A Robust Regression Method Based on Pearson Type VI Distribution

Yasin Büyükkör and A. Kemal Şehirlioğlu

Abstract In classical regression analysis, the distribution of the error is assumed to be Gaussian, and Least Squares (LS) estimation method is used for parameter estimation. In practice, even if the distribution of errors is assumed to be Gaussian, residuals are not generally Gaussian. If the data set contains outlier (s) or there are observations that are suspected to be outlier, normality assumption is violated, and parameter estimates will be biased. Many statisticians used robust method, such as the M-Estimation Method, which is a generalized version of the Maximum Likelihood (ML) Estimation method, for parameter estimation when such problems occurred. However, if the data set has skewness and excess kurtosis, traditional M-Estimators cannot achieve a good solution. In this study, using the relationship between Pearson Differential Equation (PDE) and Influence Function (IF), M-Estimation method is proposed for data sets that follow Pearson Type VI (PVI) distribution. The advantage of this method takes into account the skewness and kurtosis values of the data set and generates dynamic solutions. Objective, influence, weight functions and tail properties of the PVI distribution are obtained by using the Probability Density Function (pdf) of the PVI distribution. For the regression parameter estimates, Iteratively Re-Weighted Least Squares Estimation Method (IRWLS) is used. In many simulation studies with different scenarios and applications with real data, if the data have skewness and excess kurtosis, the proposed method has achieved better results than other M-Estimation methods in terms of Total Absolute Deviation (TAB) and Mean Square Error (MSE).

Keywords M-Estimation method · Robust regression · Pearson type VI distribution · Influence function · Iteratively re-weighted least squares method

Y. Büyükkör (✉)
Faculty of Economics and Administrative Sciences, Department of Econometrics, Karamanoğlu Mehmetbey University, 70200 Karaman, Turkey
e-mail: yasinbuyukkor@kmu.edu.tr

A. K. Şehirlioğlu
Faculty of Economics and Administrative Sciences, Department of Econometrics, Dokuz Eylül University, 35210 İzmir, Turkey
e-mail: kemal.sehirli@deu.edu.tr

M. K. Terzioğlu (ed.), *Advances in Econometrics, Operational Research, Data Science and Actuarial Studies*, Contributions to Economics,
https://doi.org/10.1007/978-3-030-85254-2_8

1 Introduction

In the history of statistics, many researchers have analyzed the data assuming Gaussian distribution. Thus, anomalies in the data (heavy-long tail, excess kurtosis, skewness, outlier, etc.) are often ignored by researchers. Such anomalies can be caused by many reasons. The main reasons are measurement and recording mistakes or mixing of two or more populations. However, in the data set, an observation (s) belonging to the data set can act like an abnormal observation. Researchers have difficulty in analyzing the data set in the presence of such anomalies and they use analysis methods 'robust' to anomalies in order to overcome such situations. Robust statistics is concerned with deviations from the assumed model and the construction of reliable and sufficiently efficient statistical procedures when these deviations occur. The term 'Robust' was first used by Box (1953). Tukey (1960) observed that even small perturbations from the assumed model cause optimal procedures to rapidly lose their effectiveness, and Tukey (1962) has led the robust methods used today. Huber (1964) developed the M-estimator, a flexible and broad class of estimators, which has an important place in the development of robust statistical methods. Hampel (1968, 1974) introduced the Influence function, which is one of the most important tools in measuring the stability of a statistical procedure and has played an important role in the development of new robust methods. M-Estimators are frequently used in Theoretical and Applied Statistics, Econometrics and Biostatistics.

In the regression analysis, the anomalies in the data while estimating the parameters can cause to lose the effectiveness of the LS estimation method. In the presence of such data, parameter estimates made with OLS are will be biased (Hampel 1968). In robust statistics, traditional M-estimation methods do not consider the skewness and kurtosis parameters of the data, the PDE contains these values. Thus, the distribution of the data set can be determined uniquely, and the error is minimized while estimating the regression parameter.

The aim of this study is to construct a new method based on PVI that can be used instead of conventional M-estimators when data have anomalies. While traditional M-estimators usually achieve a good solution for symmetric and heavy long-tailed data, they lose effectiveness when anomalies arise. Therefore, in this study, regression parameter estimates will be estimated by using the weight function of the PVI, which contains the asymmetry, kurtosis and heavy long tail occurring in the data set. Based on the similarity between PDE and IF, Objective Function, Influence Function and Weight Function will be obtained by using the pdf of PVI.

This paper is organized as follows: Sect. 2 outlines the literature review about Robust Regression and Pearson Distribution System (PDS), Sects. 3 and 4 outlines the theoretical framework of Robust Regression and PDS, Sect. 5 presents the relationship between PDE and IF, and also for the proposed method, which based on PVI, obtained Objective, IF and Weight Function, Sect. 6 presents two real-world examples and simulation study with different scenarios, also discussion on obtained results for the proposed method. In the last section, we discuss the advantages and disadvantages of the proposed method. This paper also has an Appendix section, which contains proof of the tail properties of PVI.

2 Literature Review

Robust regression analysis has been studied frequently in the literature, especially after the 1960s. After Tukey, Huber and Hampel, many researchers have been interested in robust regression analysis. To summarize briefly, Harvey (1977) suggested using the minimum absolute deviation estimator as an initial solution in the robust regression procedure. M-estimators based on the median developed by Hinich and Talwar (1975) and Andrews (1974) were also used as the initial solution. Hogg (1979) discussed the robust statistical procedures used to reduce the effects of outliers in the data set. He examined the estimation processes of regression parameters and focused on the IRWLS method, which is the method used to estimate regression parameters, and discussed the asymptotic variance formula. He discussed the data set, which is reported by Andrews (1974) and analyzed by Wood and Gorman (1971). In addition, he used M-Estimator for analysis of the data sets which 'Half-life of Plutonium-241' by Zeigler and Ferris (1973) and 'Splines' by Lenth (1977), and 'Automated data reduction' by Agee and Turner (1978). Wu (1985) discussed commonly used M-estimators for scale and regression parameters. He compared the Bell/OLS M-Estimators developed by Bell (1980) and the high breakdown point Bell/RM M-estimators developed by Siegel (1982) using several real data sets. He discussed the similarities between Tukey Bisquare M-estimator and the Bell/OLS M-estimators. Croux and Reusseeuw (1992) developed two robust scale estimates, S_n and Q_n. They focused on breakdown points and computational algorithms for the developed estimators. They compared these estimators according to calculation time. They also used these scale estimates while estimating regression parameters. Cantoni and Ronchetti (2006) have developed a new robust method to be used in skewed and heavy-tailed data. They proved that when there are deviations from the assumed model, the method they developed is more efficient than traditional methods. They demonstrated the efficiency of the method they developed by using "medical back problems" data obtained from 100 patients in a hospital in Switzerland and many simulation studies.

Allende et al. (2006) proposed an M-estimation method with an asymmetric influence function based on the G_A^0 distribution. They used the developed method to process images obtained from satellite (GPS). Mohebbi et al. (2007) examined the robust regression methods that are an alternative to LS. They compared Least Absolute Deviation (LAD), Huber and nonparametric regression methods using skewed data sets. They used MSE and TAB as comparison criteria. Chen (2013) suggested using the distributed (clustered) IRWLS estimation method, when the data set is very large. Rasheed et al. (2014) used IRWLS to estimate regression parameters in the presence of outlier or heteroscedasticity in the data set. They also compared M-Estimator, LS and Least Trimmed Squares (LTS) methods using different data sets. Khalil et al. (2016) proposed a redescending M-estimator. He compared this estimator with the Hampel, Andrews, Tukey and Qadir M-estimators. In addition to many simulation studies, they compared the methods using the data set of international telephone calls from Belgium (Rousseeuw and Leroy 1987) between 1950 and 1973. Sumarni et al. (2017) studied the location parameter of the distribution as robust using the T distribution, which has a longer tail than the normal distribution and obtained the Objective, Influence and Weight functions of the T-distribution. They obtained the asymptotic behavior and Asymptotic Relative Efficiency (ARE) for location parameter. They examined how ARE changes using different degrees of freedom. Yulita et al. (2018) compared the weight functions of Huber, Hampel, Tukey and Welsch using simple and multivariate regression analysis. They used many simulations and Human Development Index (HDI) data from India East Java Region for comparison. Considering the literature for PDS, Pearson Differential Equation, first introduced by Karl Pearson (1895), is a system that generates different probability distributions according to the different values of the parameters in the PDE. This system is called the Pearson Distribution Family (PDF) and includes 13 different distributions with 3 are main types and the 10 transition types. The main types of PDF:

- Pearson Type I Distribution (PI),
- Pearson Type IV Distribution (PIV),
- Pearson Type VI Distribution (PVI).

PI (Four Parameter Beta Distribution) is a limited distribution from both tails. The PIV, on the other hand, is a distribution whose roots are complex, but it is unlimited at both tails. The PVI distribution (Beta Distribution the Second Type) is a heavy long-tailed distribution (see Appendix.) limited in one tail (right or left). PVI contains F, Pareto, Beta and Gamma Distributions according to the values of the parameters of the distribution. In addition, due to the structure of its parameters, it can be used easily in many kurtosis and skewness values. Mainly used areas as follows:

- Loss Function (Balkema and Embrechts 2018),
- Examination of Brain Functions (Brascamp et al. 2004),
- Modeling in Epidemic Diseases (Tulupyev et al. 2013),
- Meteorology and Hydrology (Mielke and Johnson 1974),
- Financial Volatility (Moghaddam et al. 2019),
- Income Modeling (Ye et al. 2012),
- Processing of Radar Images (Salazar 2000) and
- Reliability Analysis (Kilany 2016).

3 Robust Regression

The development of robust methods has led to significant improvements in regression analysis as in all other statistical methods. Especially when the data contain outliers, it has become inevitable to use robust methods. It has been difficult for researchers to define an observation in the data set as an outlier. Barnett and Lewis (1984) stated for the outliers as "inconsistent observations for the rest of the data set". Judge et al. (1988) called the large values in regression residuals the outliers. According to Hampel et al. (1986) and Krasker et al. (1983), outliers are divided into two groups as gross errors and model errors. Gross errors are errors due to recording, writing, failure of measuring equipment, unit change or misinterpretation. Even a small number of gross errors in the data set cause a tremendous change in traditional LS estimators. In the presence of such situations, it is of great importance to use robust statistical methods. Model error may occur due to the structure of the statistical/econometric model, such as misinterpretation of a variable or removed variable, which is the great contribution of the model.

In Fig. 1, the observations within the black circle are vertical (y-direction) outliers. The values of these observations x_i are close to rest of the data. However, these values do not follow the linear relationship that most of the data have. The observations within the red circle are points of "good leverage". They have the linear relationship that most of the data but have great x_i values. Contrary to its good name, they have a great influence on the LS estimators. Observations within the green circle are points of "bad leverage". They have large x_i values and do not fit most of the data set. They have a tremendous influence on the LS estimators. They could be gross errors.

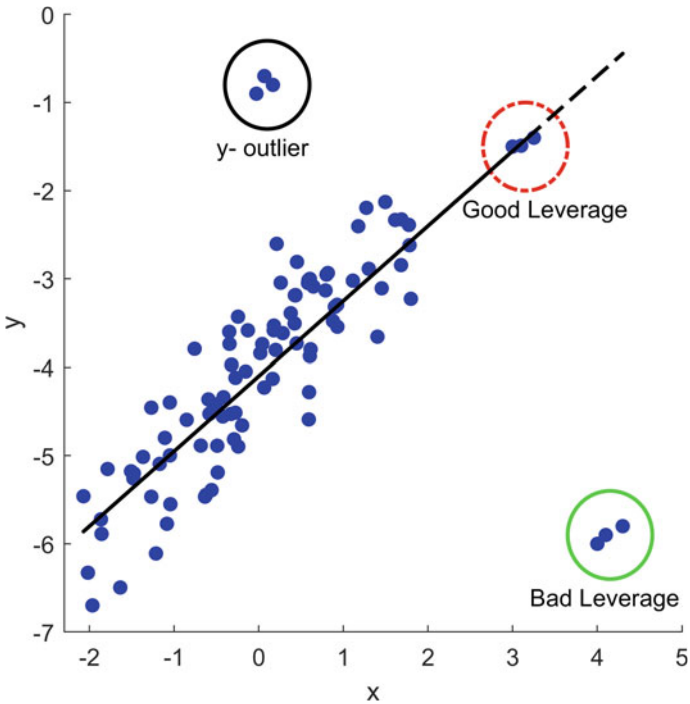


Fig. 1 Regression outliers

It is important to use robust methods in cases such as gross errors or model errors to minimize the effects of these errors. Features that robust regression estimators should have:

- If there are no outliers in the data and the distribution is normal, it should have a good performance as LS.
- When the first condition is not met, should have a better performance than the LS.
- Understanding the theory should be at least as easy as the LS method.
- It should be insensitive to trivial perturbations in the data.
- It should be easily calculated (Ryan 2008; Staudte and Sheather 2011).

3.1 Regression M-Estimator

If the distributions of the errors are heavy-tailed or there are outliers in residuals, parameter estimates made by LS will be biased (Hampel et al. 1986). Many researchers use robust methods to overcome such problems arise. One of the most popular robust methods is M-Estimators, which is based on ML proposed by Huber (1964) (Stuart 2011; Andersen 2008).

Consider the linear regression model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{y}$ is an $n \times 1$ response vector, $\boldsymbol{\theta}$ is an $p \times 1$ unknown regression parameters, $\mathbf{X}$ is an $n \times 1$ explanatory variable matrix and $\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-\mathbf{1}}$ is of full rank and $\boldsymbol{\varepsilon}$ is an $n \times 1$ error vector. In the classical LS method minimizing sum of squares:

$$\underset{\boldsymbol{\theta}\in\mathbb{R}^p}{\arg\min} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\theta}\right)^2 \tag{2}$$

Differentiating Eq. (2) with respect to $\boldsymbol{\theta}$ and system of p equations can be obtained:

$$\sum_{i=1}^{n} \left(y_i - \mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\theta}\right)x_{ij} = 0 \tag{3}$$

Solving Eq. (3) with respect to $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X}^\mathbf{T}\mathbf{X}\right)^{-\mathbf{1}}\mathbf{X}^\mathbf{T}\mathbf{y} \tag{4}$$

In Robust Regression Analysis, we can maximize or minimize the different functions or distributions of errors instead of minimizing the sum of squares of errors:

$$\sum_{i=1}^{n} \rho\left(y_i - \mathbf{x_i^T}\boldsymbol{\theta}\right) = \min! \tag{5}$$

where $\rho = -\ln f(x)$ and can be defined as Objective Function. (Susanti and Pratiwi 2014). Differentiating Eq. (5) with respect to $\boldsymbol{\theta}$:

$$\sum_{i=1}^{n} \psi\left(y_i - \mathbf{x_i^T}\boldsymbol{\theta}\right)x_{ij} = 0 \tag{6}$$

where $\psi(.)$ is Influence or Score Function. Solving Eq. (6) and obtaining i-th residuals is $e_i = y_i - \mathbf{x_i^T}\hat{\theta}$, one can rewrite the Objective and Influence Function as follows, respectively:

$$\min \sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) \tag{7}$$

$$\sum_{i=1}^{n} \psi(r_i)x_{ij} = 0 \tag{8}$$

where $r_i = e_i/s$ and s is the estimation of standard deviation (σ) must be the use for scale equivariance. Even if there are many different s estimates, the Median Absolute Deviation (MAD), which is not affected by outliers, is the most widely used for scale estimation (Draper and Smith 2014). MAD can be written as:

$$s = MAD/0.6745 = median|e_i - median(e_i)|/0.6745 \tag{9}$$

where 0.6745 is correction constant for the data actually normal (Hogg 1979).

If Eq. (8) can be written as a weighted LS estimation problem:

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X^T W X}\right)^{-1}\mathbf{X^T W y} \tag{10}$$

where $w_i = \psi(r_i)/r_i$ and $\mathbf{W} = diag\{w_i,\ i = 1, ..., n\}$ is a $n \times n$ weight matrix (Huber and Ronchetti 1981). The Weighted Least Squares method is usually used when solving Eq. (10). However, in order to calculate the weights, the first solution should be made with an appropriate method (usually LS) and the weights should be calculated with the weight function of the selected M-Estimators.

In the regression analysis, the distribution of errors is generally assumed to be Gaussian. If the distribution of the errors is actually normal, MLE and LS methods are the same. M-Estimators make parameter estimation using a different distribution or arbitrary function when the error distribution is different from Gaussian (skewed, heavy long-tailed, excess kurtosis, etc.). In this respect, LS and M-Estimator methods can be said to be MLE estimators (Rousseeuw and Leroy 1987; Andersen 2008). In this study, Huber and Tukey M-Estimators, which are the most popular M-Estimators, will be discussed.

3.1.1 Huber M-Estimator

Huber (1964) proposed an M-Estimator that consists of Objective and Influence Function, which is the most popular robust estimation method. The most important characteristic of Huber Objective Function is that it acts like Gaussian distribution in center and Laplace distribution in tails (Hogg 1979). Objective Function, Influence Function and Weight Function of Huber M-Estimator are given in Table 1.

In Table 1, k is tuning constant and default value is 1.345 for 95% efficiency under normal distribution. According to the weight function, weights, the observations in the center of the distribution are equal and 1, while inversely proportional to the absolute value of the observations as they move away from the center (Fox and Weisberg 2002: 3). The graphs of Objective, Influence and Weight Function of Huber M-Estimator can be seen in Fig. 2a.

3.1.2 Tukey's (Bisquare) M-Estimator

Tukey M-Estimator or Tukey's Bi-Weight (Bisquare) based on weight function was first proposed by Beaton ve Tukey (1974). Objective Function, Influence Function and Weight Function of Tukey M-Estimator are given in Table 2.

Table 1 Objective function, influence function and weight function of Huber M-Estimator

$\rho(r) = \begin{cases} \frac{1}{2}r^2 & , \lvert r\rvert < k \\ k\lvert r\rvert - \frac{1}{2}k^2 & , \lvert r\rvert \geq k \end{cases}$	$\psi(r) = \begin{cases} r & , \lvert r\rvert < k \\ k sign(r) & , \lvert r\rvert \geq k \end{cases}$	$w(r) = \begin{cases} 1 & , \lvert r\rvert < k \\ \frac{k}{\lvert r\rvert} & , \lvert r\rvert \geq k \end{cases}$
Objective function	Influence function	Weight function

Table 2 Objective function, influence function and weight function of Tukey M-Estimator

$\rho(r) = \begin{cases} \frac{k^2}{6}\left\{1 - \left[1 - \left(\frac{r}{k}\right)^2\right]^3\right\} & , \lvert r\rvert < k \\ \frac{k^2}{6} & , \lvert r\rvert \geq k \end{cases}$	$\psi(r) = \begin{cases} r\left[1 - \left(\frac{r}{k}\right)^2\right]^2 & , \lvert r\rvert < k \\ 0 & , \lvert r\rvert \geq k \end{cases}$	$w(r) = \begin{cases} \left[1 - \left(\frac{r}{k}\right)^2\right]^2 & , \lvert r\rvert < k \\ 0 & , \lvert r\rvert \geq k \end{cases}$
Objective function	Influence function	Weight function

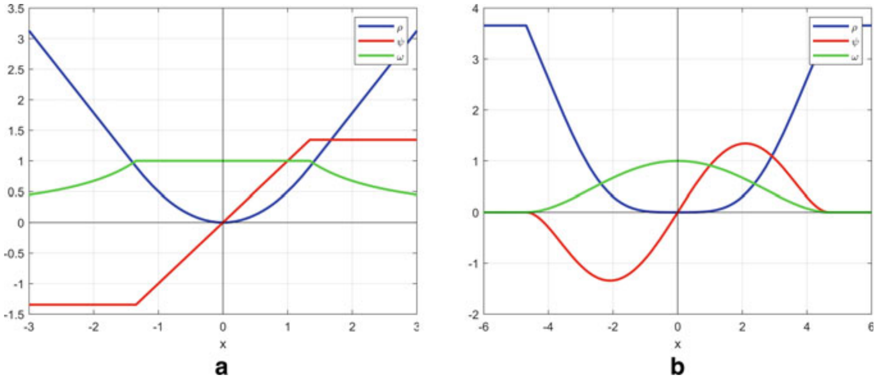


Fig. 2 **a** Huber M-estimator and **b** Tukey M-estimator

As the Huber M-Estimator, Tukey M-Estimator also has a tuning constant k, which its default value is 4.685 for 95% efficiency under normal distribution. The graphs of Objective, Influence and Weight Function of Tukey M-Estimator can be seen in Fig. 2b.

4 Pearson Distribution System

The Pearson Differential Equation (PDE) was first proposed by Karl Pearson (1895):

$$\frac{f'(x)}{f(x)} = \frac{d \ln f(x)}{dx} = \frac{x - a}{c_0 + c_1 x + c_2 x^2} \tag{11}$$

The solution of Eq. (11) defines Pearson Distribution Family (PDF), which consists of 13 different distributions with 3 main types and 10 transitional types. In Eq. (11), the parameter a is the mode of distribution and the parameters $a, c_0, c_1, c_2, \ldots$ can define type of the distribution uniquely. The function $C(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \ldots$, in the dominator of the differential equation, defining as a polynomial allows the Method of Moments (MoM) can be used in estimating the unknown parameters of the differential equation (Şehirlioğlu and Dündar 2014).

If we solve Eq. (11), the consecutive moment equation is:

$$n c_0 \mu'_{n-1} - \{(n+1)c_1 - a\}\mu'_n - \{(n+2)c_2 + 1\}\mu'_{n+1} = 0 \tag{12}$$

Substituting the values for n = 0,1,2,3 in Eq. (12) for $\mu'_1 = 0$:

$$
\begin{aligned}
c_0 &= -\frac{\mu_2\left(4\mu_2\mu_4 - 3\mu_3^2\right)}{10\mu_4\mu_2 - 12\mu_3^2 - 18\mu_2^3} = -\frac{\sigma^2(4\beta_2 - 3\beta_1)}{10\beta_2 - 12\beta_1 - 18} \\
c_1 = a &= -\frac{\mu_3\left(\mu_4 - 3\mu_3^2\right)}{10\mu_4\mu_2 - 12\mu_3^2 - 18\mu_2^3} = -\frac{\sigma\sqrt{\beta_1}(\beta_2 + 3)}{10\beta_2 - 12\beta_1 - 18} \\
c_2 &= -\frac{2\mu_4\mu_2 - 3\mu_3^2 - 6\mu_2^3}{10\mu_4\mu_2 - 12\mu_3^2 - 18\mu_2^3} = -\frac{2\beta_2 - 3\beta_1 - 6}{10\beta_2 - 12\beta_1 - 18}
\end{aligned}
\tag{13}
$$

where $\beta_1 = \mu_3^2/\mu_2^3$ (Skewness) and $\beta_2 = \mu_4/\mu_2^2$ (Kurtosis) parameters. The roots of $C(x)$ determine the type of PDS. The three main types of PDS:

- Type I (PI): Roots are real and different signs.
- Tip IV (PIV): Roots are complex.
- Tip VI (PVI): Roots are real and same sign.

Another method that can be used to determine the distribution types is the Kappa (κ) criterion. The coefficient of Kappa is a statistics obtained by using the discriminant of the $C(x)$ function (Elderton 1906; Hald 2008; Fiori and Zenga 2009; Nagahara 2008). The discriminant of the $C(x)$ and κ coefficient (Pearson 1901) can be written as follows:

$$\Delta = c_1^2 - 4c_0c_2 \tag{14}$$

$$\kappa = \frac{c_1^2}{4c_0c_2} = \frac{\beta_1(\beta_2 + 3)^2}{4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)} \tag{15}$$

The distributions according to κ criteria are given in Table 3.

An important case for PDS is the origin of the distribution must be mode point ($a = 0$). For this reason, substitution $X = x - a$ in Eq. (11), the PDE can be rewritten as follows:

$$\frac{df(X)}{dX} = \frac{Xf(X)}{C_0 + C_1X + C_2X^2} = \frac{(x-a)f(x)}{c_0 + c_1(x-a) + c_2(x-a)^2} \tag{16}$$

By solving Eq. (16), one can easily obtain the parameters which mode point coincides with the origin. The new parameters, which is $a = 0$, can be written in

Table 3 Main types of PDS

Δ	κ Statistics	c_0	c_1	c_2	Roots	Type
$\Delta > 0$	$\kappa < 0$	$c_0 \neq 0$	$c_1 \neq 0$	$c_2 \neq 0$	Real	Type I
$\Delta > 0$	$\kappa > 1$	$c_0 \neq 0$	$c_1 \neq 0$	$c_2 \neq 0$	Real	Type VI
$\Delta < 0$	$0 < \kappa < 1$	$c_0 \neq 0$	$c_1 \neq 0$	$c_2 \neq 0$	Complex	Type IV

(Şehirlioğlu and Dündar 2014: 16)

terms of original parameters as follows:

$$
\begin{aligned}
c_2 &= C_2 \\
2ac_2 + c_1 &= a(2c_2 + 1) = C_1 \\
c_2 a^2 + c_1 a + c_0 &= c_0 + a^2(1 + c_2) = C_0
\end{aligned}
\tag{17}
$$

4.1 Pearson Type I Distribution (PI)

Pearson Type I Distribution (PI) is one of the main types of PDF. The roots of $C(x)$ must be different signs and the parameters should be $c_2 > 0$ and $c_0 < 0$. The PI is also known as Beta Distribution. The pdf of PI:

$$f(x) = K(r_1 - x)^{m_1}(x + r_2)^{m_2}, \quad r_2 < x < r_1 \tag{18}$$

and

$$K = \frac{1}{B(m_1 + 1; m_2 + 1)(r_1 + r_2)^{m_1+m_2+1}} \tag{19}$$

where $B(x, y)$ is a Beta Function, r_1 is scale, r_2 is location and m_1,m_2 skewness and kurtosis parameters and K is the constant of normalization to make sure $\int f(x)dx = 1$. For the integral constant K, see more details of Pearson (1895), Nagahara (2008) and Elderton (1953). Figure 3 shows pdfs of PI with different skewness and kurtosis parameters.

4.2 Pearson Type IV Distribution (PIV)

Pearson Type IV Distribution (PIV) is the hardest distribution in PDF. For the existence of PIV, the roots of $C(x)$ must be complex. The pdf of PIV:

$$f(x) = K\left[(x + r)^2 + s^2\right]^m e^{v \arctan \tau}, \quad -\infty < x < \infty \tag{20}$$

where $K = \frac{s^{-2m-1}}{\exp\left(\frac{v\pi}{2}\right) \int_{-\pi/2}^{\pi/2} (\cos\theta)^{-2m-2} \exp(-v\theta)d\theta}$, $m = 1/2c_2, v = (a+r)/sc_2, r = real(r_1)$ and $s = imag(r_1)$. Figure 4 shows pdfs of PIV with different skewness and kurtosis parameters.

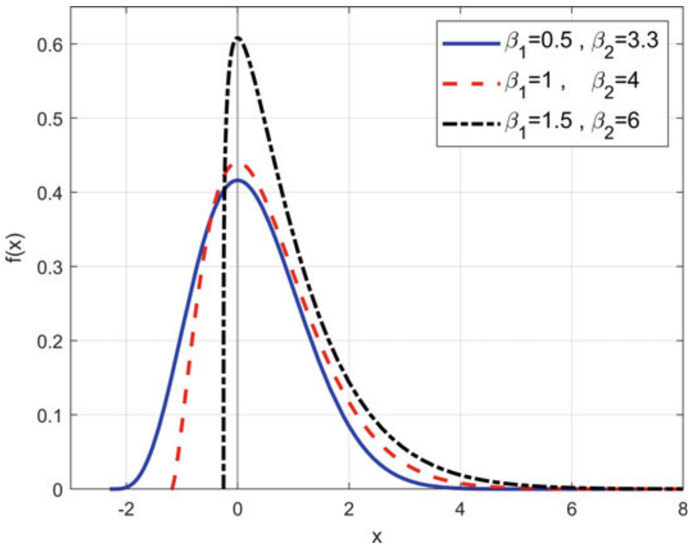


Fig. 3 Probability density functions of PI with different skewness and kurtosis

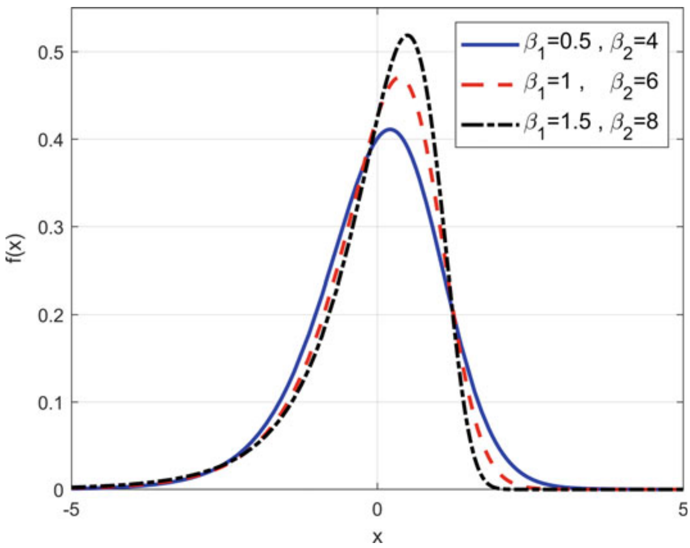


Fig. 4 Probability density functions of PIV with different skewness and kurtosis

4.3 Pearson Type VI Distribution (PVI)

Pearson Type VI Distribution (PVI) has the same sign of roots of $C(x)$. For $-r_2 < -r_1 < 0$ right skewed PVI:

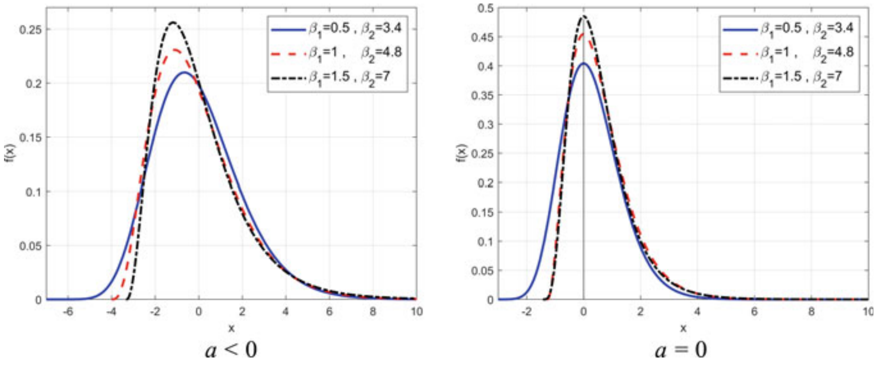


Fig. 5 Probability density functions of PVI with different skewness and kurtosis

$$f(x) = K(x + r_1)^{m_1}(x + r_2)^{m_2}, \quad -r_1 < x < \infty \tag{21}$$

and

$$K = \frac{1}{B(-m_1 - m_2 - 1; m_1 + 1)(r_2 - r_1)^{m_1+m_2+1}}. \tag{22}$$

where r_1 is scale, r_2 is location and m_1,m_2 skewness and kurtosis parameters. Use Eq. (16) for the PVI with mode at origin can be written as follows:

$$f(X) = K^*(X + R_1)^{M_1}(X + R_2)^{M_2}, \quad R_1 < R_2 < X < \infty \tag{23}$$

Figure 5 shows pdfs of PVI with different skewness and kurtosis parameters when a<0 and a=0.

5 Pearson Differential Equation as a Influence Function

The similarity between the Influence Function (IF) and the PDE, different IFs can be defined for the dynamic parameters of the PDE. Thus, regression parameter estimates can be made by using the Weight Function. Dzhun' (2011) shows the similarity between IF and PDE as follows:

$$\psi(x) = \frac{d\rho(x)}{dx} = \frac{d[-\ln f(x)]}{dx} = \frac{f'(x)}{f(x)} = \frac{x - a}{c_0 + c_1 x + c_2 x^2} \tag{24}$$

Using Eq. (24), different IFs can be easily obtained. (Wiśniewski 2014). The definition of Weight Function is $w(x) = \psi(x)\big/x$:

$$w(x) = \frac{x - a}{x\left(c_0 + c_1 x + c_2 x^2\right)} \tag{25}$$

The value of the weight function is usually closely related to the mode point of the distribution. If the data have skewness and excess kurtosis, the Weight Function should take its maximum value at the mode point. In such cases, Eq. (16) can be used for coincides to mode point and the origin. IF, where the mode point at the origin:

$$\psi(X) = \frac{X}{C_0 + C_1 X + C_2 X^2} \tag{26}$$

and the Weight Function:

$$w(X) = \frac{1}{C_0 + C_1 X + C_2 X^2} \tag{27}$$

In this study, we only consider PVI for estimating regression parameters. For this purpose, Objective, IF and Weight Function will only be obtained for PVI. Consider Eq. (23), if the constant term removed, the Objective Function of PVI:

$$\begin{aligned} &f(X) \propto (X + R_1)^{M_1} (X + R_2)^{M_2} \\ &\rho(X) = -\ln f(X) = -M_1 \ln(X + R_1) - M_2 \ln(X + R_2) \end{aligned} \tag{28}$$

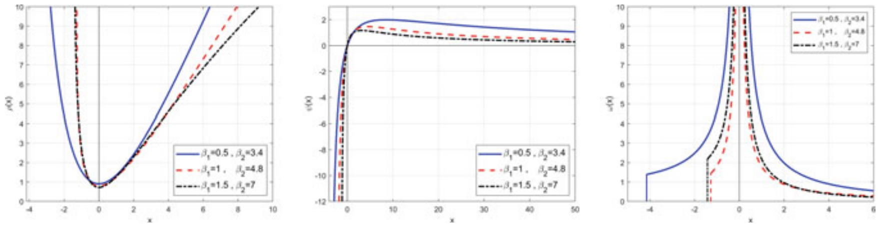
The Objective Function of PVI provides flexibility in terms of functions that will be minimized, due to its location, scale, skewness and kurtosis parameters compared with other robust methods. Based on Eq. (28), the IF and Weight Function of PVI:

$$\psi(X) = \frac{d\rho(X)}{dX} = \frac{d[-\ln f(X)]}{dX} = -\frac{M_1}{X + R_1} - \frac{M_2}{X + R_2} \tag{29}$$

and

$$w(X) = \frac{\psi(X)}{X} = \frac{M_1 R_2 + M_2 R_1 - (M_1 + M_2)X}{X(X + R_1)(X + R_2)} \tag{30}$$

Figure 6 shows Objective, Influence and Weight Functions of PVI with different skewness and kurtosis parameters.



Objective Function Influence Function Weight Function

Fig. 6 Objective, Influence and Weight Functions of PVI

6 Real Data Examples and Simulation Study

6.1 Real Data Examples

In the Real Data example, two different data sets are analyzed for the estimation method based on the weight function of the proposed PVI function. The first data set is Education Expenditure, which is commonly used in the robust literature, and the second data set is Industrial Production Index, Unemployment Rate and CPI values in Turkey. For the data sets, goodness of fit of the PVI function is applied, and standard errors of the regression parameters are obtained. Box-plot and histogram of residuals for data sets are drawn for residuals obtained from LS.

6.1.1 Education Expenditure Data

The data set published by Chatterjee and Price (1977) discussed Education Expenditures in 50 states in the United States. Information about the variables as follows: Average education expenditure per capita at public school in a state in 1975 (response variable)

- Number of residents residing in urban areas in 1970 (x1000),
- Per capita personel income in 1973,
- Number of residents under 18 years of age in 1974 (x1000).

In the data set, the value of 31st data of the response variable has been replaced with the value 400 instead of 212 due to a recording or transferring error. After that, PVI parameters were estimated and regression analysis was performed. When Figure 7a is examined, it can be seen that the 31st and 49th data points have high standardized residuals (3.69 and 3.22). Anderson-Darling goodness of fit test is applied to the residuals, and the test value is obtained as 1.046 (p = 0.000), and it is determined that the distribution of the residuals does not come from normal distribution. Also, the skewness is 1.10, and kurtosis is 5.30 of residuals obtained from LS.

In Table 4, the PVI method gives similar results with other estimation methods even at high skewness and kurtosis values. Although Huber and Tukey M-Estimation

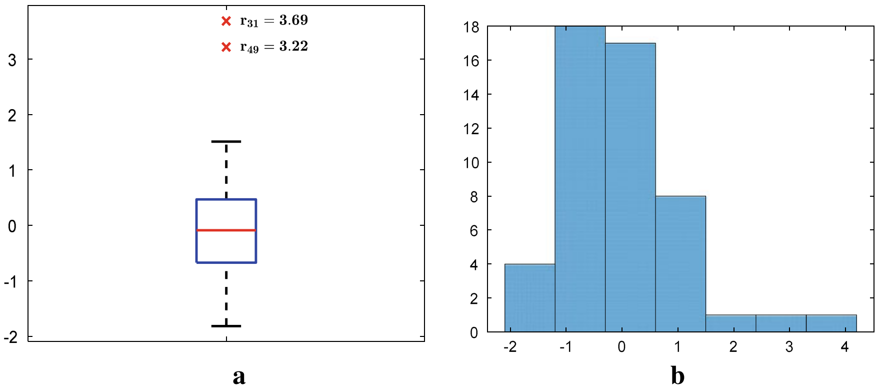


Fig. 7 Box plot and histogram of education expenditure data

Table 4 Result of education expenditure data

				Variables	
Estimation method	Results	Constant	Number of residents residing in urban areas (x1000)	Per capita personel income	Number of residents under 18 years of age (x1000)
LS	Estimate	−452.203	0.001	0.063	1.346
	Standard Error	146.891	0.061	0.014	0.375
Huber M-estimator	Estimate	−340.860	0.044	0.053	1.067
	Standard error	126.629	0.053	0.012	0.323
Tukey M-estimator	Estimate	−243.762	0.074	0.045	0.820
	Standard Error	127.490	0.053	0.012	0.326
PVI	Estimate	−307.984	0.009	0.061	0.909
	Standard error	72.326	0.030	0.007	0.185

methods give very low or 0 weight when residuals increase, the PVI method analyzed high standardized residuals as a part of the data set and weighed all residuals. Thus, when making regression diagnostic, when there is an observation that seems to be a outlier but is known to belong to the data set and even if the skewness and kurtosis values are high, the PVI method gives results at least as well as other estimation methods.

6.1.2 Economical Data

For the Economic data set, between January 2015 and February 2020, the values of 62 month Industrial Production Index (2015 = 100), Unemployment Rate and

Consumer Price Index (2003 = 100, response) are considered. In this example, by considering the variables commonly used in economic analysis in Turkey, we have been focused on the importance of determining the exact distribution of the data. When Fig. 8a is examined, it can be seen that observations 45th, 46th, 47th, 48th, 49th, 52th and 57th have high standardized residuals. In addition, if the histogram of the residuals (Fig. 8a) is examined, it is seen that the data are right skewed ($\beta_1 = 1.08$) and have excess kurtosis ($\beta_2 = 5.04$). As a result of the Anderson–Darling test, the test statistic is 1.782 ($p = 0.005$) and the distribution of the residuals is not normal.

According to the regression analysis results in Table 5, PVI method gives very similar results to other methods. Although the skewness and kurtosis values of the data are high and there are high standardized residuals, the coefficients were obtained significantly.

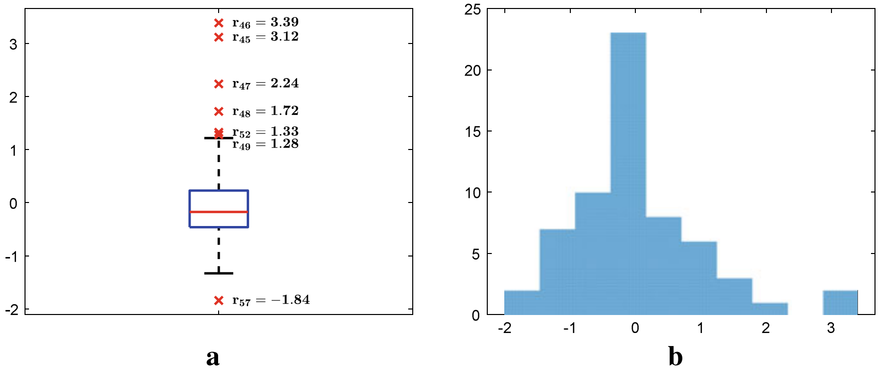


Fig. 8 Box plot and histogram of economical data

Table 5 Result of economical data

		Variables		
Estimation method	Results	Constant	Unemployment rate	Industrial production index (2015 = 100)
LS	Estimate	−29.615	0.256	1.143
	Standard error	8.052	0.077	0.338
Huber M-estimator	Estimate	−28.035	0.250	1.012
	Standard error	6.716	0.064	0.282
Tukey M-estimator	Estimate	−17.680	0.236	0.644
	Standard error	4.722	0.045	0.128
PVI	Estimate	−25.315	0.247	0.781
	Standard error	8.383	0.080	0.352

6.2 Simulation Study

In the simulation study, LS, Huber M-Estimator, Tukey M-Estimator and proposed method based on PVI are compared with different scenarios. Sample sizes *30, 50, 100* and *500* are used and *M=1000* replications are simulated. Data are generated multivariate linear regression using following model;

$$y = 2 + 2X_1 + 2X_2 + 2X_3 + 2X_4 + 2X_5 + \varepsilon, \quad X_i \sim N(0, 1) \tag{31}$$

Error term is generated 13 different scenarios and 7 distributions with 2 symmetrical and 5 asymmetrical distributions. Predetermined scenarios of the error term are as follows:

Scenario 1. $\varepsilon \sim N(0, 1)$, Standard Normal Distribution (PXIII).
Scenario 2. $\varepsilon \sim t(1)$, (PVII).
Scenario 3. $\varepsilon \sim t(10)$, (PVII).
Scenario 4. $\varepsilon \sim Exp(1)$, (PX).
Scenario 5. $\varepsilon \sim Exp(10)$, (PX).
Scenario 6. $\varepsilon \sim Gamma(1, 5)$, (PIII).
Scenario 7. $\varepsilon \sim Gamma(2, 5)$, (PIII).
Scenario 8. $\varepsilon \sim \chi^2(1)$, (PIII).
Scenario 9. $\varepsilon \sim \chi^2(5)$, (PIII).
Scenario 10. $\varepsilon \sim F(2, 10)$, (PVI).
Scenario 11. $\varepsilon \sim F(10, 10)$, (PVI).
Scenario 12. $\varepsilon \sim Weibull(1, 1)$,
Scenario 13. $\varepsilon \sim Weibull(2, 2)$.

In each scenario, residuals are obtained from LS estimation method and we choose response and explanatory variables, which is suitable for PVI. Initial weights are calculated for Huber M-Estimator and Tukey M-Estimator using the residuals obtained from LS and for the PVI method mode point must coincide at origin. We use Iteratively Re-Weighted Least Squares (IRWLS) for estimating regression parameters and calculate Total Absolute Bias (TAB) and Mean Squared Error (MSE) for comparing estimation methods. The calculating steps of IRWLS as follows (Fox and Weisberg 2002; Maronna et al. 2006; Hogg 1979):

1. Choose a suitable estimation method (usually LS) and estimate regression parameters.

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y}.$$

2. Calculate residuals using the estimated regression parameter.

$$\mathbf{e} = y - X\hat{\boldsymbol{\theta}}.$$

3. Estimate robust standard deviation (Usually MAD).
4. Calculate studentized residuals using MAD and use tuning constant if exist.

$$r_i = e_i \Big/ \left(ks\sqrt{(1 - h_i)}\right)$$

5. Choose a weight function from Table 1c, Table 2c or Eq. (30).
6. Estimate new regression parameters using Weighted LS method.

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}.$$

7. Repeat 2–6 until converge:

$$\left\|\hat{\theta}^i - \hat{\theta}^{i-1}\right\| \Big/ \left\|\hat{\theta}^i\right\| < 10^{-6} ori > 30.$$

While comparing the estimation methods, we can calculate TAB and MSE using regression parameters as follows (Wiśniewski 2014);

$$TAB = \sum_{j=0}^{5}\left|\frac{1}{M}\sum_{i=1}^{M}\hat{\theta}_{ij} - \theta_{ij}\right|$$

$$MSE = \frac{1}{M}\sum_{i=1}^{M}\left(\hat{\theta}_i - \theta\right)^T\left(\hat{\theta}_i - \theta\right).$$

Table 6 shows TAB and MSE values for each estimation method and sample sizes when the error term follows both symmetric and asymmetrical distribution. Considering these values, LS method gives the best results when the distribution of the error term is normal distribution. However, when the distribution of the error term is *Student t* distribution, which has heavy and long-tailed than the normal distribution, Huber and Tukey M-Estimators give better results than both LS and PVI. For some distributions in simulations, PVI is not suitable in large sample sizes. In every case of the error term follows asymmetrical distribution, LS, Huber M and Tukey M-Estimators give worse results than PVI since they do not consider skewness and excess kurtosis of data set. The PVI weight function ensures a flexible structure according to the parameters of the distribution.

According to the simulation results, in the presence of skewness and excess kurtosis in the data, the PVI method has the best TAB and MSE values since the weight function takes these parameters into account. However, when the distribution of errors is symmetrical, the PVI method gives worse than other methods.

7 Conclusion

Researchers use many estimation methods in Robust Regression Analysis. The most important and most widely used robust estimation method is M-Estimators. However, traditional M-Estimators do not consider the anomalies (heavy-long tail, skewness

Table 6 Simulation results

Simulation study			Symmetrical distributions			Asymmetrical distributions										
Sample size	*Method*	*Criteria*	*Normal (0,1)*	t_1	t_{10}	*Exp (1)*	*Exp (10)*	*Gamma (1,5)*	*Gamma (2,5)*	χ^2_1	χ^2_5	*F (2,10)*	*F (10,10)*	*Weibull (1,1)*	*Weibull (2,2)*
30	***LS***	***TAB***	**0.0303**	15.621	0.0922	10.178	101.781	50.819	101.022	10.416	50.487	13.022	12.908	10.187	17.646
		MSE	**0.2589**	230.663	0.3453	12.658	12.657	315.408	1138	16.044	274.566	21.606	18.987	12.668	32.608
	Huber M-Estimator	***TAB***	0.0808	0.2966	**0.0617**	0.8284	82.837	41.904	89.912	0.7200	46.236	0.9670	11.268	0.8270	16.791
		MSE	0.2902	18.084	**0.3267**	0.8349	834.860	210.001	904.979	0.7256	227.205	11.523	13.736	0.8381	29.843
	Tukey M-Estimator	***TAB***	0.1278	**0.1276**	0.1117	0.7569	75.692	38.179	85.701	0.6027	44.150	0.8499	10.614	0.7502	16.346
		MSE	0.3437	**11.645**	0.3769	0.7206	720.546	180.047	831.000	0.5310	209.838	0.9246	12.138	0.7172	28.650
	PVI	***TAB***	0.1651	0.0722	0.1427	**0.6891**	**68.899**	**35.089**	**81.770**	**0.5276**	**43.171**	**0.7715**	**10.215**	**0.6917**	**16.053**
		MSE	0.3957	13.141	0.4490	**0.6547**	**654.653**	**168.198**	**789.426**	**0.4956**	**203.966**	**0.8434**	**11.770**	**0.6737**	**28.109**
50	***LS***	***TAB***	**0.0171**	0.8723	0.0717	10.124	101.240	50.742	100.975	10.230	50.890	12.630	12.662	10.232	17.723
		MSE	**0.1366**	36.184	0.1842	11.453	1145.326	283.978	1076	12.796	265.389	18.317	16.900	11.371	31.984
	Huber M-Estimator	***TAB***	0.0644	0.2315	**0.0330**	0.8290	82.895	41.691	91.359	0.6916	46.795	0.9635	11.113	0.8441	16.951
		MSE	0.1539	0.7217	**0.1749**	0.7662	766.194	191.035	875.446	0.5761	223.527	10.395	12.744	0.7636	29.370
	Tukey M-Estimator	***TAB***	0.0933	**0.0697**	0.0682	0.7588	75.879	38.373	88.001	0.5603	45.095	0.8466	10.521	0.7765	16.646
		MSE	0.1760	**0.5445**	0.1928	0.6577	657.656	164.799	813.066	0.4031	209.263	0.8245	11.409	0.6565	28.379
	PVI	***TAB***	0.1438	0.0867	0.1268	**0.6682**	66.823	**33.031**	**80.767**	**0.4618**	**42.211**	**0.7286**	**0.9710**	**0.6834**	**15.990**
		MSE	0.2186	0.5659	0.2452	**0.5368**	**536.811**	**132.652**	**712.947**	**0.3061**	**187.777**	**0.6531**	**10.071**	**0.5333**	**26.575**
100	***LS***	***TAB***	**0.0129**	–	0.0525	10.139	**101.394**	50.372	100.787	10.223	50.460	12.355	12.545	10.082	17.708
		MSE	**0.0648**	–	0.0802	10.663	1066.307	261.131	1032	11.186	255.428	16.233	15.769	10.613	31.551
	Huber M-Estimator	***TAB***	0.0445	–	**0.0228**	0.8412	84.116	41.620	91.534	0.6817	46.692	0.9512	11.150	0.8385	17.018
		MSE	0.0711	–	**0.0755**	0.7300	729.949	179.442	853.479	0.4995	218.271	0.9471	12.339	0.7282	29.212

(continued)

Table 6 (continued)

Simulation study			Symmetrical distributions			Asymmetrical distributions									
	Tukey M-Estimator	***TAB***	0.0584	–	0.0417	0.7853	78.534	38.809	88.876	0.5540	45.643	0.8509	10.673	0.7850	16.829
		MSE	0.0764	–	0.0789	0.6420	641.947	157.784	808.076	0.3442	208.684	0.7636	11.334	0.6413	28.590
	PVI	***TAB***	0.1177	–	0.1180	**0.6485**	**64.850**	**32.109**	**79.096**	**0.4249**	**41.509**	**0.6935**	**0.9655**	**0.6432**	**15.901**
		MSE	0.1058	–	0.1130	**0.4534**	**453.435**	**112.153**	**659.232**	**0.2209**	**174.865**	**0.5346**	**0.9347**	**0.4505**	**25.913**
500	***LS***	***TAB***	–	–	0.0334	10.067	100.666	50.170	100.442	10.044	50.192	12.143	12.338	10.045	17.726
		MSE	–	–	0.0153	10.105	1010.528	252.585	1005	10.205	250.864	14.840	15.199	10.113	31.369
	Huber M-Estimator	***TAB***	–	–	**0.0189**	0.8418	84.177	42.093	91.909	0.6648	46.733	0.9429	11.027	0.8438	17.115
		MSE	–	–	**0.0144**	0.7115	711.460	177.435	841.378	0.4464	217.276	0.8946	12.119	0.7113	29.273
	Tukey M-Estimator	***TAB***	–	–	0.0254	0.7961	79.612	39.860	90.334	0.5433	46.163	0.8541	10.674	0.7995	17.061
		MSE	–	–	0.0146	0.6382	638.232	159.041	813.301	0.2987	211.924	0.7340	11.353	0.6381	29.098
	PVI	***TAB***	–	–	0.0934	**0.6209**	**62.092**	**31.087**	**78.784**	**0.3742**	**41.184**	**0.6426**	**0.9308**	**0.6250**	**15.993**
		MSE	–	–	0.0266	**0.3910**	**390.964**	**97.447**	**619.100**	**0.1474**	**169.547**	**0.4249**	**0.8639**	**0.3925**	**25.656**

and kurtosis) found in the data set. In this study, based on the similarity between PDE and IF, when the distribution of the data follows the PVI distribution, the performance of the M-Estimator based on the weighting function of the PVI distribution is considered. If the error term has symmetrical distributions, the proposed method gives worse performance than other methods. However, in cases where the distribution of the error term is asymmetric ($\beta_1 > 0$) and excess kurtosis ($\beta_2 > 3$), the weight function of the PVI distribution has a better performance than other M-Estimators since it contains these parameters. Also, due to the heavy and long tails (see Appendix) of the PVI, it performs well in asymmetrical distributions with light tails. Asymmetrical distributions used in simulations have a lighter tail than PVI.

The performance of the PVI has been analyzed by performing a simulation study on 13 scenarios with 7 different distributions with two different real-world data.

Appendix

Tail Properties of a Distribution

The Tail Function of a distribution G can be defined as follows:

$$\overline{G}(x) = G(x, \infty), \quad x \in R$$

Considering the tail function or pdf, it can be determined whether a distribution has a Heavy-Tailed, Fat-Tailed or Long-Tailed by considering the following three conditions (Bryson 1974; Foss et al. 2011).

Condition 1. Heavy-Tailed Distributions

G can be defined as Heavy-Tailed distribution, it must be satisfied;

$$\int_R e^{\lambda x} G(dx) = \infty, \quad \forall \lambda > 0$$

If the Heavy Tailness has written for pdf;

$$\lim_{x \to \infty} \sup g(x) e^{\lambda x} = \infty, \quad \forall \lambda > 0$$

Condition 2. Long-Tailed Distributions

G can be defined as Long-Tailed distribution, it must be satisfy;

$$\lim_{x\to\infty} \frac{g(x+\lambda)}{g(x)} = 1, \quad \forall\lambda > 0$$

A Long-Tailed distribution is subclass of Heavy-Tailed distributions. By using Tail Function;

$$\lim_{x\to\infty} \overline{G}(x+\lambda) \sim \overline{G}(x), \quad \forall\lambda > 0$$

Condition 3. Fat-Tailed Distributions

G can be defined as Fat-Tailed distribution, it must be satisfy;

$$\lim_{x\to\infty} P(X > x) \sim x^{-\lambda}, \quad \forall\lambda > 0$$

where $P(X > x) = \overline{G}(x)$.

In this study, we investigated tail properties of PVI. Consider Eq. (23);

Proof of Condition 1

By using pdf of PVI and Condition 1, one can easily calculate;

$$\lim_{x\to\infty} e^{\lambda x} f(x) = \infty$$

Thus PVI is a Heavy Tailed Distribution

Proof of Condition 2

For any $\lambda > 0$, we obtain the equation $f(x+\lambda) = (x + r_1 + \lambda)^{m_1}(x + r_2 + \lambda)^{m_2}$. By using Condition 2;

$$\lim_{x\to\infty} \frac{f(x+\lambda)}{f(x)} = 1$$

Thus, PVI is a Long-Tailed Distribution.

Proof of Condition 3

The Tail Function of PVI;

$$P(X > x) = 1 - \frac{x^{m_1+1}{}_2F_1\begin{bmatrix} m_1+1 & -m_2 \\ & m_1+2 \end{bmatrix}; -x}{(m_1+1)B(-m_1-m_2-1; m_2+1)}$$

where ${}_2F_1\begin{bmatrix} a & b \\ & c \end{bmatrix}; z$ is Gauss Hypergeometric Function. By using the Condition 3;

$$\lim_{x\to\infty} \overline{F}(x) = -\infty$$

Thus PVI is not a Fat-Tailed Distribution.

References

Agee WS, Turner RH (1978). Application of robust statistical methods to data reduction (No. ACD-TR-65). National range operations directorate white sands missile range N Mex analysis and computation div

Allende H, Frery AC, Galbiati J, Pizarro L (2006) M-estimators with asymmetric influence functions: the distribution case. J Stat Comput Simul 76(11):941–956

Andersen R (2008). Modern methods for robust regression (No. 152). Sage, Thousand Oaks

Andrews DF (1974) A robust method for multiple linear regression. Technometrics 16(4):523–531

Balkema G, Embrechts P (2018) Linear Regression for Heavy Tails. Risks 6(3):93

Barnett V, Lewis T (1984) Outliers in statistical data (OSD)

Beaton AE, Tukey JW (1974) The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. Technometrics 16(2):147–185

Bell RM (1980). An adaptive choice of the scale parameter for M-estimators (No. TR-3). Stanford Unıv Ca Dept of Statıstıcs, Stanford

Box GE (1953) Non-normality and tests on variances. Biometrika 40(3/4):318–335

Brascamp JW, Berg AV, Ee R (2004) Shared neural circuitry for switching between perceptual states and ocular motor states? J vis 4(8):255–255

Bryson MC (1974) Heavy-tailed distributions: properties and tests. Technometrics 16(1):61–68

Cantoni E, Ronchetti E (2006) A robust approach for skewed and heavy-tailed outcomes in the analysis of health care expenditures. J Health Econ 25(2):198–213

Chatterjee S, Price B (1977) Regression analysis by example. Wiley, New York

Chen C (2013) Distributed iteratively reweighted least squares and applications. Stat Interface 6(4):585–593

Croux C, Reusseeuw PJ (1992) Time-efficient algorithms for two highly robust estimators of scale. In: Computational statistics. Physica, Heidelberg, pp 411–428

Draper NR, Smith, H (2014). Applied regression analysis, vol 326. Wiley, Hoboken

Dzhun' IV (2011) Method for diagnostics of mathematical models in theoretical astronomy and astrometry. Kinemat Phys Celest Bodies 27:260–264

Elderton WP (1906) Frequency curves and correlation. Cambridge University, New York

Elderton WP (1953) Frequency curves and correlation. Cambridge University, New York

Fiori AM, Zenga M (2009) Karl Pearson and the origin of kurtosis. Int Stat Rev 77(1):40–50

Foss S, Korshunov D, Zachary S (2011). An introduction to heavy-tailed and subexponential distributions, vol 6, pp 0090–6778. Springer, New York

Fox J, Weisberg S (2002) Robust regression. An R S Plus Companion Appl Regres 91

Hald A (2008) A history of parametric statistical inference from Bernoulli to Fisher, 1713–1935. Springer Science & Business Media, Berlin
Hampel FR (1974) The influence curve and its role in robust estimation. J Am Stat Assoc 69(346):383–393
Hampel FR, Ronchetti EM, Rousseuw PJ, Stahel WA (1986) Robust statistics. The approach based on influence functions. Wiley, New York
Hampel FR (1968) Contribution to the theory of robust estimation. PhD thesis, University of California, Berkeley
Harvey AC (1977) A comparison of preliminary estimators for robust regression. J Am Stat Assoc 72(360a):910–913
Hinich MJ, Talwar PP (1975) A simple method for robust regression. J Am Stat Assoc 70(349):113–119
Hogg RV (1979) Statistical robustness: one view of its use in applications today. Am Stat 33(3):108–115
Huber PJ (1964) Robust version of a location parameter. Ann Math Stat 36:1753–1758
Huber PJ, Ronchetti EM (1981) Robust statistics, ser. Wiley Ser Probab Math Stat New York, NY, USA, Wiley-IEEE 52:54
Judge GG, Hill RC, Griffiths WE, Lütkepohl H, Lee TC (1988) Introduction to the theory and practice of econometrics (No. 330.015195 I61 1988). Wiley, Hoboken
Khalil U, Ali A, Khan DM, Khan SA, Qadir F (2016) Efficient Uk'S re-descending M-estimator for robust regression. Pak J Stat 32(2)
Kilany NM (2016) Weighted Lomax Distribution. Springerplus 5(1):1862
Krasker WS, Kuh E, Welsch RE (1983) Estimation for dirty data and flawed models. Handb Econ 1:651–698
Lenth RV (1977) Robust splines. Commun Stat Theory Methods 6(9):847–854
Maronna RA, Martin D, Yohai RS (2006) Wiley series in probability and statistics. Robust Stat Theory Methods 404–414
Mielke PW Jr, Johnson ES (1974) Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology. Water Resour Res 10(2):223–226
Moghaddam MD, Liu J, Serota RA (2019) Implied and realized volatility: a study of distributions and the distribution of difference. arXiv preprint arXiv:1906.02306
Mohebbi M, Nourijelyani K, Zeraati H (2007) A simulation study on robust alternatives of least squares regression. J Appl Sci 7(22):3469–3476
Nagahara Y (2008) A method of calculating the downside risk by multivariate nonnormal distributions. Asia Pac Financ Mark 15(3–4):175–184
Pearson K (1895) X. Contributions to the mathematical theory of evolution.—II. Skew variation in homogeneous material. Philos Trans R Soc Lond A 186:343–414
Pearson, K. (1901). XI. Mathematical contributions to the theory of evolution.—X. Supplement to a memoir on skew variation. Philos Trans R Soc Lond A (Containing Papers of a Mathematical or Physical Character) 197(287–299):443–459
Rasheed BA, Adnan R, Saffari SE, dano Pati K (2014) Robust weighted least squares estimation of regression parameter in the presence of outliers and heteroscedastic errors. J Teknol 71(1)
Rousseeuw PJ, Leroy AM (1987) Robust regression and outlier detection, vol 1. Wiley, New York
Ryan TP (2008). Modern regression methods, vol 655. Wiley, New York
Salazar JSI (2000) Detection schemes for synthetic-aperture radar imagery based on a beta prime statistical model.
Şehirlioğlu AK, Dündar S (2014) Pearson Dağılış Ailesi. İzmir: Ege Üniversitesi Basımevi
Siegel AF (1982) Robust regression using repeated medians. Biometrika 69(1):242–244
Staudte RG, Sheather SJ (2011) Robust estimation and testing, vol 918. Wiley, New York
Stuart C (2011) Robust regression. Department of Mathematical Sciences, Durham University, Durham, p 169

Sumarni C, Sadik K, Notodiputro KA, Sartono B (2017, March). Robustness of location estimators under t-distributions: a literature review. In: IOP conference series: earth and environmental science, vol 58, no 1. IOP Publishing, Bristol, p 012015

Susanti Y, Pratiwi H (2014) M-estimation, S estimation, and MM estimation in robust regression. Int J Pure Appl Math 91(3):349–360

Tukey JW (1962) The future of data analysis. Ann Math Stat 33(1):1–67

Tukey JW (1960) A survey of sampling from contaminated distributions. Contrib Probab Stat 448–485

Tulupyev A, Suvorova A, Sousa J, Zelterman D (2013) Beta prime regression with application to risky behavior frequency screening. Stat Med 32(23):4044–4056

Wiśniewski Z (2014) M-estimation with probabilistic models of geodetic observations. J Geodesy 88(10):941–957

Wood FS, Gorman JW (1971) Fitting equations to data: computer analysis of multifactor data for scientists and engineers. Wiley-Interscience, New York

Wu LL (1985) Robust M-estimation of location and regression. Sociol Methodol 15:316–388

Ye Y, Oluyede BO, Pararai M (2012) Weighted generalized beta distribution of the second kind and related distributions. J Stat Econ Methods 1(1):13–31

Yulita T, Notodiputro KA, Sadik K (2018) M-estimation use bisquare, hampel, huber, and welsch weight functions in robust regression

Zeigler RK, Ferris Y (1973) Half-life of Plutonium-241. J Inorg Nucl Chem 35(10):3417–3418