

Benha University

Faculty of Commerce

Department of Statistics,

Mathematics and Insurance

Regression Estimation in The Presence of Outliers:
A Comparative Study

By

Maha El-Shahat Mohamed Qura

**Supervised By**

|  |  |
| --- | --- |
| Prof. Ahmed Fathy Mustafa | Prof. Ahmed Mahmoud Gad |
| **Prof of Statistics,** **Faculty of Commerce,****Benha University** | **Prof of Statistics,**[**Faculty of Economics and Political Science, Cairo University**](https://www.facebook.com/FacultyofEconomicsandPoliticalScience) |

***THESIS***

 ***Submitted to the Department of Statistics, Mathematic and Insurance,***

***Faculty of Commerce, Benha University, in partial Fulfillment
of the Requirements for the Degree of***

***Master of Statistics, Benha University***

2015

Abstract

 In linear models, the ordinary least squares (OLS) estimators of parameters have always turned out to be the best linear unbiased estimates. When the sample data contain outliers, this may affect the least-squares estimates of parameters. So, an alternative approach - the so-called robust regression - to the problem is needed to obtain a better fit of the model or more precise estimates of parameters$.$

The goal of robust regression is to develop methods that are resistant to the possibility that one or several unknown outliers may occur anywhere in the data. In this thesis, various robust regression methods have been reviewed. This includes; Least Median Squares (LMS) estimator, Least Trimmed Squares (LTS) estimator, Least Winsorized Squares (LWS) estimator, S-estimator, Least Absolute Value (LAV) estimator, M-estimator, MM-estimator, $τ$-estimator and Robust Efficiency Weighted Least Squares (REWLS) estimator. In the last chapter of this thesis, comparisons of the properties of some robust methods based on problems with outliers in the y-direction (response direction) through a simulation study are illustrated. A real data set application of national growth of 61 countries from all over the world from 1960 to 1985 is also provided to compare the robust estimators. The efficiency and breakdown point are two traditionally used important criteria to compare robust regression techniques by using R software.

The properties of the estimates of regression coefficients with outliers of interest are the mean square errors (MSE) and the bias. The performances of different estimates are studied using simulation and real data for Y-outliers. The estimates of the regression coefficients using nine methods are compared with the least-squares.

For simulation studies, Tukey's M-estimator give a lower TAB and TMSE values than others, for all sample sizes and almost cases and when the contamination is in the direction of the response variable only.

For real data example, Hampel's M-estimators give a lower bias and MSE values than others when the contamination is in the direction of the response variable only.