Investigating impact of outliers in both independent and dependent variables on 
agricultural production data.

ABSTRACT

The production of high yielding variety (HYV) Boro rice depends on both climatic variables 
and some other non-climatic variables. Outliers may occur commonly in agriculture data. 
Regression outliers either in independent variables or in dependent variables pose a serious 
threat to traditional least squares analysis. The impact of some climatic and non-climatic 
variables like temperature, rainfall, net solar radiation, humidity and wind speed, lag-price 
and fertilizer on HYV Boro rice production have been investigated using regression 
diagnostics and robust regression techniques. In this study, we considered the annual HYV 
Boro rice production data from 1980 to 2000 for Mymensingh and Dinajpur districts in 
Bangladesh. We found that there were outliers in both the independent and dependent 
variables. The outlying observations that were found in the independent variables were 
corrected by the median of the respective variable series, the outliers in the dependent 
variables have been corrected by the robust least-trimmed squares (LTS) predicted 
observations of the HYV Boro production of the selected districts. Hence, the re-weighted 
least squares (RLS) estimation techniques have been used to judge the impact of outliers. The 
regression diagnostics for the selected districts were computed by both the OLS and RLS 
methods. Our study reveals that proper correction of outliers is very important for the 
regression models and there was improvement in the R-squared values for both the districts.

Keyword: Multiple regression; Least square estimators; Regression diagnostics; Outlier; 
Robust regression; High yielding variety (HYV); Climatic and non-climatic 
variables.