

A new uncertain linear regression model based on equation deformation

Shuai Wang

Shandong Youth University of Political Science

Yufu Ning

Shandong Youth University of Political Science

Hongmei Shi (✉ 22073319@qq.com)

Shandong Agriculture and Engineering University <https://orcid.org/0000-0001-7205-8509>

Research Article

Keywords: equation deformation method, least squares estimate, linear regression model, uncertainty theory

Posted Date: March 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-191330/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Soft Computing on August 6th, 2021. See the published version at <https://doi.org/10.1007/s00500-021-06030-7>.

A new uncertain linear regression model based on equation deformation

Shuai Wang · Yufu Ning · Hongmei Shi ✉

Received: date / Accepted: date

Abstract When the observed data are imprecise, the uncertain regression model is more suitable for the linear regression analysis. Least squares estimate can fully consider the given data and minimize the sum of squares of residual error, and can effectively solve the linear regression equation of imprecisely observed data. On the basis of uncertainty theory, this paper presents an equation deformation method for solving unknown parameters in uncertain linear regression equations. We first establish the equation deformation method of one-dimensional linear regression model, and then extend it to the case of multiple linear regression model. We also combine the equation deformation method with Cramer's rule and matrix, and propose the Cramer's rule and matrix elementary transformation method to solve the unknown parameters of the uncertain linear regression equation. Numerical examples show that the equation deformation method can effectively solve the unknown parameters of the uncertain linear regression equation.

Keywords equation deformation method · least squares estimate · linear regression model · uncertainty theory

1 Introduction

Regression analysis is an important branch of statistics. It is a kind of statistical method to study the relationship between response variables and explanatory variables. Regression analysis is also one of the most commonly used and

important statistical tools [1], and its application fields are very wide. Linear regression is an important model of regression analysis. One-dimensional linear regression can be described by linear equation, and multivariate linear regression can be described by linear combination of variables [2]. Linear regression can describe the linear relationship between random variables, but it requires precise data, and the random variables obey or are close to probability distributions. But in fact, the observed data of many practical problems are imprecise or the information obtained is not sufficient, or even some practical problems have no data. The traditional regression analysis encountered difficulties at this point. When observed data is imprecise or information is not available, we often invite domain experts to estimate the extent to which the event is likely to occur, or to predict the possible range of the event [3]. We call the data

Shuai Wang, Yufu Ning
School of Information Engineering, Shandong Youth University of Political Science, No. 31699, Jingshi East Road, Jinan, China
Key Laboratory of Information Security and Intelligent Control in Universities of Shandong, Jinan, 250100, China
E-mail: wangshuai@sdyu.edu.cn, nyf@sdyu.edu.cn

✉ Hongmei Shi
School of Information Science and Engineering, Shandong Agricultural and Engineering University, No. 866, Nongganyuan Road, Licheng District, Jinan, China
E-mail: yqbshmi@163.com

given by experts as the belief degree, but the belief degree usually does not approximate the probability distribution, so we need a new way to deal with the belief degree. Uncertainty theory [4][10] is very good at studying and analyzing of belief degree.

Liu [4] founded the uncertainty theory in 2007, and gradually perfected and developed it [5]–[7][11]. With the deepening of the research, uncertainty theory has been applied to many fields and achieved good results [18]–[23]. The first question is how to construct uncertainty distributions of uncertain variables, Liu [6] first proposed uncertain statistics in 2010 that solved this problem. Chen and Ralescu [14] estimated the distance from Tianjin to Beijing in 2012 which Liu proposed questionnaire survey and the results showed very efficient. To estimate the unknown parameters in the uncertainty distribution, Liu [6] proposed a principle of least squares in 2010. With the further study of uncertain statistics, Wang and Peng [15] puts forward the method of moments to estimate the unknown parameters, Guo, Wang and Gao [16] proposed an uncertain linear regression model in 2014. Yao and Liu [8] puts forward a point estimation method for solving unknown parameters of uncertain regression equation through the principle of least square method in 2018, which is a method of processing imprecisely observed data. In 2018, Song and Fu [12] proposed a least square method to solve the unknown parameters of uncertain multiple linear regression. Chen [13] proposed Tukey's biweight estimation for uncertain regression model with imprecise observations in 2020.

The least squares estimate can solve the parameters of linear uncertain regression equation, but it needs some advanced mathematics foundation. When there are many explanatory variables or large amount of data, the least squares estimate will encounter some difficulties. This paper presents an equation deformation method for solving unknown parameters of linear regression. The equation deformation method can solve the unknown parameters only by using the deformation of the equations, which is easy to understand and easy to use.

The remained organizational structure of this paper is as follows: In the second section, this paper first gives the calculation formula of expected value. Then the uncertain regression model is introduced and the least squares estimate for solving the unknown parameters is given. In Section 3, based on the uncertainty theory, we proposed the equation deformation method to solve the unknown parameters, and deduces the solving process in detail. Then, we extend the equation deformation method to multiple linear regression model. Then in Section 4, according to the Cramer's rule and the elementary transformation of the matrix, we proposed auxiliary solutions of the equation deformation method. In the fifth section, we verified the feasibility of the equation deformation method through a numerical example and com-

pared it with the existing method. Finally, we have made a summary.

2 Uncertain regression model

In 2007, Liu [4] founded the uncertainty theory based the three axioms of Normal Axiom, Duality Axiom and Subadditivity Axiom. In 2009, Liu [7] perfected the uncertainty theory through the Product Axiom. The uncertainty theory defines the uncertain variables and the uncertainty distribution, and the inverse uncertainty distribution is used to solve the expected value. Readers interested in uncertainty theory can read the Reference [11] for other basic concepts and theories of uncertainty theory.

In this section, we first introduced the concept and calculation method of expected value. Then, we give the uncertain regression model and introduced the least squares estimate method to solve the regression equation.

Theorem 2.1 [6] Assumed that the uncertain variable ξ has an regular uncertainty distribution Φ . then

$$E[\xi] = \int_0^1 \Phi^{-1}(\alpha) d\alpha. \quad (1)$$

Theorem 2.2 [6] Let ξ and η be independent and positive uncertain variables with regular uncertainty distribution Φ and Ψ , respectively. then

$$E[\xi\eta] = \int_0^1 \Phi^{-1}(\alpha)\Psi^{-1}(\alpha) d\alpha. \quad (2)$$

Assumed that (x_1, x_2, \dots, x_p) be a vector of explanatory variables, and the corresponding response variable be y . Now, we assumed that (x_1, x_2, \dots, x_p) have a functional relationship with y , and can be expressed in the following regression model

$$y = f(x_1, x_2, \dots, x_p | \beta) + \varepsilon \quad (3)$$

where β is the vector of the unknown parameters, and ε is a disturbance term [11].

We will call

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (4)$$

a linear regression model [11].

Now we assumed that we have a set of imprecisely observed data,

$$(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ip}, \tilde{y}_i), i = 1, 2, \dots, n \quad (5)$$

where $\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ip}, \tilde{y}_i$ are independent uncertain variables with regular uncertainty distributions $\Phi_{i1}, \Phi_{i2}, \dots, \Phi_{ip}, \Psi_i, i = 1, 2, \dots, n$, respectively.

Based on the above imprecisely observed data, Yao-liu [8] proposed the least squares estimate of β in the regression model

$$y = f(x_1, x_2, \dots, x_p | \beta) + \varepsilon. \quad (6)$$

The parameters β is the solution of the following minimization problem

$$\min_{\beta} \sum_{i=1}^n E[(\tilde{y}_i - f(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ip} | \beta))^2]. \quad (7)$$

If the minimization solution is β^* , then the fitted regression equation is $y = f(x_1, x_2, \dots, x_p | \beta^*)$.

When $p = 1$, we call

$$y = \beta_0^* + \beta_1^* x_1 \quad (8)$$

a one-dimensional linear regression equation. When $p > 1$, we call

$$y = \beta_0^* + \beta_1^* x_1 + \beta_2^* x_2 + \dots + \beta_p^* x_p \quad (9)$$

the multiple linear regression equation.

Then for each index $i (i = 1, 2, \dots, n)$, the term

$$\tilde{\varepsilon}_i = (\tilde{y}_i - f(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ip} | \beta^*)) \quad (10)$$

is called the i -th residual [9].

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

$$\hat{e} = \frac{1}{n} \sum_{i=1}^n E[\tilde{\varepsilon}_i], \quad (11)$$

and

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n E[(\tilde{\varepsilon}_i - \hat{e})^2], \quad (12)$$

where $\tilde{\varepsilon}_i$ are the i -th residual, $i = 1, 2, \dots, n$, respectively [9].

3 The equation deformation method

In this section, based on the uncertainty theory, we proposed an equation deformation method for solving unknown parameters. The idea of equation deformation method is to construct a set of equations with the same number of unknown parameters and solve them with expected value. Then, we extend the equation deformation method to multiple linear regression equations.

3.1 Equation deformation method for one-dimensional linear regression model

we always assumed that $(\tilde{x}_i, \tilde{y}_i), i = 1, 2, \dots, n$ be a set of imprecisely observed data, where \tilde{x}_i, \tilde{y}_i are independent uncertain variables with regular uncertainty distributions $\Phi_i, \Psi_i, i = 1, 2, \dots, n$, respectively.

We supposed that $(\tilde{x}_i, \tilde{y}_i), i = 1, 2, \dots, n$, satisfied the linear regression equation

$$\tilde{y}_i = \beta_0 + \beta_1 \tilde{x}_i, \quad (13)$$

where β_0 and β_1 are the unknown parameters.

Equation (13) has two unknown parameters. According to mathematical equation theory, we only need two independent equations. If Equation (13) has a solution, we can definitely solve it. Equation (13) contains n independent equations, and unknown parameters can be solved by arbitrarily selecting two equations. However, the selected equations are very one-sided and cannot represent the overall properties of the variables. In order to fully consider the influence of the value of variables and minimize the error, we construct two new equations. The detailed process is as follows.

We can take the expected values of both sides of the Equation (13), turn it into a real coefficient equation, and we get

$$E[\tilde{y}_i] = \beta_0 + \beta_1 E[\tilde{x}_i]. \quad (14)$$

Add the n equations in Equation (14), and we get

$$\sum_{i=1}^n E[\tilde{y}_i] = n\beta_0 + \beta_1 \sum_{i=1}^n E[\tilde{x}_i]. \quad (15)$$

Equation (15) is transformed into

$$\frac{1}{n} \sum_{i=1}^n E[\tilde{y}_i] = \beta_0 + \beta_1 \frac{1}{n} \sum_{i=1}^n E[\tilde{x}_i]. \quad (16)$$

According to Equation (1), Equation (16) is converted into

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \int_0^1 \Psi_i^{-1}(\alpha) d\alpha \\ &= \beta_0 + \beta_1 \frac{1}{n} \sum_{i=1}^n \int_0^1 \Phi_i^{-1}(\alpha) d\alpha. \end{aligned} \quad (17)$$

Multiply both sides of Equation (14) by $E[\tilde{x}_i]$, and we get

$$E[\tilde{y}_i]E[\tilde{x}_i] = \beta_0 E[\tilde{x}_i] + \beta_1 E[\tilde{x}_i]^2. \quad (18)$$

Add the n equations in Equation (18), and we get

$$\sum_{i=1}^n E[\tilde{y}_i]E[\tilde{x}_i] = \beta_0 \sum_{i=1}^n E[\tilde{x}_i] + \beta_1 \sum_{i=1}^n E[\tilde{x}_i]^2. \quad (19)$$

According to Equations (1) and (2), Equation (19) is converted into

$$\begin{aligned} & \sum_{i=1}^n \int_0^1 \Psi_i^{-1}(\alpha) d\alpha \int_0^1 \Phi_i^{-1}(\alpha) d\alpha \\ &= \beta_0 \sum_{i=1}^n \int_0^1 \Phi_i^{-1}(\alpha) d\alpha \\ &+ \beta_1 \sum_{i=1}^n \left[\int_0^1 \Phi_i^{-1}(\alpha) d\alpha \right]^2. \end{aligned} \quad (20)$$

Solved Equations (17) and (20), and we got the estimated values of β_0 and β_1 .

The derivation of one-dimensional linear regression of equation deformation method is relatively simple. On the basis of considering all the data, we reasonably construct two equations, and the unknown parameters can only be solved through the equation deformation and expected value. The equation deformation method does not need the foundation of advanced mathematics and is easy to understand and calculate.

3.2 Equation deformation method for multiple linear regression model

The equation deformation method can solve the unknown parameters of one-dimensional linear regression, then can it solve the unknown parameters of multiple linear regression equation? Let's go ahead and derive it.

Assumed that there is a linear functional relationship between uncertain variables $(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ip})$ and \tilde{y}_i . In order to solve the $p+1$ unknown parameters of linear regression equation, a more effective method is to establish $p+1$ equations. On the basis of the uncertainty theory, we proposed an equation deformation method to solve the unknown parameters of linear regression equation by using the expected value. The equation deformation method can fully considered the imprecisely observed data. The specific steps are as follows.

Step 1. Assume that the linear regression equation is

$$\tilde{y}_i = \beta_0 + \beta_1 \tilde{x}_{i1} + \beta_2 \tilde{x}_{i2} + \dots + \beta_p \tilde{x}_{ip} \quad (21)$$

where $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ are unknown parameters.

Add up the n equations in Equation (21), then

$$\begin{aligned} \sum_{i=1}^n \tilde{y}_i &= n\beta_0 + \beta_1 \sum_{i=1}^n \tilde{x}_{i1} + \beta_2 \sum_{i=1}^n \tilde{x}_{i2} \\ &+ \dots + \beta_p \sum_{i=1}^n \tilde{x}_{ip} \end{aligned} \quad (22)$$

Multiply both sides of Equation (21) by \tilde{x}_{i1} , then we get

$$\begin{aligned} \tilde{x}_{i1} \tilde{y}_i &= \beta_0 \tilde{x}_{i1} + \beta_1 (\tilde{x}_{i1})^2 + \beta_2 \tilde{x}_{i1} \tilde{x}_{i2} \\ &+ \dots + \beta_p \tilde{x}_{i1} \tilde{x}_{ip}. \end{aligned} \quad (23)$$

Add up the n equations in Equation (23), then

$$\begin{aligned} \sum_{i=1}^n \tilde{x}_{i1} \tilde{y}_i &= \beta_0 \sum_{i=1}^n \tilde{x}_{i1} + \beta_1 \sum_{i=1}^n (\tilde{x}_{i1})^2 \\ &+ \beta_2 \sum_{i=1}^n \tilde{x}_{i1} \tilde{x}_{i2} + \dots + \beta_p \sum_{i=1}^n \tilde{x}_{i1} \tilde{x}_{ip}. \end{aligned} \quad (24)$$

Multiply both sides of Equation (21) by \tilde{x}_{i2} , then we get

$$\begin{aligned} \tilde{x}_{i2} \tilde{y}_i &= \beta_0 \tilde{x}_{i2} + \beta_1 \tilde{x}_{i1} \tilde{x}_{i2} + \beta_2 (\tilde{x}_{i2})^2 \\ &+ \dots + \beta_p \tilde{x}_{i2} \tilde{x}_{ip}. \end{aligned} \quad (25)$$

Add up the n equations in Equation (25), then

$$\begin{aligned} \sum_{i=1}^n \tilde{x}_{i2} \tilde{y}_i &= \beta_0 \sum_{i=1}^n \tilde{x}_{i2} + \beta_1 \sum_{i=1}^n \tilde{x}_{i1} \tilde{x}_{i2} \\ &+ \beta_2 \sum_{i=1}^n (\tilde{x}_{i2})^2 + \dots + \beta_p \sum_{i=1}^n \tilde{x}_{i2} \tilde{x}_{ip}, \end{aligned} \quad (26)$$

And so on, Multiply both sides of Equation (21) by \tilde{x}_{ip} , then we get

$$\begin{aligned} \tilde{x}_{ip} \tilde{y}_i &= \beta_0 \tilde{x}_{ip} + \beta_1 \tilde{x}_{i1} \tilde{x}_{ip} + \beta_2 \tilde{x}_{i2} \tilde{x}_{ip} \\ &+ \dots + \beta_p (\tilde{x}_{ip})^2. \end{aligned} \quad (27)$$

Add up the n equations in Equation (27), then

$$\begin{aligned} \sum_{i=1}^n \tilde{x}_{ip} \tilde{y}_i &= \beta_0 \sum_{i=1}^n \tilde{x}_{ip} + \beta_1 \sum_{i=1}^n \tilde{x}_{i1} \tilde{x}_{ip} \\ &+ \beta_2 \sum_{i=1}^n \tilde{x}_{i2} \tilde{x}_{ip} + \dots + \beta_p \sum_{i=1}^n (\tilde{x}_{ip})^2. \end{aligned} \quad (28)$$

So we have p equations as follows

$$\begin{aligned} \sum_{i=1}^n \tilde{x}_{ij} \tilde{y}_i &= \beta_0 \sum_{i=1}^n \tilde{x}_{ij} + \beta_1 \sum_{i=1}^n \tilde{x}_{i1} \tilde{x}_{ij} \\ &+ \beta_2 \sum_{i=1}^n \tilde{x}_{i2} \tilde{x}_{ij} + \cdots + \beta_p \sum_{i=1}^n \tilde{x}_{ip} \tilde{x}_{ij}, \end{aligned} \quad (29)$$

$j = 1, 2, \dots, p.$

There are a total of $p + 1$ equations together with Equation (22) and Equation (29). If there are numerical solutions exists for $p + 1$ unknown parameters, the $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ can be solved by $p + 1$ equations.

Step 2. In Equations (22) and (29), $\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ip}, \tilde{y}_i, i = 1, 2, \dots, n$ are imprecisely observed data, we can't solve them directly. We can take the expected value of both sides of Equations (22) and (29), turn them into real coefficient equations, and we get

$$\begin{aligned} \sum_{i=1}^n E[\tilde{y}_i] &= n\beta_0 + \beta_1 \sum_{i=1}^n E[\tilde{x}_{i1}] \\ &+ \beta_2 \sum_{i=1}^n E[\tilde{x}_{i2}] + \cdots + \beta_p \sum_{i=1}^n E[\tilde{x}_{ip}], \\ \sum_{i=1}^n E[\tilde{y}_i \tilde{x}_{ij}] &= \beta_0 \sum_{i=1}^n E[\tilde{x}_{ij}] + \beta_1 \sum_{i=1}^n E[\tilde{x}_{i1} \tilde{x}_{ij}] \\ &+ \beta_2 \sum_{i=1}^n E[\tilde{x}_{i2} \tilde{x}_{ij}] + \cdots + \beta_p \sum_{i=1}^n E[\tilde{x}_{ip} \tilde{x}_{ij}], \end{aligned} \quad (30)$$

$j = 1, 2, \dots, p.$

Step 3. The $\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ip}, \tilde{y}_i$ are independent uncertain variables with regular uncertainty distributions $\Phi_{i1}, \Phi_{i2}, \dots, \Phi_{ip}, \Psi_i, i = 1, 2, \dots, n$, respectively.

So the first equation in Equation (30) is transformed into

$$\sum_{i=1}^n E[\tilde{y}_i] = n\beta_0 + \beta_1 \sum_{i=1}^n \sum_{k=1}^p \beta_k E[(\tilde{x}_{ik})]. \quad (31)$$

According to the Theorem 2.1, the Equation(31) is transformed into

$$\begin{aligned} \sum_{i=1}^n \int_0^1 \Psi_i^{-1}(\alpha) d\alpha \\ = n\beta_0 + \sum_{i=1}^n \sum_{k=1}^p \beta_k \int_0^1 \Phi_{ik}^{-1}(\alpha) d\alpha. \end{aligned} \quad (32)$$

The other equations in Equation (30) are transformed into

$$\begin{aligned} \sum_{i=1}^n E[\tilde{x}_{ij} \tilde{y}_i] \\ = \beta_0 \sum_{i=1}^n E[\tilde{x}_{ij}] + \sum_{i=1}^n \sum_{k=1}^p \beta_k E[\tilde{x}_{ik} \tilde{x}_{ij}]. \end{aligned} \quad (33)$$

$j = 1, 2, \dots, p.$

According to the Theorem 2.1 and Theorem 2.2, the Equation (33) is transformed into

$$\begin{aligned} \sum_{i=1}^n \int_0^1 \Phi_{i1}^{-1}(\alpha) \Psi_i^{-1}(\alpha) d\alpha &= \beta_0 \sum_{i=1}^n \int_0^1 \Phi_{ij}^{-1}(\alpha) d\alpha \\ &+ \sum_{i=1}^n \sum_{k=1}^p \beta_k \int_0^1 \Phi_{ik}^{-1}(\alpha) \Phi_{ij}^{-1}(\alpha) d\alpha. \end{aligned} \quad (34)$$

$j = 1, 2, \dots, p.$

Equations (32) and (34) contains $p + 1$ equations. By solved the above $p + 1$ equations, the estimated values of the unknown parameters $\beta_1, \beta_2, \dots, \beta_p$ can be obtained. We can get the fitting equation of the multiple linear regression.

4 Other auxiliary solutions

In the equation deformation method of multiple linear regression model, we can regard Equation (30) as a linear system of equations. With the help of the knowledge of linear algebra [17], we further discuss the equation deformation method, and proposed the Cramer's rule and elementary transformation of matrix for solving the unknown parameters of linear regression equation.

4.1 Cramer's rule

Equation (30) is equal to

$$\begin{aligned} n\beta_0 + \beta_1 \sum_{i=1}^n E[\tilde{x}_{i1}] + \beta_2 \sum_{i=1}^n E[\tilde{x}_{i2}] \\ + \cdots + \beta_p \sum_{i=1}^n E[\tilde{x}_{ip}] &= \sum_{i=1}^n E[\tilde{y}_i], \\ \beta_0 \sum_{i=1}^n E[\tilde{x}_{ij}] + \beta_1 \sum_{i=1}^n E[\tilde{x}_{i1} \tilde{x}_{ij}] + \beta_2 \sum_{i=1}^n E[\tilde{x}_{i2} \tilde{x}_{ij}] \\ + \cdots + \beta_p \sum_{i=1}^n E[\tilde{x}_{ip} \tilde{x}_{ij}] &= \sum_{i=1}^n E[\tilde{x}_{i1} \tilde{y}_i], \end{aligned} \quad (35)$$

$j = 1, 2, \dots, p.$

We assumed that the coefficient matrix corresponding to the above equation is A , the determinant of matrix A is

$$|A| = \begin{vmatrix} n & \sum_{i=1}^n E[\tilde{x}_{i1}] & \cdots & \sum_{i=1}^n E[\tilde{x}_{ip}] \\ \sum_{i=1}^n E[\tilde{x}_{i1}] & \sum_{i=1}^n E[(\tilde{x}_{i1})^2] & \cdots & \sum_{i=1}^n E[\tilde{x}_{i1}\tilde{x}_{ip}] \\ \sum_{i=1}^n E[\tilde{x}_{i2}] & \sum_{i=1}^n E[\tilde{x}_{i1}\tilde{x}_{i2}] & \cdots & \sum_{i=1}^n E[\tilde{x}_{i2}\tilde{x}_{ip}] \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^n E[\tilde{x}_{ip}] & \sum_{i=1}^n E[\tilde{x}_{i1}\tilde{x}_{ip}] & \cdots & \sum_{i=1}^n E[(\tilde{x}_{ip})^2] \end{vmatrix}. \quad (36)$$

If

$$|A| \neq 0, \quad (37)$$

the linear system of equations has a unique numerical solution according to cramer's rule.

Replace the JTH column of determinant $|A|$ with the constant term on the right side of the linear system of equations, and the resulting determinant is denoted as $|A_j|$, i.e

$$|A_j| =$$

$$\begin{vmatrix} \cdots & \sum_{i=1}^n E[\tilde{x}_{ij-1}] & \sum_{i=1}^n E[\tilde{y}_i] & \sum_{i=1}^n E[\tilde{x}_{ij+1}] & \cdots \\ \cdots & \sum_{i=1}^n E[\tilde{x}_{i1}\tilde{x}_{ij-1}] & \sum_{i=1}^n E[\tilde{x}_{i1}\tilde{y}_i] & \sum_{i=1}^n E[\tilde{x}_{i1}\tilde{x}_{ij+1}] & \cdots \\ \cdots & \sum_{i=1}^n E[\tilde{x}_{i2}\tilde{x}_{ij-1}] & \sum_{i=1}^n E[\tilde{x}_{i2}\tilde{y}_i] & \sum_{i=1}^n E[\tilde{x}_{i2}\tilde{x}_{ij+1}] & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \sum_{i=1}^n E[\tilde{x}_{ip}\tilde{x}_{ij-1}] & \sum_{i=1}^n E[\tilde{x}_{ip}\tilde{y}_i] & \sum_{i=1}^n E[\tilde{x}_{ip}\tilde{x}_{ij+1}] & \cdots \end{vmatrix}. \quad (38)$$

According to Cramer's rule, a set of numerical solutions of the linear system of equations are

$$\beta_0 = \frac{|A_1|}{|A|}, \beta_1 = \frac{|A_2|}{|A|}, \cdots, \beta_j = \frac{|A_{j+1}|}{|A|}, \cdots, \beta_p = \frac{|A_{p+1}|}{|A|}. \quad (39)$$

Solved the Equation of (39), we can get the estimated values of the unknown parameters $\beta_0, \beta_1, \cdots, \beta_p$. Therefore, the linear regression equation can also be solved by cramer's rule.

4.2 Elementary transformation of matrices

We start with Equation (30), which can be expressed as the following matrix equation

$$A\beta = b. \quad (40)$$

where

$$\beta = (\beta_0, \beta_1, \beta_2, \cdots, \beta_p)^T, \quad (41)$$

$$b = \left(\sum_{i=1}^n E[\tilde{y}_i], \sum_{i=1}^n E[\tilde{x}_{i1}\tilde{y}_i], \cdots, \sum_{i=1}^n E[\tilde{x}_{ip}\tilde{y}_i] \right)^T, \quad (42)$$

and

$$A = \begin{pmatrix} n & \sum_{i=1}^n E[\tilde{x}_{i1}] & \cdots & \sum_{i=1}^n E[\tilde{x}_{ip}] \\ \sum_{i=1}^n E[\tilde{x}_{i1}] & \sum_{i=1}^n E[(\tilde{x}_{i1})^2] & \cdots & \sum_{i=1}^n E[\tilde{x}_{i1}\tilde{x}_{ip}] \\ \sum_{i=1}^n E[\tilde{x}_{i2}] & \sum_{i=1}^n E[\tilde{x}_{i1}\tilde{x}_{i2}] & \cdots & \sum_{i=1}^n E[\tilde{x}_{i2}\tilde{x}_{ip}] \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^n E[\tilde{x}_{ip}] & \sum_{i=1}^n E[\tilde{x}_{i1}\tilde{x}_{ip}] & \cdots & \sum_{i=1}^n E[(\tilde{x}_{ip})^2] \end{pmatrix}. \quad (43)$$

The augmented matrix of the linear system of equations is

$$B = (A, b). \quad (44)$$

If the rank of the coefficient matrix A is equal to the rank of the augmented matrix B , then the linear system of equations has a unique set of numerical solutions. The unknown parameters $\beta_0, \beta_1, \beta_2, \cdots, \beta_p$ can be estimated by elementary row transformation of the augmented matrix B . Therefore, the linear regression equation can also be solved by elementary transformation of matrix.

5 Numerical example

In order to verify the feasibility of the equation deformation method, we give an example of imprecisely observed data and compared it with the least square estimation. Furthermore, we numerically analyzed the estimated expected values and variances of the disturbance terms by using the methods of References [9] and [24].

Assumed that $(\beta_0, \beta_1) = (5, 2)$, and the one-dimensional linear regression model is

$$y = 5 + 2x + \varepsilon. \quad (45)$$

For the above one-dimensional linear regression model, we designed two sets of imprecisely observed data $(\tilde{x}_{ji}, \tilde{y}_i)$, $i = 1, 2, \cdots, 10$, $j = 1, 2$ as shown in Table 1. The first set of imprecise data $(\tilde{x}_{1i}, \tilde{y}_i)$, $i = 1, 2, \cdots, 10$ is the normal dataset. The second set of imprecise observations $(\tilde{x}_{2i}, \tilde{y}_i)$, $i = 1, 2, \cdots, 10$ is called a singular dataset, which contains outliers for $i = 3, 5, 8$. Now we compare the estimated values of unknown parameters by the equation deformation method and the least squares estimate.

Table 1 Two sets of imprecisely observed data

i	\tilde{y}_i	\tilde{x}_{1i}	\tilde{x}_{2i}
1	$\mathcal{L}(5, 7)$	$\mathcal{L}(0,1)$	$\mathcal{L}(0,1)$
2	$\mathcal{L}(7, 9)$	$\mathcal{L}(1,2)$	$\mathcal{L}(1,2)$
3	$\mathcal{L}(15, 17)$	$\mathcal{L}(5,6)$	$\mathcal{L}(8,9)$
4	$\mathcal{L}(9, 11)$	$\mathcal{L}(2,3)$	$\mathcal{L}(2,3)$
5	$\mathcal{L}(25, 27)$	$\mathcal{L}(10,11)$	$\mathcal{L}(5,6)$
6	$\mathcal{L}(13, 15)$	$\mathcal{L}(4,5)$	$\mathcal{L}(4,5)$
7	$\mathcal{L}(17, 19)$	$\mathcal{L}(6,7)$	$\mathcal{L}(6,7)$
8	$\mathcal{L}(19, 21)$	$\mathcal{L}(7,8)$	$\mathcal{L}(3,4)$
9	$\mathcal{L}(33, 35)$	$\mathcal{L}(14,15)$	$\mathcal{L}(14,15)$
10	$\mathcal{L}(23, 25)$	$\mathcal{L}(9,10)$	$\mathcal{L}(9,10)$

For the first set of imprecise data, the regression equations obtained by equation deformation method and least square estimation are consistent with the original equation, which will not be discussed in detail here. The results obtained from the second set of imprecise data are discussed in detail as follows

The fitting linear regression equations obtained by the equation deformation method and least squares estimate are shown in the Table 2.

Table 2 The linear regression equations

Model	Regression equations
equation deformation method	$y_1 = 6.5965 + 1.9304x_1$
least squares estimate	$y_2 = 7.4839 + 1.7747x_2$

As can be seen from Table 2, the two fitting equations are significantly different, and the constant terms are greatly different, and the fitting effect is greatly different.

The bias between the estimated values of β_0 and β_1 and the corresponding values of the original equation is shown in Table 3. The bias here is the estimated values minus the original values.

Table 3 The bias of the second set of data

Model	β_0	β_1
equation deformation method	1.5965	- 0.0696
least squares estimate	2.4839	- 0.2253

It can be seen from Table 3 that the equation deformation method can better deal with the data with singular values, and the deviation of the corresponding coefficient is small.

According to Equations (11) and (12), the expected value and variance of the disturbance term are

$$\hat{e} = 0.0002, \hat{\sigma}^2 = 18.1316.$$

The expected value of the disturbance term is almost zero, which indicates that the fitting effect is good. The vari-

ance of the disturbance term is large because the data we set contains three singular values and the degree of dispersion is large.

Numerical examples show that the equation transformation method is feasible, easy to understand and simple to calculate.

6 Conclusion

In this paper, we discussed the uncertain regression model, and put forward the equation deformation method to solve the unknown parameters in the linear regression equation. Then the equation deformation method is extended to multiple linear regression model. We also proposed the Cramer's rule and the elementary transformation method of matrix, both of which can solve the unknown parameters of the linear regression equation, but required the reader have a basic knowledge of linear algebra.

Equation deformation method does not require advanced mathematical knowledge such as calculus, so readers can better understand and use. In the case of fewer unknown parameters, the equation deformation method is relatively simple to solve the unknown parameters. However, when the number of unknown parameters is large, the calculation amount of equation deformation method is large.

This paper presents an equation deformation method for solving unknown parameters in the linear regression equation. The next work is to try to use Matlab or Python programming to solve the numerical solution of unknown parameters, so as to better solve the linear regression equation with more complex data and more variables.

Acknowledgements This work was supported by the National Natural Science Foundation of China (No. 11701338) and the Natural Science Foundation of Shandong Province (No. ZR2014GL002).

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

Authorship Contributions

Shuai Wang contributed to the conception of the study. Hongmei Shi contributed significantly to analysis and manuscript preparation; Shuai Wang performed the data analyses and wrote the manuscript; Yufu Ning helped perform the analysis with constructive discussions.

References

1. Chatterjee, S., Hadi, S., Regression analysis by example, 4th edn, Canada: Wiley, 2006.
2. Ma L, Regression analysis, China Machine Press, China: Beijing, 2014.
3. Kahneman D, Tversky A, Prospect theory: an analysis of decision under risk. *Econometrica* 47(2):263-292, 1979.
4. Liu B, Uncertainty Theory, 2nd edn, Springer-Verlag, Germany: Berlin, 2007.
5. Liu B, Uncertainty Theory, 4th edn, Springer-Verlag, Germany: Berlin, 2015.
6. Liu B, Uncertainty Theory: A Branch of Mathematics for Modeling Human Uncertainty, Springer-Verlag, Germany: Berlin, 2010.
7. Liu B, Some research problems in uncertainty theory, *Journal of Uncertain Systems* 3(1):3-10, 2009.
8. Yao K, Liu B, Uncertain regression analysis: An approach for imprecise observations. *Soft Computing* 22(17):5579-5582, 2018.
9. Lio W, Liu B, Residual and confidence interval for uncertain regression model with imprecise observations. *Journal of Intelligent & Fuzzy Systems* 35(2):2573-2583, 2018.
10. Liu B, Why is there a need for uncertainty theory? *Journal of Uncertain Systems* 6(1):3-10, 2012.
11. Liu B, Uncertainty theory, 5th edn., Springer, Berlin, 2017.
12. Song Y, Fu Z, Uncertain multivariable regression model. *Soft Computing* 22:5861-5866, 2018.
13. Chen D, Tukey's biweight estimation for uncertain regression model with imprecise observations. *Soft Computing* 24:16803-16809, 2020.
14. Chen X, Ralescu D, B-spline method of uncertain statistics with application to estimating distance. *Journal of Uncertain Systems* 6(4):256-262, 2012.
15. Wang X, Peng Z, Method of moments for estimating uncertainty distributions. *Journal of Uncertainty Analysis and Applications* 2:5, 2014.
16. Guo H, Wang X, Gao Z, Uncertain linear regression model and its application. *Journal of Intelligent Manufacturing*, <https://doi.org/10.1007/s10845-014-1022-4>.
17. Department of mathematics, Tongji University, Linear algebra, 6nd ed., Higher Education Press, China: Beijing, 2014.
18. Wang X, Ning Y, A new stability analysis of uncertain delay differential equations. *Mathematical Problems in Engineering*, Article ID 1257386, pp.1-8, 2019.
19. Wang X, Ning Y, An Uncertain currency model with floating interest rates. *Soft Computing* 21(22):6739-6754, 2017.
20. Ning Y, Yan L, Xie Y, Mean-TVaR model for portfolio selection with Uncertain returns. *Information-An International Interdisciplinary Journal* 16(2):977-985, 2013.
21. Ning Y, Liu J, Yan L, Uncertain aggregate production planning. *Soft Computing* 17(4):617-624, 2013.
22. Ning Y, Pang N, Wang X, An Uncertain aggregate production planning model considering investment in vegetable preservation technology. *Mathematical Problems in Engineering*, Article ID 8505868, 1-10, 2019.
23. Liu Y, Ha M, Expected value of function of Uncertain variables. *Journal of Uncertain Systems* 4(3):181-186, 2010.
24. Liu Z, Yang Y, Least absolute deviations estimation for uncertain regression with imprecise observations. *Fuzzy Optimization and Decision Making* 19(1):33-52, 2020.