IRAQI Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (2024) 14 (2): 54-65

UKJAES

University of Kirkuk Journal For Administrative and Economic Science

ISSN 2222-2995
University of Kirkuk Journal For Administrative and Economic Science

Ali Sozan S. Haydar. A Comprehensive Comparative Analysis of Weighted Least Squares and Robust Regression Models with Application. ***University of Kirkuk Journal For Administrative and Economic Science*** **(2024) 14 (2):54-65.**

A Comprehensive Comparative Analysis of Weighted Least Squares and Robust Regression Models with Application

Sozan S. Haydar Ali [1]

[1] *Sulaymani University: College Administration & Economics, Statistics & Informatics Department, Sulaymani, Iraq*

sozan.haider@univsul.edu.iq[1]

Abstract: Regression analysis is indeed a statistical technique used to model the relationship between a dependent variable (response variable) and one or more independent variables (explanatory variables). The conventional method, ordinary least squares (OLS), is typically employed for estimating linear regression parameters under certain assumptions. However, when these assumptions are violated, such as in the presence of outliers in the data, the accuracy of the estimates and results may be compromised. The main objective of this study is to compare two regression models: weighted least squares (WLS) and robust regression (RR), particularly employing the M-Bisquare weighting function in the robust M-estimator regression (RR) model. Additionally, this study aims to showcase that both the weighted least squares (WLS) and robust regression (RR) models can serve as alternative methods to ordinary least squares (OLS), specifically tailored for parameter estimation in datasets containing outliers. The sample population for this research comprises 130 women experiencing infertility, with the variables categorized into two groups: the first group encompasses five explanatory variables (Age, Prolactin (PL), Follicle Stimulating Hormone (FSH), Luteinizing Hormone (LH), and Thyroid Stimulating Hormone (TSH)), while the second group consists of the response variable, the number of eggs (oocyte).. The comparison between weighted least squares and robust regression models was conducted using statistical software such as Eviews and SPSS. The weight least square (WLS) model and robust regression (RR) model were used as a suitable statistical method for the study sample. By analyzing the study data, it was found that the robust regression (RR) model and the weighting function (M- Bisquare) with is the best fit model for this medical data compared to weight least square (WLS) model.

Keywords: Outliers, Robust regression, Weight least square, Weighting function.

تحليل مقارنة شاملة لنماذج المربعات الصغرى المرجحة والانحدار الحصين مع التطبيق

أ.م.د. سوزان صابر حيدر علي [١]

[١] *جامعة السليمانية - كلية الإدارة والاقتصاد، قسم الإحصاء والمعلوماتية، السليمانية، العراق*

المستخلص: تحليل الانحدار هي تقنية إحصائية تستخدم لنمذجة العلاقة بين متغير تابع (متغير الاستجابة) ومتغير واحد أو أكثر مستقل (المتغيرات التوضيحية). تساعد في فهم كيفية تغير قيمة المتغير التابع عند تغيير متغير أو أكثر من المتغيرات المستقلة. وتُستخدم الطريقة التقليدية، المربعات الصغرى العادية (OLS)، عادةً لتقدير معلمات الانحدار الخطي في ظل افتراضات معينة. ومع ذلك، عند انتهاك هذه الافتراضات، كما هو الحال في وجود قيم متطرفة في البيانات، قد تتعرض دقة التقديرات والنتائج للخطر. الهدف الرئيسي من هذه الدراسة هو مقارنة نموذجي الانحدار: المربعات الصغرى المرجحة (WLS) والانحدار الحصين (RR)، وخاصة عند استخدام وظيفة الترجيح M-Bisquare في نموذج الانحدار الحصين لمقدر M (RR). بالإضافة إلى ذلك، تهدف هذه الدراسة إلى إظهار أن كلاً من نماذج المربعات الصغرى الموزونة (WLS) ونماذج الانحدار الحصين (RR) يمكن أن تكون بمثابة طرق بديلة للمربعات الصغرى العادية (OLS)، المصممة خصيصًا لتقدير المعلمة في مجموعات البيانات التي تحتوي على القيم المتطرفة. وفي الجانب التطبيقي لدراستنا، تتكون العينة من ١٣٠ امرأة تعاني من العقم، وقد تم تصنيف المتغيرات إلى مجموعتين: المجموعة الأولى تشمل خمسة متغيرات توضيحية (العمر، البرولاكتين (PL)، الهرمون (FSH)، الهرمون (LH)، والهرمون المحفز للغدة الدرقية (TSH))، بينما تتكون المجموعة الثانية من متغير التابع وهو عدد البويضات (البويضة). وتم إجراء المقارنة بين التقنيتين باستخدام البرامج الإحصائية مثل Eviews وSPSS ، ومن خلال تحليل بيانات الدراسة وجد أن نموذج الانحدار الحصين (RR) ودالة الترجيح (M- Bisquare) هو النموذج الأنسب لهذه البيانات الطبية مقارنة بنموذج المربعات الموزونة (WLS).

الكلمات المفتاحية: القيم الشاذه (المتطرفه)، نموذج الانحدار الحصين، نموذج المربعات الصغرى المرجحة، داله Weighting.

Corresponding Author: E-mail: sozan.haider@univsul.edu.iq

1. Introduction [11][12][17]

Data classification methods are among the most commonly utilized statistical techniques. The term "regression" was coined by Francis Galton in the nineteenth century to describe a biological phenomenon. Consequently. Weighted Least Squares and Robust Regression Models" highlights the significance of advanced regression techniques in statistical analysis. It discusses the importance of addressing heteroscedasticity and outliers, introducing the concepts of Weighted Least Squares and Robust Regression as powerful methods for improving the accuracy and reliability of regression models. This comprehensive analysis compares the efficacy of Weighted Least Squares and Robust Regression models, emphasizing their applications in statistical research and highlighting their strengths and limitations**.** Fransiska et al. (2022) conducted a study to address violations of heteroscedasticity using Weighted Least Squares (WLS) and Quantile Regression techniques. Their objective was to compare the effectiveness of these methods in overcoming heteroscedasticity and identifying the most precise approach for their specific research context. Their findings indicated that the WLS method outperformed Quantile Regression, yielding estimates with a higher level of precision (98%). This suggests that WLS is the preferred method for addressing heteroscedasticity in their research. Sulaimon Mutiu, O. (2015) conducted this study to model the loss of properties resulting from fire outbreaks in Ogun State, Nigeria, utilizing Simple Weighted Least Squares Regression. The researchers collected secondary data on fire outbreaks and the monetary value of property losses across the twenty Local Government Areas of Ogun State for the year 2010. They employed Simple Weighted Least Squares Regression as their statistical method and analyzed the data using SPSS 21.0 software. The sample for this study consists of data from the twenty Local Government Areas of Ogun State for the year 2010. The researchers utilized secondary data on fire outbreaks and property losses for their analysis. The study concludes that there exists a very strong positive relationship between the number of fire outbreaks and the loss of properties in Ogun State. This relationship is not only significant but also accounts for approximately 91.2% of the property loss in the state. Therefore, an increase in the number of fire outbreaks leads to a corresponding increase in the monetary loss of properties, indicating the significant influence of fire outbreaks on property loss in Ogun State. Alma (2011) illustrated the definition and behavior of outliers in linear regression and compared several robust regression methods via simulation study. The simulation study aimed to determine the best method for all linear regression scenarios. According to Alma, four robust regression methods were comparatively

evaluated against the OLS regression method. The S estimator demonstrated reasonable efficiencies, bounded the influence of high leverage outliers, and exhibited high breakdown. Mohammed (2013) studied three different methods of robust regression: Maximum likelihood type estimate M-estimator, Modified Maximum likelihood type estimate MM-estimator, and Least Trimmed Squares LTS-estimator, comparing their results with the OLS method. Mohammed concluded that the estimated values of parameters using OLS were significantly different from those obtained using robust methods, particularly the LTS-estimator, which showed better results based on the MSE criterion, and the M-estimator, which performed better according to the MAPE criterion. Susanti et al. (2014) presented M estimation, S estimation, and MM estimation in robust regression to determine a regression model. M estimation, an extension of the maximum likelihood method, is robust, while S estimation and MM estimation are developments of the M estimation method. The authors applied these methods to maize production data and found that the best model for maize production data was determined to be S estimation.

A. Introduction Classic Regression Model [4][14][18]

Regression analysis emerged as a statistical process for estimating relationships among variables. Its scope encompasses estimating regression model parameters, indicating the strength and direction of relationships, and evaluating the estimated model. Regression analysis encompasses various techniques for modeling and analyzing multiple variables. Its primary focus lies in exploring the relationship between a dependent variable, often denoted as y or the response variable, and one or more explanatory variables, denoted as xi or predictors. Furthermore, the linear effect of the explanatory variables on response variable, the application of Ordinary Least Squares estimator method (OLSEM) in simple or multiple linear regressions hinges on several crucial assumptions: Normality of the error terms, Homoscedasticity, No auto-correlation problem, and No or little multicollinearity problem).

The linear regression model between a response variable y and k explanatory variables: $X_1, X_2, \ldots \ldots X_k$then the multiple linear regression models can be expressed as

$$Y_i = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon_i \qquad (1)$$

Where:

Y_i: represents the value of the response variable in the i^{th} observation

β_o : denotes the intercept parameter.

β_k: signifies the slope parameter linked with the k^{th} variable.

X_{ji}: stands for the k^{th} explanatory variable associated with the i^{th} observation. $i = 1,2,..n\ j = 1,2,..k$

ε_i: refers to the random error term of the i^{th} observation, $\varepsilon_i \sim N(0, \sigma^2{}_e)$

B. Ordinary Least Squares Estimator Method (OLSEM) [13][14]

The least-squares method is employed to estimate the regression coefficients in a multiple linear regression model. We present multiple regression analysis using matrix notation. This model examines the relationship between one dependent variable and multiple explanatory variables. Furthermore, multiple linear regression can be viewed as an expansion of simple linear regression, where there are k explanatory variables, or conversely, simple linear regression can be seen as a specific case of multiple linear regression when k = 1. Additionally, we utilize the least squares method to determine the estimated coefficients ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots \ldots \hat{\beta}_k$) that minimize the sum of squares error.

The regression model in terms of the observations, Equation (1), may be written in matrix notation as

$$y = X\beta + \varepsilon \qquad (2)$$

Where:

$$Y=\begin{pmatrix}Y_1\\Y_2\\\cdot\\\cdot\\Y_n\end{pmatrix},X=\begin{pmatrix}1 & X_{11} & \dots\dots & X_{1k}\\1 & X_{21} & \dots\dots & X_{2k}\\\vdots & \vdots & & \vdots\\\cdot & \cdot & & \cdot\\1 & X_{n1} & & X_{nk}\end{pmatrix},\ \beta=\begin{pmatrix}\beta_0\\\beta_1\\\cdot\\\cdot\\\beta_k\end{pmatrix},\text{and }\varepsilon=\begin{pmatrix}\varepsilon_1\\\varepsilon_1\\\cdot\\\cdot\\\varepsilon_n\end{pmatrix}$$

$\underline{Y}$: The vector of response variable from $n \times 1$

X: Matrix of explanatory variables from $n \times k$ where i=1,2,…,n numbers of observation j=1,2,….k number of variable

$\underline{\beta}$: The vector of parameters from $(k+1) \times 1$

$\underline{\varepsilon}$: The vector of error

The objective is to determine the vector least squares estimators denoted as, $\hat{\beta}$, which minimizes

$$L=\sum_{i=1}^{n}\varepsilon_i^2=\varepsilon'\varepsilon=(Y-X\beta)'(Y-X\beta)$$

For given y and x. A minimum will always exist as (L) is a real-valued:

$$L=\sum_{i=1}^{n}\varepsilon_i^2=\varepsilon'\varepsilon=Y'Y-\beta'X'Y-Y'X\beta+\beta'X'X\beta \qquad (3)$$

As $(\beta'X'y)$ represents a (1×1) matrix, also known as a scalar, its transpose $(\beta'X'Y)'=Y'X\beta$ yields the same scalar. In the context of least squares estimators, it is essential to ensure that

$$L=Y'Y-2\beta'X'Y+\beta'X'X\beta$$

Differentiate (L) with respect to β

$$\frac{\partial L}{\partial \beta}=-2X'Y+2X'X\beta$$

The normal equation is: $\frac{\partial L}{\partial \beta}=0$, then:

$$(X'Y)=(X'X)\beta \rightarrow \hat{\beta}=(X'X)^{-1}(X'Y) \qquad (4)$$

2. The Weighted Least Square Estimator Method (WLS) [8][18][20]

A. Patterns of Weighted Least Squares (WLS): Exploring Variance Relationships

In this method, the deviation between the observed and expected values of yi is multiplied by a weight wi where wi is chosen to be inversely proportional to the variance of yi. The multiple linear regression models can be expressed as:

$$Y_i=\beta_o+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k+\varepsilon_i \qquad (5)$$

The model in terms of the observations, Equation (1), may be written in matrix notation as

$$\underline{Y}=\underline{X\beta}+\underline{\varepsilon} \qquad (6)$$

To estimate parameters using WLS, the following objective function is minimized: The Weighted Least Squares function is

$$\varepsilon'W\varepsilon=(Y-X\beta)'W(Y-X\beta)$$

The objective is to determine the vector weight least squares estimators, denoted as, $\hat{\beta}$ which minimize (L)

$$L=\varepsilon'W\,\varepsilon=(Y-X\beta)'W(Y-X\beta)$$

The matrix W is a diagonal matrix containing W_i:

$$W = \begin{pmatrix} W_1 & 0 & 0 & \dots & 0 \\ 0 & W_2 & 0 & \dots & 0 \\ 0 & 0 & W_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & W_n \end{pmatrix} \quad \text{Then: } W = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\sigma_3^2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{\sigma_n^2} \end{pmatrix}$$

Differentiate (L) with respect to β

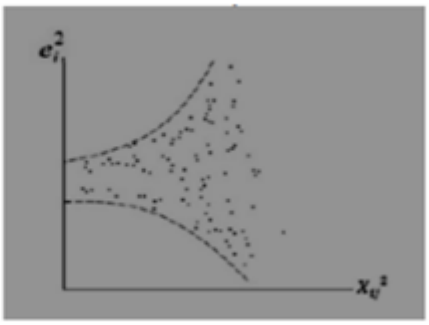
$$\frac{\partial L}{\partial \beta} = -2X'WY + 2X'WX\beta \quad \text{The normal equation is:}$$

$$\frac{\partial L}{\partial \beta} = 0 \rightarrow (X'WY) = X'WX\beta \rightarrow \hat{\beta} = (X'WX)^{-1}(X'WY) \qquad (7)$$

B. Patterns of Weighted Least Squares (WLS): Exploring Variance Relationships

The weight is determined by analysing the residual pattern of the independent variable. These patterns include:

1. The Error variance is proportional to X_{ij}^2 for $1 \leq j \leq k$

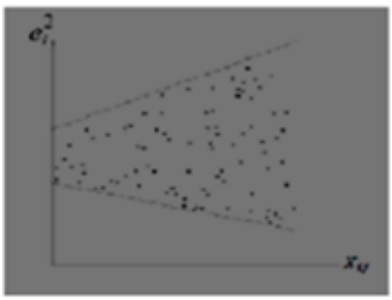


$$E(\varepsilon^2) = \sigma^2 X_{ij}^2$$

Figure 1. The error variance is directly proportional to X_{ij}^2.

If the pattern exhibits a quadratic relationship as depicted in Figure 1, it suggests that the error variance is directly proportional to X_{ij}^2. Therefore, the weights utilized in the WLS method are given by $\frac{1}{X_{ij}^2}$.

2. The Error Variance is proportional to X_{ij}



$$E(\varepsilon^2) = \sigma^2 X_{ij}$$

Figure 2. The error variance is directly proportional to X_{ij}.

Should the trend indicate a linear correlation akin to that depicted in Figure 2, it can be inferred that the error variance correlates directly with X_{ij}, consequently, the weights applied in the WLS technique would be $\frac{1}{\sqrt{X_{ij}}}$.

3. The Error Variance is proportional to $[E(Y_i)]^2$

$$E(\varepsilon^2) = \sigma^2[E(Y_i)]^2$$

Figure 3. The error variance is directly proportional to $[E(Y_i)]^2$.

If the variance of the error is directly proportional to $[E(Y_i)]^2$, the Weighted Least Squares (WLS) method involves conducting Ordinary Least Squares (OLS) regression while disregarding heteroscedasticity. This is done to derive the Y_i value, which is then employed as a weight. Consequently, the regression equation transforms into:

$$\frac{\beta_o}{\hat{Y}_i} + \beta_1 \frac{X_{i1}}{\hat{Y}_i} + \beta_2 \frac{X_{i2}}{\hat{Y}_i} + \cdots + \beta_k \frac{X_{ik}}{\hat{Y}_i} + \frac{\varepsilon_i}{\hat{Y}_i}$$

3. Robust Regression Model (RR) [2][7][19]

A. Introduction of the Robust Regression Model (RR)

In multiple regression analyses, the ordinary least squares (OLS) method is typically employed for parameter estimation, given that certain assumptions are met to derive regression weights for data analysis. These assumptions include the requirement for a continuous response variable, normally distributed residual errors, homoscedasticity (equal variance across explanatory variable levels), and uncorrelatedness between explanatory variables and with each other. However, when data contain missing entries, outliers, non-normal distributions, or multicollinearity among variables, the clarity of sample estimates and results may be compromised. The ordinary least squares estimation of multiple regression is particularly susceptible to outliers, non-normality, multicollinearity, heteroscedasticity, and missing data, given that outliers violate the assumption of normally distributed residuals in least squares regression. Consequently, the OLS estimator is highly sensitive to outliers and deviations from normality assumptions. This sensitivity means that estimation results are directly influenced by each observation, necessitating detailed data analysis. Indeed, the importance of this issue is such that even a single observation can dramatically alter parameter estimates, potentially resulting in entirely different outcomes if omitted from the data. Outliers wield significant influence on model fitting, particularly in regression analysis, prompting the use of robust regression models to enhance robustness. Robust regression estimators is an alternative to least-squares regression models when residual distribution deviates from normality or outliers exert substantial influence on the model. Robust regression serves as a valuable method for addressing these challenges and providing more reliable results in such scenarios. The main objective of robust regression analysis is to fit a model. This method is a great tool for analyzing the data which is affected by outliers. Then, the robust regression model is insensitive to outliers and provides better predictions, and it is an important method for analyzing data that are infected with outliers. It can be used to identify outliers and to provide resistant results in the presence of outliers. Moreover, robust regression analyses have been developed to improve least squares estimation in the presence of outliers. Then, robust regression is the most suitable model that seeks to identify outliers and minimize their impact on the coefficient estimates. Moreover, some of these robust regression techniques are maximum likelihood Estimator (M- Estimator), S-Estimator (S- Estimator), and Modified Maximum likelihood Estimator (MM- Estimator).

B. Maximum likelihood estimator M-estimation [2][7][14]

M-estimation, considered the prevalent approach in robust regression, was introduced by Huber in 1964. It represents a broadened concept of maximum likelihood estimation within the framework of location models. Essentially, M-estimation can be regarded as "maximum likelihood type estimators," almost as effective as OLS, but diverging from the goal of minimizing the sum of squared errors typically seen in OLS. Instead, the fundamental principle of the M-estimation method revolves around minimizing the residual function. The objective function of M-estimation is formulated as:

$$\hat{\beta}_{RE} = \min_B \sum_{i=1}^{n} \rho(e_i) = \min_B \sum_{i=1}^{n} \rho(Y_i - \sum_{j=1}^{k} X_{ij}\beta_j) \qquad (8)$$

$$\min_B \sum_{i=1}^{n} \rho(\frac{e_i}{s}) = \min_B \sum_{i=1}^{n} \rho\left[\frac{Y_i - \sum_{j=1}^{k} X_{ij}\beta_j}{s}\right] = \min_B \sum_{i=1}^{n} \rho(u_i)$$

In the given context, "s" represents an estimate of scale derived from a linear combination of the residuals, where $\rho(e_i)$ is termed as an objective function. It's imperative to estimate "s" concurrently. One viable approach is to utilize the Median Absolute Deviation (MAD) scale estimator

$s = c \times \text{Median}|e_i - \text{median}(e_i)|$, Assuming Gaussian noise, where c = 1.4826,

$$s = \frac{MAD}{0.6745} = \frac{\text{Median}|e_i|}{0.6745} \qquad \text{Or} \qquad s = \frac{\text{Median}|e_i - \text{median}(e_i)|}{0.6745} \qquad (9)$$

$$s = 2.1\ \text{Median}|e_i| \qquad \text{Or} \qquad s = 1.5\ \text{Median}|e_i| \qquad (10)$$

Where (s) (median absolute deviation) is an estimate of scale often formed from linear combination of the residuals, the constant 0.6745 makes (s) an approximately unbiased estimate of (σ) if N is large and the distribution is normal. For ρ function:

$$\min_B \sum_{i=1}^{n} \rho\left[\frac{Y_i - \sum_{j=1}^{k} X_{ij}\beta_j}{s}\right] \qquad (11)$$

The function ρ determines the impact of each residual on the objective function. As per Alma (2011), an appropriate ρ should possess the following characteristics:

$\rho(e) \geq 0\ , \rho(0) = 0, \rho(e) = \rho(-e)\ and\ \rho(e_i) \geq \rho(e_i')$ for $|e_i| \geq |e_i'|$, and ρ is continuous

A robust M-estimator accomplishes this objective by minimizing the sum of an objective function that increases less rapidly than the one used in the least squares estimator, where $\rho(e) = (e^2)$

So, we minimize equations (9) with respect to each of the p parameters $\beta_1, \beta_2, , , , \beta_p$ and this resulted in a system of p equations:

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{e_i}{s}\right) = \sum_{i=1}^{n} X_{ij}\psi\left(\frac{Y_i - \sum_{j=1}^{k} X_{ij}\beta_j}{s}\right) = 0 \qquad (12)$$

$$j = 1,2, \ldots . p \quad \text{and} \quad i = 1.2 \ldots . n$$

Where: $\psi = \rho'$ x_{ij} is the i-th observation of the j-th parameter. Definition and the weighting function and $w(e_i) = \frac{\psi(e_i)}{e_i} = \frac{\psi(\frac{Y_i - \sum_{j=1}^{k} X_{ij}\beta_j}{s})}{\frac{Y_i - \sum_{j=1}^{k} X_{ij}\beta_j}{s}}$ So that equation (12) can be written:

$$\sum_{i=1}^{n} X_{ij} W_i\left(Y_i - \sum_{j=1}^{k} X_{ij}\hat{\beta}_j\right) = 0 \qquad j = 1,2,3, \ldots \ldots, k \qquad (13)$$

Since $s \neq 0$, we define the weight matrix W = dig($w_i : i = 1,2, \ldots \ldots, n$)as:

$$W = \begin{pmatrix} W_1 & 0 & 0 & \dots & 0 \\ 0 & W_2 & 0 & \dots & 0 \\ 0 & 0 & W_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & W_n \end{pmatrix}$$, In matrix notation, equation (8) can be written as:

w_i: I It can be found form table (1) in the section 3.3.2 Weight Function W(u)

$$\rightarrow (X'WX)\hat{\beta} = (X'WY) \quad \rightarrow \hat{\beta} = (X'WX)^{-1}\ (X'WY) \qquad (14)$$

C. Objective Function ρ(u)[1][21]

The objective function, denoted as ρ(u), plays a pivotal role in M-estimators within robust regression models, guiding the estimation process by assessing the impact of deviations from expected values. In comparing various M-estimators such as the least-squares, Huber, and Tukey bisquare, distinct patterns emerge in their objective functions. While, the bisquare function eventually levels off for residuals exceeding a certain threshold. This differential response highlights the varying sensitivity of these estimators to outliers. The tuning constant 'k' further influences the behavior of the estimators, balancing resistance to outliers with efficiency, particularly crucial in normal error distributions. The objective and weight functions for the Bisquare estimators are also given in table (1),

D. Weight Function W(u) [1][21]

In robust regression analysis, the weight function 'w(u)' crucially determines the contribution of each observation to the estimation process, particularly regarding the treatment of outliers. M-estimators, by design, assign reduced weights to unusual observations, contrasting starkly with the equal weighting assigned by Ordinary Least Squares (OLS). This discrepancy underscores the enhanced robustness of M-estimators against heavy-tailed error distributions and non-constant error variance. Notably, tuning constants play a significant role in shaping the behavior of these estimators, striking a delicate balance between relative efficiency and breakdown point. The weight functions, exemplified by the bisquare (1987), exemplify this robustness by assigning smaller weights to extreme residuals, thus mitigating their influence on the estimation process. Table (1) shows the weight functions for Bisquare.

Table (1): Objective and weight functions for M- Bisquare in RR

Objective Function ρ(u)	Weight Function w(u)	Default
$\rho_B(e) = \begin{pmatrix} (\frac{k^2}{6}\left[\left(1-\left(\frac{e}{k}\right)^2\right)^3\right]) & \text{for} \quad \lvert e\rvert \le k \\ (\frac{k^2}{6}) & \text{for} \quad \lvert e\rvert > k \end{pmatrix}$	$w_B(e) = \begin{pmatrix} \left[1-\left(\frac{e}{k}\right)^2\right]^2 & \lvert e\rvert \le k \\ 0 & \lvert e\rvert > k \end{pmatrix}$	K=4.685

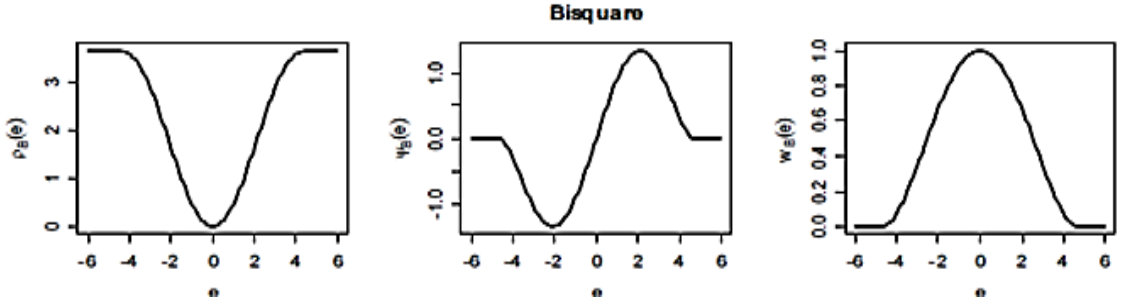


Figure (4): Objective and weight functions for M- Bisquare in RR

4. Criteria for Evaluating the Accuracy [3][14][9]

When evaluating the precision of an estimation method, it's crucial to account for potential sources of inaccuracy, such as mismatches between the data characteristics and the procedures employed. For instance, utilizing inadequate procedures for the data type or using inappropriate data examples can lead to inaccuracies. Statisticians commonly employ various criteria to assess estimation methods, aiming to ensure their effectiveness and reliability. Criteria for assessing the accuracy of an estimation method include:

A. Mean Absolute Error (MAE)

In statistical analysis, the Mean Absolute Error (MAE) serves as a metric to quantify the discrepancies between paired observations representing identical phenomena. MAE computation involves determining the average absolute difference between these paired values. MAE is calculated as:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|Y_i - \hat{Y}_i|}{n} \qquad (13)$$

B. Relative Absolute Error (RAE)

Relative Absolute Error (RAE) offers a means to compare errors across models, even when measured in different units. Typically, it's defined as follows:

$$\mathrm{RAE} = \frac{\sum_{i=1}^{n}|Y_i - \hat{Y}_{i_i}|}{\sum_{i=1}^{n}|Y_i - \bar{Y}|} \qquad (14)$$

Where: Y: is response variable (actual), $\hat{Y}_i$: is the predicted value of Y, $\bar{Y}$: is the mean of the response variable and n: is the sample size.

5. Data analysis

A. Data description

The sampling of this paper consists of (130) women suffering from infertility in Royal hospital in Sulaymaniyah city, and a set of variables were taken as follows: The first group represents (5) explanatory variables which are {Age, Prolactin (PL), Follicle Stimulating Hormone (FSH), Luteinizing Hormone (LH) and Hormone Thyroid stimulation (TSH)}. While the second group represents the response variable, which is number of eggs (oocyte). The weight least square and robust regression model were used as suitable statistical methods for the study sample. To compare the weight least square and robust regression model by using Statistical program (Eviews & SPSS).

Table (2): Shows the variables

Response variable	**Explanatory variables**				
Y	X_1	X_2	X_3	X_4	X_5
Number of Oocyte	Age	PL	FSH	LH	TSH

B. Application of the Ordinary Least Square

Table (3): The Regression Coefficients According to the Ordinary Least Square (OLS)

Explanatory variables	Constant	X_1(Age)	X_2(PL)	X_3(FSH)	X_4(LH)	X_5(TSH)
Coefficient	46.214	-0.753	-0.034	-0.931	0.241	0.661
t-Statistic	15.053	-10.306	-0.46	-6.79	1.837	4.235
P-value	0.000	0.000	0.646	0.000	0.069	0.000
Scale of comparison	MAE	3.376242		RAE	0.434036842	

The hypothesis test: H_o: $\beta_j = 0$ V/S H_1: At least two of them aren't equal, j=1, 2, 3,……, k
It can be seen from Table (3) that there is statistically significance relationship between (Age, FSH, TSH) with the response variable which is number of oocyte because the p-value of these was less than the common alpha 0.05. A one-point increase in age corresponds to a decrease of (0.753) oocytes, while a similar increase in FSH result in decreases of (0.931) oocytes. Conversely, a one-point rise in TSH leads to an increase of (0.661) oocytes. However, no significant correlation was found between PL and LH and the response variable which is oocyte count because the p-value of both variables was more than 0.05. the mean absolute error (MAE) equals to (3.376242) and relative absolute error (RAE) equals to (0.434036842).

C. Application of the Weight Least Square

Table (4): The Regression Coefficients According to the Weight Least Square (WLS)

Explanatory variables	Constant	X_1(Age)	X_2(PL)	X_3(FSH)	X_4(LH)	X_5(TSH)
Coefficient	57.970	-0.895	-0.204	-1.447	0.332	0.898
t-Statistic	21.686	-15.158	-8.621	-20.996	4.456	7.187
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Scale of comparison	MAE	2.660219154		RAE	0.341987705	

The hypothesis test: H_o: $\beta_j = 0$ V/S H_1: At least two of them aren't equal, j=1, 2, 3,……, k
It can be seen from Table (4) that there is statistically significance relationship between (Age, PL, FSH, LH and TSH) with the response variable which is number of oocyte because the p-value of these was less than the common alpha 0.05.
A one-point increase in age corresponds to a decrease of (0.895) oocytes, while a similar increase in PL and FSH results in decreases of (0.204 and 1.447) oocytes, respectively. Conversely, a one-point rise in LH and TSH leads to increases of (0.332 and 0.898) oocytes, respectively. The mean absolute error (MAE) equals to (2.660219154) and relative absolute error (RAE) equals to (0.341987705).

D. Application of the Robust Regression Model (RR)

Table (5): Shows the Regression Coefficients According to the Robust Regression Model (RR)

Explanatory variables	Constant	X_1(Age)	X_2(PL)	X_3(FSH)	X_4(LH)	X_5(TSH)
Coefficient	49.726	-0.728	-0.220	-1.292	0.125	1.324
t-Statistic	73.105	-44.992	-13.583	-42.539	4.295	38.281
P-value	0.000	0.000	0.000	0.000	0.000	0.000
Scale of comparison	MAE	2.03908		RAE	0.26213642	

The hypothesis test: H_o: $\beta_j = 0$ V/S H_1: At least two of them aren't equal, j=1, 2, 3,……, k
It can be seen from Table (5) that there is statistically significance relationship between (Age, PL, FSH, LH and TSH) with the response variable which is number of oocyte because the p-value of these was less than the common alpha 0.05.
A one-point increase in age corresponds to a decrease of (0.728) oocytes, while a similar increase in PL and FSH results in decreases of (0.220 and 1.292) oocytes, respectively. Conversely, a one-point rise in LH and TSH leads to increases of (0.125 and 1.324) oocytes, respectively. The mean absolute error (MAE) equals to (2.03908) and relative absolute error (RAE) equals to (0.26213642).

Table (6): Models Compassion table

Methods	MAE	RAE
OLS	3.376242	0.434036842
WLS	2.660219154	0.341987705
RR- M-Bisquare	2.03908	0.26213642

6. Discussion

Weighted least squares and robust regression models were employed as appropriate statistical techniques for analyzing the study sample. Our primary objective in this study is to compare the Weighted Least Squares (WLS) and Robust Regression (RR) models. Specifically, we utilize the M-Estimator Method to estimate the procedures of the Robust Regression Model, focusing on the M-Estimator with M-Bisquare-Estimation. Different criteria or measures, such as Mean Absolute Error (MAE) and Relative Absolute Error (RAE), are employed to determine the best model for the data.The results of the estimation for both criteria for the estimation methods are presented in Table 6. Here, it is observed that the Classic Regression model, represented by the Ordinary Least Squares (OLS) Method, has MAE and RAE values of 3.376242 and 0.434036842 respectively. These values are higher than those of the Weighted Least Squares (WLS) method and M-Bisquare estimation method (RR). Subsequently, the Weighted Least Squares method shows MAE and RAE values of 2.660219154 and 0.341987705 respectively. In Table 8, the results of the estimation for both criteria for the estimation methods are presented again. This time, it demonstrates the preference for the robust regression method, particularly the M-Bisquare estimation. Here, the MAE and RAE criteria for the robust regression model (M-Bisquare estimation) are 2.03908 and 0.26213642 respectively, which are lower than those of the OLS Method and WLS method. Based on the study's results, it is evident that the robust regression model (M-Bisquare estimation) is the best fit for the data compared to the Weighted Least Squares (WLS) and Classic Regression (OLS) models. This determination is made considering criteria such as Mean Absolute Error (MAE) and Relative Absolute Error (RAE), where the robust regression model outperforms the Ordinary Least Squares (OLS) Method and Weighted Least Squares (WLS).

7. Conclusion

A. According the results of the Table 6, it's evident that the most suitable model among classic regression (OLS), weighted least squares (WLS), and robust regression (RR) models was robust regression (RR) with M-Bisquare estimation methods. This conclusion is drawn from comparing MAE and RAE, where all criteria values were smaller for robust regression (RR) compared to classic regression (OLS) and weighted least squares (WLS) models. Hence, utilizing robust regression (RR) with M-Bisquare weighting functions emerges as a favorable method for parameter estimation.
B. According to the result of weighted least squares (WLS) and robust regression (RR) model that (Age, PL, FSH, LH and TSH) have effects on increasing or decreasing of the response variable. Moreover, age, PL and FSH have negative impact on the oocyte and other explanatory variables have positive impact on the response variable.
C. Depends on the significant of variables, the impact of explanatory variables (Age, PL, FSH, LH and TSH) on number of oocyte can be ordered according to the result of robust regression (RR) with M-Bisquare weighting functions. As a result, the order were (Age, FSH, TSH, PL and LH) respectively.

D. The findings of this study suggest that weighted least squares (WLS) and robust regression (RR) models serve as viable alternatives to ordinary least squares (OLS), particularly when dealing with outlier-laden data.
E. Ultimately, the optimal approach for estimating parameters and regression coefficients affecting the number of oocytes (eggs) is found to be robust regression (RR) with M-Bisquare estimation. This method outperforms classic regression (OLS) and weighted least squares (WLS) models based on MAE and RAE criteria.

It is essential to apply all robust regression methods, including S-estimator and MM-estimator, and compare them with the weighted least squares method, particularly when handling extensive datasets. This comprehensive evaluation is crucial for assessing the effectiveness of each method within the robust regression model and weighted least squares framework

References:

1. Abed Mohammed, S.I.Z.A.R., 2013. "Comparing some Robust Methods with OLS Method in Multiple Regression with Application". Journal of Concrete & Applicable Mathematics, 11(1).
2. Adedia, D., 2014. "Comparison of robust regression estimators" (Doctoral dissertation).
3. Ahmed, D. H., Mohamad, S. H., & Karim, R. H. R. 2023. "Using Single Exponential Smoothing Model and Grey Model to Forecast Corn Production in Iraq during the period (2022-2030)". University of Kirkuk Journal for Administrative and Economic Science, 13(3).
4. Ahmed, N. M., Hamdin, A. O., & Mohamad, S. H., 2023. "Copula Probability Model for Determining Dependency and Generations", University of Kirkuk Journal for Administrative and Economic Science, Vol. 13, No.2.
5. Alma, Ö.G., 2011. Comparison of robust regression methods in linear regression. Int. J. Contemp. Math. Sciences, 6(9), pp.409-421.
6. Bhar, L., 2008. "Robust regression". New Delhi: IASRI, 70(350), pp.70-78.
7. Fox, J. weisberg, 2002. Robust regression: Appendix to an R and S-PLUS companion to applied regression.
8. Fransiska, W., Nugroho, S., & Rachmawati, R. 2022. "A comparison of weighted least square and quantile regression for solving heteroscedasticity in simple linear regression". Journal of Statistics and Data Science, 1(1), 19-29.
9. Hamad A. S., Faqe M. M., & Mohamad S. H. 2023. "Forecasting Life-Expectancy in Iraq During the Period (2022-2035) Using Fuzzy Markov Chain". University of Fallujah , Journal of Business Economics for Applied Research, Volume 5, Issue 3, Pages 347-372. https://shorturl.at/aJQSX
10. Hussein, M. M. F., Saeed, A. A., & Mohamad, S. H. 2023. "Comparison Markov Chain and Neural network models for forecasting population growth data in Iraq", University of Kirkuk Journal for Administrative and Economic Science. Economic Science, Vol.13, Issue(4), P1-14.
11. Karim Hama Ali, F., A Abdullah, S., & Husen Mohamad, S. (2023). "Relationship Between Socio-Demographic Characteristics and Food Labeling Consumption in Sulaimani City by Using Chi-Square Test". Al-Qadisiyah Journal for Agriculture Sciences, 13(1), 139-146.
12. Kewan, J.N.T.S.E., 2015. "Estimation of multivariate multiple linear regression models and applications (Doctoral dissertation)".
13. Li, Y., 2010. "An Empirical Analysis of Family Cost of Children: A Comparison of Ordinary Least Square Regression and Quantile Regression".
14. Mohamad, S. H., (August. 7. 2021), University of University of Sulaimani MSc Thesis (A comparison between Robust regression model and quantile regression model Applied in a medical Field), College administration and economic – statistics and informatics department.
15. Mohammed, S.A.,2013. "Comparing Some Robust Methods with OLS Method in Multiple Regression with application. Journal of Concrete & Applicable Mathematics", 11(1).
16. Nahar, J. and Purwani, S., 2017. "Application of robust M-estimator regression in handling data outliers". 4th ICRIEMS Proceedings, pp.53-60.
17. Omer, A., Faraj, S. M., & Mohamad, S. H. 2023. "An application of two classification methods: hierarchical clustering and factor analysis to the plays PUBG". Iraqi Journal of Statistical Sciences, 20(1), 25-42.
18. Rosopa, P. J. 2006. "A Comparison of Ordinary Least Squares, Weighted Least Squares, And Other Procedures When Testing for The Equality Of Regression".
19. Schumacker, R.E., Monahan, M.P. and Mount, R.E., 2002. "A comparison of OLS and robust regression using S-PLUS. Multiple Linear Regression Viewpoints", 28(2), pp.10-13.
20. Sulaimon Mutiu, O. 2015. "Application of weighted least squares regression in forecasting". Int. J. Recent. Res. Interdiscip. Sci, 2(3), 45-54.
21. Susanti, Y. and Pratiwi, H., 2014. "M estimation, S estimation, and MM estimation in robust regression". International Journal of Pure and Applied Mathematics, 91(3), pp.349-360.