Jack-knife residuals and Atkinson's measure methods are more useful for detecting outliers.

Acknowledgements

We thank the National Bureau of Statistics and the Central Bank of Nigeria (CBN) for granting us access to their statistical bulletins from which the data used were extracted.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Bollen, K.A. and Jackman, R.W. (1990) Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases. In: Fox, J. and Scott, L.J., Eds., *Modern Methods of Data Analysis*, Sage, Newbury Park, 257-291.

[2] Hawkins, D.M. (1983) Discussion of Paper by Beckman and Cook. *Technometrics*, **25**, 155-156. https://doi.org/10.1080/00401706.1983.10487843

[3] Dixon, W.J. (1950) Analysis of Extreme Values. *The Annals of Mathematical Statistics*, **21**, 488-506. https://doi.org/10.1214/aoms/1177729747

[4] Barnett, V. and Lewis, T. (1994) Outliers in Statistical Data. John Wiley, New York.

[5] Montgomery, D.C., Peck, E.A. and Vining, G.G. (2013) Introduction to Linear Regression Analysis. John Wiley and Sons, Hoboken.

[6] Abuzaid, A.H., Hussin, A.G. and Mohamed, I.B. (2008) Identifying Single Outlier in Linear Circular Regression Model Based on Circular Distance. *Journal of Applied Probability and Statistics*, **3**, 107-117.

[7] Zhang, Y., Meratnia, N. and Havinga, P.J.M. (2010) Outlier Detection Techniques for Wireless Sensor Networks Survey. *IEEE Communication Survey and Tutorial*, **12**, 159-170. https://doi.org/10.1109/SURV.2010.021510.00088

[8] Rousseeuw, P.J. (1984) Least Median of Squares Regression. *Journal of the American Statistical Association*, **79**, 871-880. https://doi.org/10.1080/01621459.1984.10477105

[9] Aggarwal, C.C. and Yu, P.S. (2013) Outlier Detection for High Dimensional Data. https://www.researchgate.net/publication/2401320_Outlier_Detection_for_High_Dimensional_Data

[10] Arning, A., Agrawal, R. and Raghavan, P. (1996) A Linear Method for Deviation Detection in Large Databases. *KDD*-1996, Portland, 2-4 August 1996, 164-169.

[11] Hodge, V.J. and Austin, J. (2018) An Evaluation of Classification and Outlier Detection Algorithms. https://arxiv.org/pdf/1805.00811.pdf

[12] Sebert, D.M., Montgomery, D.C. and Rollier, D.A. (1998) Clustering Algorithm for

Identifying Multiple Outliers in Linear Regression. *Computational Statistics and Data Analysis*, **27**, 461-484. https://doi.org/10.1016/S0167-9473(98)00021-8

[13] Worden, K., Manson, G. and Fieller, N.R.J. (2000) Damage Detection Using Outlier Analysis. *Journal of Sound and Vibration*, **229**, 647-667. https://doi.org/10.1006/jsvi.1999.2514

[14] Kitagawa, G. (1984) Bayesian Analysis of Outliers via Akaike's Predictive Likelihood of a Model. *Communication Statistics—Simulation Computation*, **13**, 107-126. https://doi.org/10.1080/03610918408812361

[15] Fung, W.-K. and Bacon-Shone, J. (1993) Quasi-Bayesian Modelling of Multivariate Outliers. *Computational Statistics and Data Analysis*, **16**, 271-278. https://doi.org/10.1016/0167-9473(93)90129-H

[16] Belsey, D.A., Kuh, E. and Welsch, R.E. (1980) Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley, Hoboken. https://doi.org/10.1002/0471725153

[17] Cook, R.D. (1977) Detection of Influential Observation in Linear Regression. *Technometrics*, **19**, 15-18. https://doi.org/10.1080/00401706.1977.10489493

[18] Williams, D.X. (1973) Letter to the Editor. *Applied Statistics*, **22**, 407-408. https://doi.org/10.1111/j.1467-9876.1973.tb00281.x

[19] Meloun, M. and Militky, J. (2001) Detection of Single Influential Points in OLS Regression Model Building. *Analytica Chimica Acta*, **439**, 169-191. https://doi.org/10.1016/S0003-2670(01)01040-6

Appendix

Data 1.

Rain (mm)	Wheat (kg)	Rain (mm)	Wheat (kg)	Rain (mm)	Wheat (kg)	Rain (mm)	Wheat (kg)	Rain (mm)	Wheat (kg)	Rain (mm)	Wheat (kg)
12	260	19	330	22	360	27	385	29	398	50	406
13	270	20	334	23	370	27	388	30	400		
14	310	20	344	24	370	28	392	30	400		
16	320	20	348	26	380	28	393	31	403		
18	323	22	352	26	383	29	395	31	403		

Data 2.

S/no.	Year	Crop production	Livestock	Forestry	Fishing	GDP
1	1961	224.4	31.0	37.0	9.4	497.7
2		288.2	31.2	35.0	9.6	558.7
3		329.3	29.7	31.0	7.7	586.0
4		288.3	28.4	29.7	8.7	580.0
5	1962	229.2	35.2	34.6	11.4	527.1
6		292.4	35.4	33.6	11.6	601.5
7		336.4	33.7	28.9	9.3	617.2
8		294.5	32.2	27.7	10.5	615.5
9	1963	257.3	37.0	34.5	11.7	575.5
10		328.2	37.2	33.5	11.9	561.0
11		377.6	35.4	28.8	9.6	680.7
12		330.6	33.9	27.6	10.8	680.3
13	1964	268.6	33.8	41.0	11.7	608.7
14		342.6	43.0	39.8	12.0	700.9
15		394.2	32.4	34.3	9.6	720.9
16		345.1	31.0	32.9	10.8	725.4
17	1965	262.4	38.1	38.9	18.2	648.6
18		334.8	38.3	37.8	18.6	736.2
19		385.2	36.5	32.5	14.9	760.9
20		337.2	34.9	31.2	16.8	758.7
21	1966	260.1	40.1	41.6	20.7	700.4
22		331.8	40.3	40.3	21.2	789.9
23		381.8	38.4	34.8	17.0	799.8
24		334.3	36.1	33.3	19.1	820.0
25	1967	288.9	40.1	38.4	28.9	758.4
26		368.5	40.3	37.2	29.5	856.4

Continued

27	1968	423.9	38.4	32.1	23.6	869.4
28		371.2	36.1	30.7	26.6	890.5
29		239.8	34.8	20.3	30.2	614.9
30		305.9	34.9	19.7	30.9	698.8
31		351.9	33.3	17.0	24.8	710.0
32		308.1	31.8	16.3	27.9	728.9
33	1969	217.6	33.8	20.6	30.7	594.5
34		277.6	34.0	20.0	31.4	675.2
35		319.4	32.4	17.2	25.2	680.3
36		279.6	31.0	16.5	28.4	705.6
37	1970	269.8	37.3	22.0	34.9	803.2
38		344.2	37.6	21.3	35.8	899.1
39		396.0	35.7	18.4	28.6	903.1
40		346.7	34.1	17.6	32.2	944.0
41	1971	392.4	45.6	40.7	74.3	1211.6
42		500.5	45.9	39.5	75.8	1337.3
43		575.9	43.7	34.1	60.9	1337.3
44		504.2	41.8	32.6	68.6	1339.3
45	1972	467.5	49.5	46.0	86.3	1392.7
46		596.4	49.8	44.6	88.1	1556.7
47		686.1	47.4	38.5	70.7	1681.3
48		600.7	45.4	36.9	79.7	1677.1
49	1973	449.0	60.3	56.9	105.2	1735.8
50		572.8	60.7	55.2	107.4	1691.3
51		659.0	57.7	47.6	86.2	1816.9
52		576.9	55.3	45.6	97.1	1794.9
53	1974	446.7	86.2	64.3	119.3	1884.3
54		569.9	86.6	62.4	121.8	2071.0
55		655.6	82.4	53.8	97.8	2179.9
56		574.0	78.5	51.5	110.1	2127.1
57	1975	545.5	184.8	45.1	200.8	2252.5
58		695.9	185.8	43.8	204.9	4627.2
59		800.6	176.8	37.7	164.5	4706.5
60		701.0	169.8	36.1	165.3	4528.1
61	1976	738.6	202.6	89.6	279.1	4981.2
62		942.1	203.7	87.0	284.9	5352.5
63		1083.9	193.9	75.0	228.6	5386.7

Continued

64		949.0	185.6	71.6	257.6	5195.1
65		818.0	260.4	94.6	174.7	6530.9
66	1977	1043.5	261.3	91.8	178.3	6714.7
67		1200.6	249.2	79.2	143.1	6663.3
68		1051.1	238.5	75.8	161.2	6424.7
69		1013.6	319.1	82.6	204.2	6853.1
70	1978	1293.0	321.1	80.1	208.4	7928.0
71		1487.6	305.8	69.1	167.3	7876.2
72		1302.0	292.5	66.2	188.5	7515.7
73		1001.0	339.2	82.8	368.4	8099.9
74	1979	1276.8	341.0	80.3	376.0	8490.9
75		1469.1	324.5	69.3	301.8	8636.4
76		1286.2	310.6	66.3	349.0	8309.1
77		1103.3	384.9	85.0	496.0	9103.7
78	1980	1407.4	387.0	82.5	506.2	10,330.8
79		1619.2	368.2	71.1	406.3	1486.3
80		1417.6	352.4	66.1	457.7	10,061.7
81		1314.1	482.4	87.3	323.9	11,095.9
82	1981	1676.2	485.0	84.7	330.6	12,105.0
83		1928.5	461.5	73.0	265.4	12,377.5
84		1688.4	441.7	70.0	298.9	11,908.9
85		2006.3	440.5	294.4	192.2	13,241.6
86	1982	2559.3	442.5	285.6	196.2	11,241.9
87		2944.5	421.1	246.3	157.5	11,958.3
88		2577.9	402.1	236.9	177.4	11,743.6
89		2242.3	690.7	295.9	235.2	12,675.9
	1983	2860.2	692.5	287.0	242.1	11,420.9
91		3290.6	660.9	247.6	192.7	12,344.8
92		2881.0	632.5	237.1	217.1	12,176.5
93		2559.6	905.2	321.2	344.9	13,127.1
94	1984	3285.1	910.1	311.7	352.0	12,226.5
95		3756.5	866.1	268.8	282.5	13,361.6
96		3288.8	828.9	257.5	318.3	13,226.1
97		3365.0	1153.9	350.3	303.2	14,293.2
98	1985	4292.6	1160.1	339.9	309.4	13,859.0
99		4938.6	1104.0	293.1	248.4	15,021.6
100		4323.8	1056.7	280.7	279.7	14,989.6

Continued

101		3923.8	1248.5	372.5	188.8	15,750.4
102	1986	5005.1	1255.3	361.5	192.7	15,647.0
103		5758.6	1194.5	311.7	154.6	17,115.3
104		5041.6	1143.3	298.6	174.2	18,049.3
105		4066.6	1288.0	399.0	266.6	15,874.4
106	1987	5186.0	1295.0	387.2	274.2	17,460.1
107		5966.6	1232.4	333.8	220.1	17,418.8
108		5223.8	1179.5	319.8	247.9	18,393.6
109		6208.9	1459.6	402.5	232.3	24,995.7
110	1988	7918.8	1467.5	391.6	237.0	25,452.0
111		9110.7	1396.6	387.7	190.2	25,416.5
112		7976.5	1336.6	323.4	214.3	27,358.3
113		9681.4	1549.6	472.1	407.3	32,232.7
114	1989	12,349.6	1558.0	458.1	415.7	34,961.2
115		14,208.4	1482.6	395.1	333.6	35,329.8
116		12,439.5	1419.0	378.4	375.9	36,561.7
117		11,252.2	2055.3	552.0	843.4	53,255.4
118	1990	14,353.4	2066.4	535.7	550.8	54,337.8
119		16,513.8	1966.4	462.0	690.9	53,793.9
120		144,459.0	1882.1	442.4	778.3	55,410.5
121		1360.9	2465.8	595.5	1120.7	65,929.7
122	1991	17,357.0	2479.1	577.9	1143.8	67,103.8
123		19,969.4	2359.2	498.3	918.1	66,260.6
124		17,483.4	2258.0	477.3	1034.2	69,255.9
125		15,911.0	1715.0	618.5	1249.5	75,449.4
126	1992	20,296.2	2729.7	600.2	1275.2	75,243.7
127		23,350.9	2597.7	517.6	1023.6	77,318.5
128		20,444.0	2486.3	495.7	1153.1	80,126.2
129		24,009.1	4013.9	759.3	1647.6	133,927.1
130	1993	30,626.1	4035.8	736.6	1681.7	133,280.6
131		35,235.7	3841.4	635.4	1348.7	130,710.2
132		30,849.2	3675.7	608.5	1520.5	134,716.0
133		39,007.5	6375.5	1006.9	1951.2	158,745.6
134	1994	49,758.2	6410.0	977.0	1991.5	171,233.1
135		57,247.3	6100.0	842.6	1598.4	170,640.0
136		50,120.6	5838.3	806.9	1800.7	175,251.1

Continued

137	1995	59,061.5	9465.7	1518.6	2681.8	211,794.3
138		75,339.1	9517.0	1473.5	2737.1	225,286.2
139		86,678.5	9056.6	1270.8	2196.9	227,716.4
140		75,887.8	8668.1	1217.0	2474.9	235,956.4
141	1996	104,606.4	16,963.2	2096.2	5087.5	475,135.4
142		133,817.8	17,035.0	2033.0	5172.0	482,976.8
143		1539.8	16,210.9	1753.3	4151.2	481,117.3
144		134,792.4	15,515.5	1679.0	4676.6	493,962.0
145	1997	141,959.5	22,731.2	2632.0	7979.1	670,619.8
146		181,084.2	22,854.3	2554.0	8143.7	675,141.6
147		208,339.4	21,748.8	2202.5	6536.4	670,697.4
148		182,403.0	20,815.9	2109.3	7363.6	686,280.4
149	1998	160,649.3	25,279.9	3186.9	9635.5	686,351.0
150		204,924.9	25,416.8	3092.0	9834.4	700,532.3
151		235,788.3	24,187.3	2666.9	7693.4	699,923.5
152		206,417.3	23,149.8	2554.0	8892.3	715,165.8
153	1999	177,413.6	27,595.5	4031.4	11,685.8	647,960.4
154		226,309.6	27,345.0	3911.9	11,928.9	678,288.7
155		360,371.7	26,402.9	3373.6	9572.9	685,015.5
156		227,957.8	25,270.3	3230.8	10,784.4	897,156.2
157	2000	188,577.0	28,651.9	4900.6	13,478.6	777,923.9
158		240,549.6	38,807.0	4755.3	13,756.7	799,246.2
159		276,755.0	27,413.5	4101.0	11,041.0	801,411.1
160		242,301.5	26,237.6	3827.4	12,438.9	816,338.7
161	2001	198,896.3	30,014.3	6217.6	14,364.2	1,165,093.3
162		253,714.9	30,176.8	6033.3	14,650.3	1,144,268.2
163		291,899.5	28,717.1	5203.1	11,758.8	1,124,629.8
164		255,560.7	27,485.2	4982.8	13,246.9	1,148,136.0
0.8	2002	266,058.3	39,839.6	7610.4	19,978.0	1,164,239.0
166		339,385.1	40,055.4	7384.7	20,390.1	1,182,576.3
167		390,466.3	38,117.8	6368.6	16,365.8	1,181,000.0
168		341,856.8	36,482.7	6099.0	18,437.0	1,197,270.8
169	2003	606,640.0	47,242.2	9196.4	24,033.7	1,625,546.2
170		773,832.5	47,498.1	8923.7	24,529.5	1,735,802.5
171		890,302.8	45,200.4	7695.9	19,688.2	1,792,349.4
172		779,468.2	43,261.5	7370.0	22,179.8	1,758,882.6

Continued

173	2004	651,425.4	52,157.4	11,201.4	28,295.5	2,039,515.9
174		830,961.1	15,439.9	10,869.2	28,878.9	2,127,893.3
175		956,029.9	49,903.2	9373.7	23,179.2	2,171,579.1
176		837,012.8	47,762.8	9876.8	26,112.6	2,148,243.2
177	2005	618,328.3	55,588.3	11,560.0	36,012.1	2,831,255.6
178		811,555.8	59,516.5	12,676.1	38,494.5	2,592,273.2
179		1,062,751.7	62,650.6	13,211.3	13,473.9	2,985,541.8
180		98,460.7	66,137.2	13,810.1	42,106.0	3,201,996.4
181	2006	749,944.0	73,375.5	15,212.4	37,696.9	3,169,613.4
182		986,377.9	77,684.6	15,388.1	40,821.5	3,399,351.8
183		1,292,413.0	79,508.8	15,504.6	47,099.6	3,924,775.0
184		1,199,547.3	82,688.3	15,620.1	44,360.0	4,078,498.8
185	2007	931,325.0	87,245.7	19,471.8	44,769.1	3,986,279.5
186		1,232,947.2	89,714.4	18,637.2	46,705.8	4,426,083.8
187		1,619,235.4	104,601.5	17,928.5	53,776.4	4,986,489.4
188		1,506,111.4	97,341.0	17,423.6	61,203.9	5165742.0
189	2008	1,087,907.7	104,146.3	22,167.2	49,557.1	4,740,806.8
190		1,440,102.2	99,366.7	21,231.4	60,332.3	4,853,839.3
191		1,889,913.4	118,850.9	20,494.6	58,139.3	5,524,363.8
192		1,606,557.7	111,787.6	19,918.7	66,494.3	5,538,294.6
193	2009	1,148,460.0	111,750.0	23,920.0	54,710.0	5,421,316.8
194		1,554,120.9	114,084.0	24,401.1	57,072.4	5,704,400.3
195		2,074,683.0	135,850.8	24,417.0	64,352.7	6,340,199.4
196		1,767,306.1	120,471.0	21,534.9	62,474.5	6,376,225.2