

Article

Robust Total Least Squares Estimation Method for Uncertain Linear Regression Model

Hongmei Shi ¹, Xingbo Zhang ¹, Yuzhen Gao ¹, Shuai Wang ^{2,3,*} and Yufu Ning ^{2,3}

¹ School of Information Science and Engineering, Shandong Agriculture and Engineering University, Jinan 250100, China; z2013454@sdaeu.edu.cn (H.S.); z2013452@sdaeu.edu.cn (X.Z.); z2017043@sdaeu.edu.cn (Y.G.)

² School of Information Engineering, Shandong Youth University of Political Science, Jinan 250103, China; nyf@sdyu.edu.cn

³ New Technology Research and Development Center of Intelligent Information Controlling in Universities of Shandong, Jinan 250103, China

* Correspondence: wangshuai@sdyu.edu.cn

Abstract: In data analysis and modeling, least squares and total least squares are both mathematical optimization techniques. It is noteworthy that both the least squares method and the total least squares method are used to deal with precise and random data. However, when the given data are not random, or when the data are imprecise, and only the range of the data is available, the traditional linear regression method cannot be used. This paper presents an uncertain total least squares estimation method and an uncertain robust total least squares linear regression method based on uncertainty theory and total least squares method. The uncertain total least squares estimation can fully consider errors in the given data and the uncertain robust total least squares linear regression method can effectively eliminate outliers in the data. It is possible to obtain a more reasonable fitting effect with both of these methods, as well as to solve the predicted value and the confidence interval with these two methods. In terms of robust total least squares linear regression estimation, both uncertain total least squares regression estimation and uncertain robust total least squares regression estimation are feasible based on numerical examples. There are more accurate fitting equations and more reliable results with uncertain robust least squares linear regression estimation.



Citation: Shi, H.; Zhang, X.; Gao, Y.; Wang, S.; Ning, Y. Robust Total Least Squares Estimation Method for Uncertain Linear Regression Model. *Mathematics* **2023**, *11*, 4354. <https://doi.org/10.3390/math11204354>

Academic Editors: Jin Peng,
Dan Ralescu and Carlos
Conceicao Antonio

Received: 2 September 2023

Revised: 9 October 2023

Accepted: 18 October 2023

Published: 20 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: uncertain robust total least squares estimation; total least squares method; uncertainty theory; linear regression

MSC: 62P99

1. Introduction

Linear regression is a widely used statistical method in mathematical statistics, which reveals the mutual dependence between two or more variables through regression analysis. This method is particularly useful for determining the relationship between variables and predicting future trends or values based on historical data, playing a crucial role in research and the solution of practical problems. It is one of the problems encountered in many engineering applications. The least squares method is a widely used mathematical optimization technique in data analysis and modeling. It aims to minimize the sum of squares of errors to establish the relationship between variables and derive the most accurate fitting function for predicting and inferring future data. The least squares method is one of the common methods of linear fitting. In machine learning, especially regression models, least squares is often used [1]. Although the least square method can be used to solve the regression equation easily and minimize the sum of squares of errors between the obtained data and the actual data, it only takes into account the errors of the response variable. When the explanatory variable and the response variable are selected differently,

the results of linear regression are often different. In order to solve this problem, many scholars have conducted in-depth studies and proposed the total least squares method, which can simultaneously consider errors in explanatory variables and response variables and obtain good results [2]. But the total least squares linear fitting does not take into account the possible gross errors or outliers in the data, which are actually objective in the measurement. Some experts and scholars derive a robust total least squares linear fitting [3,4]. For the traditional total least squares method and the robust total least squares method mentioned above, they can perform better linear regression on random data and obtain better fitting functions. In practice, however, the data we obtain are not necessarily random or not accurate values or even rough ranges. In this case, traditional total least squares and robust total least squares cannot be used for linear regression. What to do? The uncertain regression analysis proposed by Liu [5] can help us solve these problems.

The uncertainty theory established by Liu [5–8] is a new interdisciplinary and comprehensive discipline. Its research purpose is to establish a set of axiomatic mathematical systems to research uncertainty and provide rigorous mathematical tools, theoretical basis and method guidance for dealing with uncertain information in the real world. The basic content of uncertainty theory is the three concepts of uncertain measure, uncertain variable and uncertainty distribution and the two algorithms of uncertain operation laws and uncertain expected value. The uncertainty theory is effective in addressing real-world uncertainty problems. Extensive research has been conducted by numerous experts and scholars, who have also widely applied this theory [9–11]. In 2010, Liu [5] initiated research into uncertain statistics. He introduced the least squares principle as a means to estimate unknown parameters in uncertain distributions. This groundbreaking work has since inspired many experts and scholars to delve into the field of uncertainty statistics [12–14]. Uncertain statistics is a mathematical tool to collect, analyze and process data using uncertainty theory. Uncertain regression analysis is an important research topic and it is a technical tool to research the relationship between variables by using uncertainty theory. In 2018, Yao and Liu [15] proposed least squares estimates for unknown parameters of uncertain regression equations and Lio and Liu [16] further suggested an approach to estimate the uncertainty distribution of the disturbance term in an uncertain regression model. In 2020, Wang, Li and Guo [17] proposed a new uncertain regression model and Liu, Yang [18] proposed the least absolute deviation estimation for uncertain regression with imprecise observations. Wang et al. [19,20] proposed new uncertain linear regression models.

Deterministic models and uncertain models are symmetric. So far, we have many methods to solve the uncertain regression model, but most of these methods only consider the error of the dependent variable, which is obviously one-sided and the accuracy and reliability of the regression equation will be affected. In 2022, Shi et al. [21] proposed an uncertain total least squares estimation. The uncertain total least squares estimation improves the accuracy of the regression equation and has a good effect. However, when there are many variables, the solution process of this method is very tedious. Based on previous studies [2] and uncertain statistics, this paper first proposes an uncertain total least squares estimation method, which applies the traditional singular value decomposition to uncertain regression analysis and takes into account the errors of independent variables and dependent variables. The regression equation obtained is more reasonable. In fact, gross errors or outliers in given data are objective and unavoidable and the total least squares estimation does not take into account the possible gross errors or outliers in the data. Therefore, based on the robust total least squares estimation [3,4], we propose a robust uncertain total least squares estimation when considering gross errors or outliers in the observed data. This method not only considers the errors in all the observed values, but also can eliminate the gross errors or outliers in the data, with higher accuracy and more reliable results.

In this paper, we propose an uncertain total least squares estimation model for linear regression and then derive uncertain robust population least squares estimation when considering the existence of gross errors or outliers in the data. The main organizational

structure of this paper is as follows: In Section 2, this paper introduces the uncertain regression model, which is an important part of uncertain statistics and also the basis of the model proposed in this paper. In Section 3, we propose an uncertain total least squares estimation model, which fully considers the errors of independent variables and dependent variables and results in better regression equations. In Section 4, because there are often gross errors or outliers in the data, this paper proposes an uncertain robust total least squares estimation model, which is more stable and reliable. In Section 5, a numerical example verifies the feasibility of uncertain total least squares estimation and uncertain robust total least squares estimation and compared with uncertain least squares estimation the data analysis shows that the regression equation of the model proposed in this paper is better. Finally, the proposed model is discussed and summarized.

2. Uncertain Regression Model

In 2007, Liu [6] founded the uncertainty theory. Uncertain statistics is an important part of uncertainty theory and can better deal with non-random or imprecise data. Readers who are interested in uncertainty theory can refer to Reference [8] for further information. This section focuses on introducing the uncertain regression model.

Assume that the vector of explanatory variables is $(x_1, x_2, \dots, x_n)$ and the response variable is y . The functional relationship between $(x_1, x_2, \dots, x_n)$ and y can be used to express it as in Equation (1)

$$y = f(x_1, x_2, \dots, x_n | \beta) + \varepsilon, \quad (1)$$

where β is an unknown vector of parameters and ε is a disturbance term [8]. The model is called an uncertain regression model.

In particular, Liu [8] called

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (2)$$

a linear regression model.

Now, suppose that we have a set of imprecise data as in Equation (3)

$$(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{in}, \tilde{y}_i), i = 1, 2, \dots, m, \quad (3)$$

where $\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{in}, \tilde{y}_i$ are independent uncertain variables with regular uncertainty distributions $\Phi_{i1}, \Phi_{i2}, \dots, \Phi_{in}, \Psi_i, i = 1, 2, \dots, m$, respectively. The following formulas are defined and derived from this premise.

Yao-Liu [15] introduced the least squares estimation method for the unknown parameter β in a linear regression model. This parameter can be obtained by solving the following minimization problem

$$\min_{\beta} \sum_{i=1}^n E[(\tilde{y}_i - f(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{in} | \beta))^2]. \quad (4)$$

The operator symbol E in Equation (4) is the symbol of the uncertain expected value, which is the average value of the uncertain variable in the sense of uncertain measure [8].

If the minimization solution is β^* , then the regression equation that best fits the data is given by $y = f(x_1, x_2, \dots, x_n | \beta^*)$. For each index i ($i = 1, 2, \dots, m$), the term

$$\tilde{\varepsilon}_i = \tilde{y}_i - f(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{in} | \beta^*) \quad (5)$$

is called the i -th residual.

Let the disturbance term ε be an uncertain variable, whose expected value and variance can be estimated as

$$\hat{\varepsilon} = \frac{1}{n} \sum_{i=1}^m E[\tilde{\varepsilon}_i], \quad (6)$$

and

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^m E[(\tilde{\varepsilon}_i - \hat{\varepsilon})^2], \quad (7)$$

where $\tilde{\varepsilon}_i$ are the i -th residual, $i = 1, 2, \dots, m$, respectively [16].

Given a new vector of explanatory variables $(x_1, x_2, \dots, x_n)$, the forecast uncertain variable of response variable $\hat{y}$ is

$$\hat{y} = f(x_1, x_2, \dots, x_n | \beta^*) + \mathcal{N}(\hat{\varepsilon}, \hat{\sigma}). \quad (8)$$

According to Lio-Liu [16], the forecast value is defined as the expected value of the uncertain variable $\tilde{y}$, which means

$$\hat{u} = f(x_1, x_2, \dots, x_n | \beta^*) + \hat{\varepsilon}. \quad (9)$$

Taking α (e.g., 95%) as the confidence level [16], the confidence interval of $\tilde{y}$ is

$$\hat{u} \pm \frac{\hat{\sigma}\sqrt{3}}{\pi} \ln \frac{1+\alpha}{1-\alpha}. \quad (10)$$

3. Uncertain Total Least Squares Estimation

Based on the traditional total least squares method [2], combined with uncertain statistics, this paper proposes uncertain total least squares estimation. This method can deal with imprecise data well and is an important linear regression model for uncertain statistics. As an extension of the traditional total least squares method, it expands the application field of linear regression models and it can produce good fitting results.

In this study, we consider a dataset consisting of imprecise data $(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{im}, \tilde{y}_i)$ for $i = 1, 2, \dots, n$. Each $\tilde{x}_{ij}$ and $\tilde{y}_i$ represents an independent uncertain variable, with corresponding regular uncertainty distributions $\Phi_{i1}, \Phi_{i2}, \dots, \Phi_{im}, \Psi_i$, where i ranges from 1 to n .

We assume that uncertain variables $(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{im})$ ($i = 1, 2, \dots, n$) and $\tilde{y}_i$ ($i = 1, 2, \dots, n$) have the following linear function relationship,

$$\begin{aligned} \tilde{y}_1 &= \beta_0 + \beta_1 \tilde{x}_{11} + \beta_2 \tilde{x}_{12} + \dots + \beta_m \tilde{x}_{1m}, \\ \tilde{y}_2 &= \beta_0 + \beta_1 \tilde{x}_{21} + \beta_2 \tilde{x}_{22} + \dots + \beta_m \tilde{x}_{2m}, \\ \tilde{y}_3 &= \beta_0 + \beta_1 \tilde{x}_{31} + \beta_2 \tilde{x}_{32} + \dots + \beta_m \tilde{x}_{3m}, \\ &\dots \\ \tilde{y}_n &= \beta_0 + \beta_1 \tilde{x}_{n1} + \beta_2 \tilde{x}_{n2} + \dots + \beta_m \tilde{x}_{nm}, \end{aligned} \quad (11)$$

where $\beta_0, \beta_1, \beta_2, \dots, \beta_m$ are unknown parameters. The model represented by Equation (11) is referred to as a multiple linear regression model. In Equation (11), $(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{im})$ ($i = 1, 2, \dots, n$) and $\tilde{y}_i$ ($i = 1, 2, \dots, n$) are uncertain variables.

According to the uncertainty theory, we can calculate the expected values of both sides of Equation (11) and we obtain the following equations with constant coefficients

$$\begin{aligned} E(\tilde{y}_1) &= \beta_0 + \beta_1 E(\tilde{x}_{11}) + \beta_2 E(\tilde{x}_{12}) + \dots + \beta_m E(\tilde{x}_{1m}), \\ E(\tilde{y}_2) &= \beta_0 + \beta_1 E(\tilde{x}_{21}) + \beta_2 E(\tilde{x}_{22}) + \dots + \beta_m E(\tilde{x}_{2m}), \\ E(\tilde{y}_3) &= \beta_0 + \beta_1 E(\tilde{x}_{31}) + \beta_2 E(\tilde{x}_{32}) + \dots + \beta_m E(\tilde{x}_{3m}), \\ &\dots \\ E(\tilde{y}_n) &= \beta_0 + \beta_1 E(\tilde{x}_{n1}) + \beta_2 E(\tilde{x}_{n2}) + \dots + \beta_m E(\tilde{x}_{nm}). \end{aligned} \quad (12)$$

According to matrix theory, Equation (12) can be expressed as the following matrix equation.

$$\begin{bmatrix} E(\tilde{y}_1) \\ E(\tilde{y}_2) \\ E(\tilde{y}_3) \\ \vdots \\ E(\tilde{y}_n) \end{bmatrix} = \begin{bmatrix} 1 & E(\tilde{x}_{11}) & E(\tilde{x}_{12}) & \cdots & E(\tilde{x}_{1m}) \\ 1 & E(\tilde{x}_{21}) & E(\tilde{x}_{22}) & \cdots & E(\tilde{x}_{2m}) \\ 1 & E(\tilde{x}_{31}) & E(\tilde{x}_{32}) & \cdots & E(\tilde{x}_{3m}) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & E(\tilde{x}_{n1}) & E(\tilde{x}_{n2}) & \cdots & E(\tilde{x}_{nm}) \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}. \quad (13)$$

Assuming that

$$\beta_{(m+1) \times 1} = [\beta_0, \beta_1, \beta_2, \cdots, \beta_m]^T, \quad (14)$$

$$\tilde{Y}_{n \times 1} = [E(\tilde{y}_1), E(\tilde{y}_2), E(\tilde{y}_3), \cdots, E(\tilde{y}_n)]^T, \quad (15)$$

$$\tilde{X}_{n \times (m+1)} = \begin{bmatrix} 1 & E(\tilde{x}_{11}) & E(\tilde{x}_{12}) & \cdots & E(\tilde{x}_{1m}) \\ 1 & E(\tilde{x}_{21}) & E(\tilde{x}_{22}) & \cdots & E(\tilde{x}_{2m}) \\ 1 & E(\tilde{x}_{31}) & E(\tilde{x}_{32}) & \cdots & E(\tilde{x}_{3m}) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & E(\tilde{x}_{n1}) & E(\tilde{x}_{n2}) & \cdots & E(\tilde{x}_{nm}) \end{bmatrix}. \quad (16)$$

Then, Equation (13) can be expressed as

$$\tilde{Y}_{n \times 1} = \tilde{X}_{n \times (m+1)} \beta_{(m+1) \times 1}. \quad (17)$$

where, $\tilde{Y}_{n \times 1}$ is the observed value vector, $\beta_{(m+1) \times 1}$ is the unknown parameter, $\tilde{X}_{n \times (m+1)}$ is the coefficient matrix. Considering the existence of errors in the imprecise data, we establish the following EIV (error-in-variable) model

$$\tilde{Y}_{n \times 1} + \varepsilon_{n \times 1} = (\tilde{X}_{n \times (m+1)} + E_{n \times (m+1)}) \beta_{(m+1) \times 1}, \quad (18)$$

where, $\varepsilon_{n \times 1}$ is the random error matrix of $\tilde{Y}_{n \times 1}$ and $E_{n \times (m+1)}$ is the random error matrix of coefficient matrix $\tilde{X}_{n \times (m+1)}$.

The uncertain total least squares estimation method is extended based on the least squares estimation, considering the error of observation vector and coefficient matrix. The criterion of the uncertain total least squares estimation is

$$\min = \varepsilon_{n \times 1}^T \varepsilon_{n \times 1} + \text{vec}(E_{n \times (m+1)})^T \text{vec}(E_{n \times (m+1)}), \quad (19)$$

where $\text{vec}(\cdot)$ represents the matrix column vectorization operator.

Equation (18) is transformed into

$$\begin{bmatrix} \tilde{X}_{n \times (m+1)} + E_{n \times (m+1)} & \tilde{Y}_{n \times 1} + \varepsilon_{n \times 1} \end{bmatrix} \cdot \begin{bmatrix} \beta_{(m+1) \times 1} \\ -1 \end{bmatrix} = \mathbf{0} \quad (20)$$

and

$$([\tilde{X}_{n \times (m+1)} \quad \tilde{Y}_{n \times 1}] \quad [E_{n \times (m+1)} \quad \varepsilon_{n \times 1}]) \cdot \begin{bmatrix} \beta_{(m+1) \times 1} \\ -1 \end{bmatrix} = \mathbf{0}. \quad (21)$$

We use matrix singular value decomposition (SVD) [22] to calculate the unknown parameters of the total least squares estimation. By singular value decomposition of the augmented matrix C , we can obtain

$$C = [\tilde{X}_{n \times (m+1)} \quad \tilde{Y}_{n \times 1}] = U \Sigma V^T. \quad (22)$$

where $U = [u_1, u_2, \dots, u_n]$ is an orthogonal matrix composed of n eigenvalue vectors of the matrix CC^T . $V = [v_1, v_2, \dots, v_{m+1}]$ is an orthogonal matrix composed of $m+1$ eigenvalue vectors of the matrix C^TC . $\Sigma = \text{diag}[\sigma_1, \sigma_2, \dots, \sigma_{m+1}]$ is a diagonal matrix of order $n \times (m+1)$, where the elements on the diagonal are the singular value ($\sigma_{ii} = \sigma_i, i = 1, 2, \dots, m+1$) and the elements on the nondiagonal are all 0.

By singular value decomposition of the augmented matrix $\hat{C}$, we can obtain

$$\hat{C} = ([\tilde{X}_{n \times (m+1)} \quad \tilde{Y}_{n \times 1}] \quad [E_{n \times (m+1)} \quad \varepsilon_{n \times 1}]) = U \begin{bmatrix} \Sigma' & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} V^T, \quad (23)$$

where $\Sigma' = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_m)$.

The left and right singular vector matrices in Equation (23) can be decomposed into

$$\begin{aligned} U &= [U_1, U_2], \\ V &= [V_1, V_2] \\ U_1 &= [u_1, u_2, \dots, u_{n-1}], U_2 = [U_n], \\ V_1 &= [v_1, v_2, \dots, v_m], V_2 = [v_{m+1}]. \end{aligned} \quad (24)$$

Equation (21) is transformed into

$$\hat{C} \begin{bmatrix} \beta_{(m+1) \times 1} \\ -1 \end{bmatrix} = \mathbf{0}. \quad (25)$$

Then, the total least squares estimate of the unknown parameter is

$$\beta_{TLS} = -V_2 V_{m+1, m+1}^{-1}. \quad (26)$$

The residual matrix is

$$E_{min} = [E_{n \times (m+1)} \quad \varepsilon_{n \times 1}] = -\sigma_{m+1} U_2 V_2^T. \quad (27)$$

In the derivation of the uncertain total least squares estimation, we applied the matrix singular value decomposition to calculate the least squares solution of unknown parameters of the regression equation. Singular value decomposition (SVD) is an important matrix decomposition in linear algebra. It is an extension of the unitary diagonalization of normal matrices in matrix analysis. It has important applications in signal processing, statistics and other fields. This method is theoretical and requires readers to have a strong knowledge of matrix theory.

4. Uncertain Robust Total Least Squares Estimation

Linear fitting is one of the problems encountered in many engineering applications. It can be described as follows: for given data $(\tilde{x}_i, \tilde{y}_i), (i = 1, 2, \dots, n)$, we need to find a best fit line $y = ax + b$ that passes through or is as close to these points as possible. We can use the total least squares estimator proposed in this paper to obtain estimates of parameters a and b .

To obtain the most accurate estimates of the linear parameters a and b in total least squares estimation, it is crucial to eliminate any gross errors or outliers present in the observed data. Building upon the principles of total least squares estimation, this paper introduces the uncertain robust total least squares estimation. The algorithm steps for this approach are outlined below:

Step 1. Calculate the initial values of a and b according to the total least squares estimation.

Step 2. Based on the calculated values of a and b , the distance d_i from each point to the fitted line is calculated.

$$d_i = \frac{|ax_i - y_i + b|}{\sqrt{a^2 + 1}}. \quad (28)$$

Step 3. Calculate the standard deviation of distance d_i according to the following equation

$$\gamma = \sqrt{\frac{(d_i - \bar{d})^T (d_i - \bar{d})}{n - 1}}, \quad (29)$$

where

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i. \quad (30)$$

Step 4. In general, we specify a distance threshold of 2 standard deviations, that is, when $d_i > 2\gamma$, the point is considered an outlier and is removed. Otherwise, it is retained.

Step 5. Recalculate a and b using all the remaining points.

Step 6. Repeat steps 2 to 5 until d_i of all remaining points is within the specified threshold; that is, d_i is less than two times the standard deviation.

Step 7. Calculate the final values of parameters a and b based on the uncertain total least squares estimation.

5. Numerical Example

In order to verify the feasibility and superiority of total least squares estimation and robust total least squares estimation, we provide an example of imprecise data and compare them with least squares estimation. Data analysis shows that total least squares estimation and robust total least squares estimation have better effect and higher accuracy. In addition, we also analyze the estimated expected value and variance of the disturbance term of the regression equation and calculate the forecast value and confidence interval.

In Table 1, we assume that the imprecise data (x_i, y_i) , $i = 1, 2, \dots, 12$ are independent uncertain variables with regular linear uncertainty distributions.

Table 1. Imprecise data (linear uncertainty distribution).

i	1	2	3	4
x_i	$\mathcal{L}(2,4)$	$\mathcal{L}(3,5)$	$\mathcal{L}(4,6)$	$\mathcal{L}(5,6)$
y_i	$\mathcal{L}(4,6)$	$\mathcal{L}(6,8)$	$\mathcal{L}(3,5)$	$\mathcal{L}(10,12)$
i	5	6	7	8
x_i	$\mathcal{L}(6,8)$	$\mathcal{L}(7,9)$	$\mathcal{L}(8,10)$	$\mathcal{L}(9,10)$
y_i	$\mathcal{L}(12,14)$	$\mathcal{L}(14,16)$	$\mathcal{L}(22,24)$	$\mathcal{L}(18,20)$
i	9	10	11	12
x_i	$\mathcal{L}(10,12)$	$\mathcal{L}(11,13)$	$\mathcal{L}(12,14)$	$\mathcal{L}(13,15)$
y_i	$\mathcal{L}(20,22)$	$\mathcal{L}(22,24)$	$\mathcal{L}(24,26)$	$\mathcal{L}(26,28)$

Table 2 shows the linear regression equations derived using least squares estimations, total least squares estimations and robust total least squares estimations.

Table 2. The linear regression equations.

Model	Regression Equations
least squares estimation	$y = -1.8428 + 2.1298x$
total least squares estimation	$y = -1.363 + 2.0455x$
robust total least squares estimation	$y = -0.9815 + 2.0093x$

From Table 2, it can be seen that there is little difference between the coefficients of regression equations for least squares estimation, for total least squares estimation and for robust total least squares estimation and that the method proposed here is a feasible one.

Table 3 shows a comparison between the expected value and variance of the disturbance terms when least squares estimation, total least squares estimation and robust total least squares estimation are undertaken.

Table 3. The expected value and variance of the disturbance term.

Model	Expected Value	Variance
least squares estimation	0.0000	59.8156
total least squares estimation	0.0000	9.9545
robust total least squares estimation	0.0000	0.8889

In Table 3, it is shown that the expected value of the disturbance term in regression equations of least squares estimation, total least squares estimation and robust total least squares estimation are all the same, so we are able to say that all three methods produce the same regression results. With robust total least squares estimation, the disturbance term in the regression equation has the smallest variance, the best effect and the highest level of accuracy compared to all other estimation methods.

Forecast values and confidence intervals can also be calculated using robust total least squares estimation. Assume that $x \sim \mathcal{L}(14, 15)$ is imprecise data and according to [16], we obtain a forecast value of $\hat{u}$ as 28.1534.

The confidence level is assumed to be $\alpha = 95\%$ and the interference term is assumed to have an uncertain normal distribution $\mathcal{N}(\hat{e}, \hat{\sigma})$. According to Reference [16], the confidence interval of variable y is

$$28.1534 \pm 1.9053.$$

Uncertain total least squares estimation takes into account the errors of independent variables and dependent variables, which makes the linear regression equation more efficient. Uncertain robust total least squares estimation removes the outliers from the data on the basis of total least squares estimation, so the regression equation has better effect and higher accuracy. In the example, compared with the data results of the existing methods, robust total least squares estimation has a better regression effect than other methods.

6. Conclusions

There are many practical applications for linear regression and we always expect to obtain better linear fitting equations when we apply linear regression techniques. The problem can be solved using the mathematical statistics and probability theory when the given data are random. In cases where the data are not random, we can use the estimation method that has been proposed in this paper to estimate the data. It is demonstrated in this paper that total least squares estimation takes into account both the errors of the explanatory variables as well as errors of the response variables, which makes it more effective. It is also important to note that a robust total least squares linear estimation method has also been proposed in order to address the issue of outliers not being taken into account in previous linear fitting methods. With the help of this method, gross errors or

outliers in the data can be removed and the regression equation obtained is more accurate and more reliable than those obtained using other estimation methods. Robust total least squares linear estimation is more stable in practice.

In the derivation of the uncertain robust total least squares estimation method, this paper uses the singular value decomposition method of the matrix. In spite of the fact that this method is widely used and has great effectiveness, it is difficult to understand and requires a certain level of knowledge on linear algebra and matrix theory.

Author Contributions: Conceptualization, S.W.; methodology, Y.N.; validation, Y.G.; data curation, X.Z.; writing—original draft preparation, H.S.; writing—review and editing, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Ding, K.; Shen, Y.; Ou, J. Linear fitting by total least squares Method. *J. Liaoning Tech. Univ. (Nat. Sci. Ed.)* **2010**, *29*, 44–47.
- Lu, T.; Tao, B.; Zhou, S. Linear regression and modeling method based on total least squares. *J. Wuhan Univ. (Inform. Sci. Ed.)* **2008**, *33*, 504–507.
- Wang, Q.; Yang, D.; Yang, T. The solution of robust total least squares for linear regression models. *J. Geod. Geodyn.* **2015**, *35*, 239–242.
- Guan, Y.; Zhou, S.; Zhang, L.; Lu, T. A robust method for fitting a line to point clouds based on TLS. *Geotech. Investig. Surv.* **2012**, *2*, 60–62.
- Liu, B. *Uncertainty Theory: A Branch of Mathematics for Modeling Human Uncertainty*; Springer: Berlin/Heidelberg, Germany, 2010.
- Liu, B. *Uncertainty Theory*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 10–95.
- Liu, B. Why is there a need for uncertainty theory? *J. Uncertain Syst.* **2012**, *6*, 3–10.
- Liu, B. *Uncertainty Theory*, 5th ed.; Springer: Berlin/Heidelberg, Germany, 2021.
- Wang, X.; Ning, Y. An Uncertain currency model with floating interest rates. *Soft Comput.* **2017**, *21*, 6739–6754. [[CrossRef](#)]
- Ning, Y.; Liu, J.; Yan, L. Uncertain aggregate production planning. *Soft Comput.* **2013**, *17*, 617–624. [[CrossRef](#)]
- Ning, Y.; Pang, N.; Wang, X. An Uncertain aggregate production planning model considering investment in vegetable preservation technology. *Math. Probl. Eng.* **2019**, *2019*, 8505868. [[CrossRef](#)]
- Wang, X.; Peng, Z. Method of moments for estimating uncertainty distributions. *J. Uncertain. Anal. Appl.* **2014**, *2*, 5. [[CrossRef](#)]
- Guo, H.; Wang, X.; Gao, Z. Uncertain linear regression model and its application. *J. Intell. Manuf.* **2014**, *28*, 559–564. [[CrossRef](#)]
- Chen, D. Tukey's biweight estimation for uncertain regression model with imprecise observations. *Soft Comput.* **2020**, *24*, 16803–16809. [[CrossRef](#)]
- Yao, K.; Liu, B. Uncertain regression analysis: An approach for imprecise observations. *Soft Comput.* **2018**, *22*, 5579–5582. [[CrossRef](#)]
- Lio, W.; Liu, B. Residual and confidence interval for uncertain regression model with imprecise observations. *J. Intell. Fuzzy Syst.* **2018**, *35*, 2573–2583. [[CrossRef](#)]
- Wang, X.; Li, H.; Guo, H. A new Uncertain regression model and its application. *Soft Comput.* **2020**, *24*, 6297–6305. [[CrossRef](#)]
- Liu, Z.; Yang, Y. Least absolute deviations estimation for uncertain regression with imprecise observations. *Fuzzy Optim. Decis. Mak.* **2020**, *19*, 33–52. [[CrossRef](#)]
- Wang, S.; Ning, Y.; Shi, H.; Chen, X. A new uncertain linear regression model based on slope mean. *J. Intell. Fuzzy Syst.* **2021**, *40*, 10465–10474. [[CrossRef](#)]
- Wang, S.; Ning, Y.; Huang, H.; Chen, X. Uncertain least squares estimation model based on relative error. *J. Intell. Fuzzy Syst.* **2023**, *44*, 8281–8290. [[CrossRef](#)]
- Shi, S.; Sun, X.; Wang, S.; Ning, Y. Total least squares estimation model based on uncertainty theory. *J. Ambient. Intell. Humaniz. Comput.* **2023**, *14*, 10069–10075. [[CrossRef](#)]
- Schaffrin, B. A note on constrained total least squares estimation. *Linear Algebra Its Appl.* **2006**, *17*, 245–258. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.