

International Journal of Dynamical Systems and Differential Equations

ISSN online: 1752-3591 - ISSN print: 1752-3583

<https://www.inderscience.com/ijdsde>

Improvements and application of normal cloud positive regression model

Jiandong He, Xiaoting Tan, Chaolin Liu

DOI: [10.1504/IJDSDE.2025.10071144](https://doi.org/10.1504/IJDSDE.2025.10071144)

Article History:

Received:	05 March 2024
Last revised:	22 June 2024
Accepted:	02 December 2024
Published online:	27 June 2025

Improvements and application of normal cloud positive regression model

Jiandong He*

Jiaxing Vocational and Technical College,
Jiaxing 314000, Zhejiang, China
Email: hjd@jxvtc.edu.cn

*Corresponding author

Xiaoting Tan

College of Mathematics and Statistics,
Chongqing University,
Chongqing City, China
and
School of Mathematics and Statistics,
Wuhan University,
Wuhan City, China
Email: 1153659376@qq.com

Chaolin Liu

College of Mathematics and Statistics,
Chongqing University,
Chongqing City, China
and
Provincial Demonstration Center for Experimental Mathematics
Education,
Chongqing University,
Chongqing City, China
Email: lcl@cqu.edu.cn

Abstract: The normal cloud model, as an important model for simultaneously studying the randomness and fuzziness of uncertainty problems, not only covers various disciplines and fields in practical applications but has also been promoted by scholars in theoretical research since its publication. This article introduces the normal cloud model and the cloud least squares method based on discrete envelope distance. In practical applications, the cloud least squares method may generate low accuracy or results that contradict the actual situation. To solve these problems, this article presents the improvements of the normal cloud linear regression model, which leads to the normal cloud ridge regression (NCR) and normal cloud positive regression models. Their advantages have been illustrated through simulation data analysis respectively. Finally, the model was applied to actual stock data to demonstrate the value of our modification once again.

Keywords: normal cloud model; normal cloud ridge regression model; cloud least squares method; normal cloud positive regression; NCP; mean square error.

Reference to this paper should be made as follows: He, J., Tan, X. and Liu, C. (2025) ‘Improvements and application of normal cloud positive regression model’, *Int. J. Dynamical Systems and Differential Equations*, Vol. 14, Nos. 1/2, pp.123–142.

Biographical notes: Jiandong He is a Professor at the Department of Basic Courses, Jiaying Vocational and Technical College. His main research interests include statistical modelling and mathematical modelling pedagogy.

Xiaoting Tan is a graduate student at the School of Mathematics and Statistics, Wuhan University. Her main research interests include biostatistics. She earned her Bachelor’s in Statistics from the College of Mathematics and Statistics, Chongqing University.

Chaolin Liu is an Associate Professor at the School of Mathematics and Statistics, Chongqing University. His main research interests include statistical modelling, statistical computing, and fuzzy statistics.

1 Introduction

As a bidirectional cognitive computing model for qualitative and quantitative transformation, the normal cloud model reveals the internal correlation between randomness and fuzziness in statistics. Since the cloud model was proposed by Prof D.Y. Li in 1995 (Li and Liu, 2004), after years of research and development, it has been covered more and more widely in scientific research. In terms of theoretical research (Yang et al., 2018; Xu et al., 2014), a large number of scholars have linked various theories of randomness and fuzziness to cloud models by analogy, combined with the characteristics of cloud models themselves, and creatively studied various theories of cloud models, making them increasingly mature in theoretical research (Liu, 2022; Nishat et al., 2024). In terms of practical application (Zhang, 2024; Sirbiladze et al., 2014), scholars in different disciplines take into account the randomness and fuzziness of indicators and problems (Bo and Fang, 2015; Song and Zhu, 2019), and look at problems from a new perspective with the help of the normal cloud model (Chhatria and Tripathy, 2023; Xu, 2022), so that the normal cloud model has been applied to various disciplines and fields (Guo et al., 2022; Guo and Guo, 2024).

Considering the correlation between ambiguity and randomness in uncertainty regression, the traditional linear regression model is extended and generalised. Gong et al. (2017) proposed the cloud least squares method based on discrete envelope distance as a generalisation of the traditional regression model and fuzzy regression model. Normal cloud linear regression considers the deviation between data and its estimated value as the result of the fuzzy and random association of system parameters. The three digital features of a normal cloud can reflect the uncertainty degree of the regression model, which has the advantage of being more in line with the connotation of natural language values.

Similar to classical linear regression, cloud least squares method plays a crucial role in normal cloud linear regression. However, in practical applications, people may handle large-scale regression problems that involve many independent variables. In addition, there are some similar properties between the independent variables, which may lead to low accuracy and unstable results in the application of cloud least squares method. How to solve the problems of low accuracy and instability has become an urgent issue for statisticians.

For classical linear regression, current literature mainly focuses on model evaluation and does not consider the issue of multicollinearity. As is well known, multicollinearity between regression variables in linear regression analysis can lead to highly unstable least squares estimates of regression parameters. For multicollinearity data, some coefficients may not be statistically significant and may have incorrect symbols. To solve this problem, some local estimation methods have been proposed to improve the accuracy of the least-squares method. Hoerl and Kennard (1970) proposed ridge regression estimation, which has become the most common method to overcome the special weakness of least squares estimation. In addition, many scholars have studied non negative coefficient regression and gradually enriched the theory of non-negative coefficient regression from various aspects such as different model construction methods and algorithm calculation methods. Peng (2023) analysed the advantages and disadvantages of different non-negative coefficient linear regression models based on domestic and foreign securities market data, indicating that non-negative coefficient regression in classical linear regression is becoming increasingly abundant and gradually improving.

In this article, we proposed a normal cloud ridge estimation method and a normal cloud positive regression (NCP) estimation method for the normal cloud regression model. This method overcomes the potential for low accuracy and instability in cloud least squares determination. At the same time, we also analysed the superiority of the new method compared to existing methods, and demonstrated the feasibility of the new method through simulation data analysis. Finally, we take stock prices as an example to illustrate the advantages and value of this model in practical applications.

The article is organised as follows. In Sections 2 and 3, related work and evaluation criteria of the normal cloud linear regression model are introduced. In Section 4, the normal cloud ridge regression (NCR) model and NCP model are obtained, and we also discuss their properties and simulated analysis for these two estimators. Furthermore, a real data example is given in Section 5. Finally, some conclusions are given in Section 6.

2 Related work

In the real world, uncertainty mainly includes randomness and fuzziness. When the uncertainty of observed variables is determined by membership functions rather than probability distributions, the corresponding regression model is called a fuzzy linear regression model. Randomness and fuzziness are two fundamental characteristics in uncertainty problems. Classic linear regression models and fuzzy regression models have studied these two types of uncertainties separately, but the study of the correlation between the two has not received enough attention, that is, considering both randomness and fuzziness simultaneously. In order to deal with the widespread randomness and fuzziness in qualitative concepts, Li and Du (2014) first proposed a cloud model that uses

a second-order Gaussian distribution method to reflect the randomness of qualitative concepts, while also calculating the fuzziness that reflects qualitative concepts. Li et al. (2009) proposed a new cognitive model loud model, which can synthetically describe the randomness and fuzziness of concepts. They also showed that a normal cloud model can not only be viewed as a generalised normal distribution with weak constraints but also avoid the flaw of fuzzy sets to quantify the membership degree of an element as an accurate value between 0 and 1 and, therefore, may be more adaptive for uncertainty description of linguistic concepts. Ma et al. (2023) proposed an approach based on a fuzzy cloud cognitive diagnosis framework (FC-CDF) to predict examinees' performance in an e-learning environment and utilised innovatively the normal cloud models to measure the expectation, degree of variation and variation frequency of learners' skill proficiency, and each NCM is transformed into an interval fuzzy number to characterise the uncertainty of the skill proficiency for every learner. Gong et al. (2018) considered the fuzzy linear regression model with fuzzy input, fuzzy output, and crisp parameters and combined centroid point and radius of gyration point for defuzzification from the viewpoint of geometric quality. They merged the least square method with the new GCP distance and proposed the least square GCP distance method to estimate regression coefficients. Chen et al. (2013) proposed a simple two-stage approach for constructing a fuzzy regression model based on the distance concept and showed a mathematical programming model to determine the fuzzy adjustment term in the proposed fuzzy regression model to minimise the total estimation error based on the distance concept. The results showed that the proposed fuzzy regression model has the highest explanatory power based on the total estimation error using various criteria. Choi et al. (2013) have successfully combined ridge regression with the fuzzy regression model to address the multicollinearity in fuzzy regression models and the proposed algorithm uses the α -level estimation method to evaluate the parameters of the ridge fuzzy regression model. Li et al. (2023) proposed an improved method for estimating the similarity between LR-type fuzzy numbers. They also showed an algorithm for computing the proposed similarity measure between LR-type fuzzy numbers and computed the similarity between trapezoidal fuzzy numbers, Gaussian fuzzy numbers, and normal fuzzy numbers. Min et al. (2022) discussed the similarities and differences between the fuzzy transform and least-squares fuzzy transform on the least-squares fuzzy transform that combines a fuzzy transform with a least-squares algorithm, and showed that the fuzzy transform outperforms the least-squares fuzzy transform in terms of image reconstruction and image denoising. On the basis of cloud model theory, this article extends and generalises the existing normal cloud linear regression methods to address the correlation between fuzziness and randomness in uncertainty regression. The NCR method and NCP method are proposed to estimate parameters, and the effectiveness and feasibility of the methods are discussed.

3 Basic theory and evaluation criteria of normal cloud linear regression model

The normal cloud model is the uncertainty transformation model between a qualitative concept expressed in language value and its quantitative representation. To describe such a model that combines randomness and fuzz, the normal cloud model expresses the numerical characteristics of a qualitative concept with an independent set of parameters,

reflecting the uncertainty of the concept, namely, expected value $Ex(Expected Value)$, entropy $En(Entropy)$, and hyper entropy $He(Hyper Entropy)$. The formula of the normal cloud model is $x = NC(Ex, En, He)$.

For a multiple linear regression model $y = a_0 + a_1x_1 + a_2x_2 + \dots + a_px_p$, presume that both the independent and dependent variables are in a normal cloud form with both randomness and fuzziness, and the model coefficient $a_j(j = 1, 2, \dots, p)$ is a real number, then the regression model is called a normal cloud linear regression model. Namely, if $(x_{1i}, x_{2i}, \dots, x_{pi}, y_i)$ ($i = 1, 2, \dots, n$) is the experimental data of $(x_1, x_2, \dots, x_p, y)$, then the data of the variable is in the form of $x_{ji} = NC(Ex_{ji}, En_{ji}, He_{ji})$, $y_i = NC(Ex_i, En_i, He_i)$.

According to the definition of the discrete envelope distance between normal clouds (Liu, 2022), the mean square error between the estimated and observed values of the normal cloud-dependent variable in the normal cloud linear regression model is:

$$E_\alpha = \frac{1}{4m} \sum_{i=1}^n \sum_{k=1}^m \left([y_{i,r1}(k/m) - \hat{y}_{i,r1}(k/m)]^2 + [y_{i,r2}(k/m) - \hat{y}_{i,r2}(k/m)]^2 + [y_{i,l1}(k/m) - \hat{y}_{i,l1}(k/m)]^2 + [y_{i,l2}(k/m) - \hat{y}_{i,l2}(k/m)]^2 \right) \\ = \frac{1}{4m} \sum_{i=1}^n \sum_{k=1}^m \left(\begin{aligned} & \left[a_0 + a_1(x_{1i})_{r1}(k/m) + a_2(x_{2i})_{r1}(k/m) + \dots + a_p(x_{pi})_{r1}(k/m) - y_{i,r1}(k/m) \right]^2 \\ & + \left[a_0 + a_1(x_{1i})_{r2}(k/m) + a_2(x_{2i})_{r2}(k/m) + \dots + a_p(x_{pi})_{r2}(k/m) - y_{i,r2}(k/m) \right]^2 \\ & + \left[a_0 + a_1(x_{1i})_{l1}(k/m) + a_2(x_{2i})_{l1}(k/m) + \dots + a_p(x_{pi})_{l1}(k/m) - y_{i,l1}(k/m) \right]^2 \\ & + \left[a_0 + a_1(x_{1i})_{l2}(k/m) + a_2(x_{2i})_{l2}(k/m) + \dots + a_p(x_{pi})_{l2}(k/m) - y_{i,l2}(k/m) \right]^2 \end{aligned} \right) \quad (1)$$

The cloud least squares method is to estimate the linear regression coefficient of the normal cloud $\hat{a}_j$, whose partial derivative is equal to zero, by solving E_α with $a_j(j = 1, 2, \dots, p)$. Let the coefficient vector be

$$A = [a_0, a_1, a_2, \dots, a_p]^T$$

The normal cloud independent variable envelope endpoint value vector and matrix are respectively:

$$X_{i,k}(\alpha) = [1, (x_{1i})_k(\alpha), (x_{2i})_k(\alpha), \dots, (x_{pi})_k(\alpha)], \quad k \in \{r1, r2, l1, l2\} \\ X_k(\alpha) = \begin{bmatrix} 1, (x_{11})_k(\alpha), (x_{21})_k(\alpha), \dots, (x_{p1})_k(\alpha) \\ 1, (x_{12})_k(\alpha), (x_{22})_k(\alpha), \dots, (x_{p2})_k(\alpha) \\ \dots \\ 1, (x_{1n})_k(\alpha), (x_{2n})_k(\alpha), \dots, (x_{pn})_k(\alpha) \end{bmatrix}, \quad k \in \{r1, r2, l1, l2\}$$

The normal cloud-dependent variable envelope endpoint value vector is:

$$Y_k(\alpha) = [1, y_{1,k}(\alpha), y_{2,k}(\alpha), \dots, y_{n,k}(\alpha)]^T, \quad k \in \{r1, r2, l1, l2\}$$

Then the mean square error of equation (1) can be expressed as:

$$E_\alpha = \frac{1}{4m} \sum_{k=1}^m \left\{ \|X_{r1}(k/m)A - Y_{r1}(k/m)\|^2 + \|X_{r2}(k/m)A - Y_{r2}(k/m)\|^2 \right. \\ \left. + \|X_{l1}(k/m)A - Y_{l1}(k/m)\|^2 + \|X_{l2}(k/m)A - Y_{l2}(k/m)\|^2 \right\} \quad (2)$$

Set $\frac{\partial E_\alpha}{\partial A} = 0$, we can get:

$$\frac{2}{4m} \sum_{k=1}^m \begin{pmatrix} X_{r1}^T(k/m)X_{r1}(k/m)A - X_{r1}^T(k/m)Y_{r1}(k/m) \\ + X_{r2}^T(k/m)X_{r2}(k/m)A - X_{r2}^T(k/m)Y_{r2}(k/m) \\ + X_{l1}^T(k/m)X_{l1}(k/m)A - X_{l1}^T(k/m)Y_{l1}(k/m) \\ + X_{l2}^T(k/m)X_{l2}(k/m)A - X_{l2}^T(k/m)Y_{l2}(k/m) \end{pmatrix} = 0 \quad (3)$$

If the matrix

$$\sum_{k=1}^m \begin{pmatrix} X_{r1}^T(k/m)X_{r1}(k/m) + X_{r2}^T(k/m)X_{r2}(k/m) \\ + X_{l1}^T(k/m)X_{l1}(k/m) + X_{l2}^T(k/m)X_{l2}(k/m) \end{pmatrix}$$

is reversible, then the least square estimation of the normal cloud linear regression can be obtained as:

$$\hat{A} = W^{-1}V \quad (4)$$

W, V are respectively:

$$W = \sum_{k=1}^m \begin{pmatrix} X_{r1}^T(k/m)X_{r1}(k/m) + X_{r2}^T(k/m)X_{r2}(k/m) \\ + X_{l1}^T(k/m)X_{l1}(k/m) + X_{l2}^T(k/m)X_{l2}(k/m) \end{pmatrix} \\ V = \sum_{k=1}^m \begin{pmatrix} X_{r1}^T(k/m)Y_{r1}(k/m) + X_{r2}^T(k/m)Y_{r2}(k/m) \\ + X_{l1}^T(k/m)Y_{l1}(k/m) + X_{l2}^T(k/m)Y_{l2}(k/m) \end{pmatrix}$$

The normal cloud linear regression least squares parameter estimation method based on discrete envelope distance is called the cloud least squares method (CLS: Cloud Least Squares), $\hat{A}$ is the cloud least squares estimator, and the normal cloud linear regression model obtained according to the cloud least squares is: $\hat{Y} = X\hat{A}$

Property 1: The estimator $\hat{A}$ is an unbiased estimate of A .

Proof:

$$E(\hat{A}) = W^{-1}E(V) \\ = W^{-1} \left[\sum_{k=1}^m \begin{pmatrix} X_{r1}^T(k/m)E[Y_{r1}(k/m)] + X_{r2}^T(k/m)E[Y_{r2}(k/m)] \\ + X_{l1}^T(k/m)E[Y_{l1}(k/m)] + X_{l2}^T(k/m)E[Y_{l2}(k/m)] \end{pmatrix} \right] \quad (5)$$

$$\begin{aligned}
 E[Y_{r1}(k/m)] &= E \begin{bmatrix} y_{1,r1}(k/m) \\ y_{2,r1}(k/m) \\ \dots \\ y_{n,r1}(k/m) \end{bmatrix} \\
 &= \begin{bmatrix} a_0 + a_1(x_{11})_{r1}(k/m) + a_2(x_{21})_{r1}(k/m) + \dots + a_p(x_{p1})_{r1}(k/m) \\ a_0 + a_1(x_{12})_{r1}(k/m) + a_2(x_{22})_{r1}(k/m) + \dots + a_p(x_{p2})_{r1}(k/m) \\ \dots \\ a_0 + a_1(x_{1n})_{r1}(k/m) + a_2(x_{2n})_{r1}(k/m) + \dots + a_p(x_{pn})_{r1}(k/m) \end{bmatrix} \\
 &= X_{r1}(k/m)A
 \end{aligned}$$

We can also get $E[Y_{r2}(k/m)]$, $E[Y_{l1}(k/m)]$, $E[Y_{l2}(k/m)]$, by substituting (5), we get:

$$\begin{aligned}
 E(\hat{A}) &= W^{-1} \left[\sum_{k=1}^m \begin{pmatrix} X_{r1}^T(k/m)X_{r1}(k/m)A + X_{r2}^T(k/m)X_{r2}(k/m)A \\ + X_{l1}^T(k/m)X_{l1}(k/m)A + X_{l2}^T(k/m)X_{l2}(k/m)A \end{pmatrix} \right] \\
 &= W^{-1}WA = A
 \end{aligned}$$

We can get the conclusion: the estimator $\hat{A}$ is an unbiased estimate of A . q.e.d

To evaluate the performance of the normal cloud linear regression model, the error of the model must be estimated. Similar to classical linear regression, the mean square error of the model was used for comparison. Although the two names sound the same, the classical regression analysis uses mean square error comparison to compare the distance between the observed value and the fitted value by using the point-to-point gap to evaluate the fitting effect. In comparison, the normal cloud fitting focuses on the gap between the actual cloud droplet and the estimated cloud droplet by using the discrete envelope distance difference to explain whether the envelope curve between the actual cloud and the synthetic cloud is similar. Thus, the gap between the actual cloud droplets and the estimated cloud droplets is illustrated, and the quality of the model is further explained. Therefore, when E_{α} the obtained fitting value and the actual value are relatively smaller, we believe that the normal cloud linear regression model is good.

Definition 1: For two normal cloud regression models obtained from the same sample data, the mean square error of Model 1 is $E_{\alpha1}$, and the mean square error of Model 2 is $E_{\alpha2}$. If $E_{\alpha1} < E_{\alpha2}$ so, then Model 1 is superior to Model 2 in the sense of mean square error.

4 Improvement of normal cloud regression model

4.1 The significance of improvement of normal cloud regression model

Since the performance of a normal cloud model is evaluated by considering whether the envelope curves are similar in practice, the mean square error of a normal cloud linear regression model is as follows:

$$\begin{aligned}
E_{\alpha}(\hat{Y}) &= \frac{1}{4m} \sum_{k=1}^m \left\{ \left\| X_{r1}(k/m)\hat{A} - X_{r1}(k/m)A \right\|^2 + \left\| X_{r2}(k/m)\hat{A} - X_{r2}(k/m)A \right\|^2 \right. \\
&\quad \left. + \left\| X_{l1}(k/m)\hat{A} - X_{l1}(k/m)A \right\|^2 + \left\| X_{l2}(k/m)\hat{A} - X_{l2}(k/m)A \right\|^2 \right\} \\
&= \frac{1}{4m} \sum_{k=1}^m \left\{ \left\| X_{r1}(k/m)(\hat{A} - A) \right\|^2 + \left\| X_{r2}(k/m)(\hat{A} - A) \right\|^2 \right. \\
&\quad \left. + \left\| X_{l1}(k/m)(\hat{A} - A) \right\|^2 + \left\| X_{l2}(k/m)(\hat{A} - A) \right\|^2 \right\}
\end{aligned} \tag{6}$$

In this way, the mean square error $E_{\alpha}(\hat{Y})$ estimated by the normal cloud linear regression model can be reflected by the mean square error $MSE(\hat{A}) = E(\hat{A} - A)^T(\hat{A} - A)$ of the coefficient estimator. The smaller $MSE(\hat{A})$ means smaller $E_{\alpha}(\hat{Y})$, which means the model works better.

In traditional linear regression, the mean square error of the coefficient estimator obtained by the least squares method is proportional to the sum of the reciprocal eigenvalues of the matrix $X^T X$. When there is multicollinearity between the independent variables, the minimum eigenvalue is very small, resulting in a very large mean square error and poor simulation effect.

Different from traditional linear regression, when the multicollinearity between variables is not serious to a certain extent in normal cloud linear regression, most of the effects of such multicollinearity are balanced out according to the calculation of the envelope endpoint value in the solution process of cloud least squares method. In this case, the cloud least squares method is still preferable.

However, when the multicollinearity between variables is very serious, the influence brought by multicollinearity is still not very satisfactory after most of it is neutralised. Meanwhile, to seek the minimum mean square error, cloud least squares sometimes ignore the problem itself, resulting in a large change in the result after a slight transformation of the data. That is, the result is unstable. At the same time, it may cause problems such as unreasonable results. To solve such problems, combined with the traditional regression improvement method, the current study puts forward the NCR model, which has a certain improvement compared with the cloud least squares method in solving very serious multicollinearity problems under the evaluation standard of mean square error. Meanwhile, the fitting results of different data generated by the computer for the same problem are relatively stable and feasible.

In addition, in some practical problems, negative coefficients are of no practical significance, and the regression of non-negative coefficients has to be considered in the regression process, while the positive and negative coefficients generated by the cloud least squares method are uncertain. To solve such problems, combined with the positive regression in classical regression, the NCP is proposed in the current study to make the coefficient estimators have practical significance.

4.2 The NCR model

4.2.1 Model construction

In 1970, Hoerl and Kennard proposed ridge estimators in the traditional sense of regression (Li et al., 2023; Min et al., 2022), which provides a good solution for the presence of multicollinearity in traditional data. Concerning the ridge estimation under

traditional regression, the normal cloud ridge estimation is presented here. Consider coefficient estimators:

$$\hat{A}(k) = (W + kI)^{-1}V \quad (7)$$

The coefficient estimator called normal cloud ridge regression is the NCR model. $\hat{Y} = X\hat{A}(k)$ is NCR model. Where I is the order unit matrix and k is the ridge parameter, and the mean square error of the model is reduced by selecting the appropriate ridge parameter, where the corresponding coefficient estimator is the cloud least squares estimator.

By definition, the NCR estimator has the following properties:

Property 2: For arbitrary $k > 0$, $\hat{A}(k)$ is a biased estimate of A .

Proof: According to equation (7), there are:

$$\begin{aligned} \hat{A}(k) &= (W + kI)^{-1}V = (W + kI)^{-1}WW^{-1}V = (W + kI)^{-1}W\hat{A} \\ E(\hat{A}(k)) &= (W + kI)^{-1}WE(\hat{A}) = (W + kI)^{-1}WA \end{aligned}$$

So when we have $k > 0$, $\hat{A}(k)$ is a biased estimate of A . q.e.d

Property 3: There exists $k > 0$ that makes $MSE(\hat{A}(k)) < MSE(\hat{A})$, then we get $E_\alpha(\hat{Y}(k)) < E_\alpha(\hat{Y})$, that is, in the sense of mean square error, there is a ridge regression estimation model better than the cloud least squares method.

Proof:

$$\begin{aligned} MSE(\hat{A}(k)) &= E(\hat{A}(k) - \hat{A})^T (\hat{A}(k) - \hat{A}) \\ &= E\left[\hat{A}(k) - E(\hat{A}(k)) + E(\hat{A}(k)) - \hat{A}\right]^T \left[\hat{A}(k) - E(\hat{A}(k)) + E(\hat{A}(k)) - \hat{A}\right] \\ &= E\left[\hat{A}(k) - E(\hat{A}(k))\right]^T \left[\hat{A}(k) - E(\hat{A}(k))\right] \\ &\quad + E\left[\hat{A}(k) - E(\hat{A}(k))\right]^T \left[E(\hat{A}(k)) - \hat{A}\right] \\ &\quad + E\left[E(\hat{A}(k)) - \hat{A}\right]^T \left[\hat{A}(k) - E(\hat{A}(k))\right] \\ &\quad + E\left[E(\hat{A}(k)) - \hat{A}\right]^T \left[E(\hat{A}(k)) - \hat{A}\right] \\ &= trCov(\hat{A}(k)) + \|E(\hat{A}(k)) - \hat{A}\|^2 \end{aligned}$$

Notes that $g(k) = MSE(\hat{A}(k))$, $g_1(k) = trCov(\hat{A}(k))$ and $g_2(k) = \|E(\hat{A}(k)) - \hat{A}\|^2$, we get $g(k) = g_1(k) + g_2(k)$, $k > 0$, among which:

$$\begin{aligned} g_1(k) &= trCov(\hat{A}(k)) \\ &= tr(Cov((W + kI)^{-1}WA)) \\ &= tr((W + kI)^{-1}WCov(A)W(W + kI)^{-1}) \end{aligned}$$

Note that W is a real symmetric matrix, so there are: the eigenvalues of the W are respectively $\lambda_0, \lambda_1, \dots, \lambda_p$, and then the eigenvalues of the covariance matrix $\text{Cov}(A)$ are $\alpha_0, \alpha_1, \dots, \alpha_p$ (here α_j is related to λ_j , but not related to k), so it can be obtained (here is used in higher algebra is that the symmetric matrix W can be expressed as an orthogonal matrix P and the eigenvalue matrix $\text{diag}(\lambda_0, \lambda_1, \dots, \lambda_p)$ of the form of the contract property $W = P \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_p) P^T$ and the trace operation property of the matrix):

$$\begin{aligned} g_1(k) &= \sum_{j=0}^p \frac{\alpha_j \lambda_j^2}{(\lambda_j + k)^2} \\ g_2(k) &= \|E(\hat{A}(k)) - \hat{A}\|^2 = \|((W + kI)^{-1}W - I)A\|^2 = k^2 \|(W + kI)^{-1}A\|^2 \\ &= k^2 \left(\begin{pmatrix} \lambda_0 + k & & 0 \\ & \ddots & \\ 0 & & \lambda_p + k \end{pmatrix}^{-1} \right)^T \left(\begin{pmatrix} \lambda_0 + k & & 0 \\ & \ddots & \\ 0 & & \lambda_p + k \end{pmatrix}^{-1} P^T A \right) \end{aligned}$$

Note that $P^T A = [b_0, b_1, \dots, b_p]^T$, (b_j is related to α_j , but not related to k), so:

$$g_2(k) = \sum_{j=0}^p k^2 \frac{b_j^2}{(\lambda_j + k)^2}$$

Take the derivation of $g_1(k)$ and $g_2(k)$ respectively, and you get

$$g'_1(k) = \sum_{j=0}^p \frac{-2\alpha_j \lambda_j^2}{(\lambda_j + k)^3}, \quad g'_2(k) = \sum_{j=0}^p \frac{2kb_j^2 \lambda_j}{(\lambda_j + k)^3}$$

$g'_1(0) < 0$, $g'_2(0) = 0$ and $g'_1(k)$, $g'_2(k)$ are continuous when $k \geq 0$, so when $k > 0$ and small enough, we get $g'(k) = g'_1(k) + g'_2(k) < 0$, so $g(k)$ is monotone decreasing.

So there exists $k > 0$ that makes $g(k) < g(0)$, us get $MSE(\hat{A}(k)) < MSE(\hat{A})$. q.e.d

The ridge parameters k can still be selected by the cross-validation method. That is, the mean square error of the model is regarded as a function of the ridge parameters, and the ridge parameters are selected so that the mean square error is small enough, and the regression coefficient does not have unreasonable symbols. At the same time, the value k should not be too large, that is, the coefficient estimators should not be too biased.

It should be noted that although the normal cloud ridge estimation should consider that the value of the ridge parameter k should not be too large when selecting a certain ridge parameter value, since it is a matrix W obtained by the addition of 4 m matrices, even if the value k is not too large, it is roughly 4 m times that of the traditional ridge regression.

4.2.2 Simulated data and analysis

Take $n = 10$, using R statistical software, we can generate n groups of random datasets a_1, b_1 , which are independently and equally distributed $N(20, 4^2)$, by the normal cloud data generation method, we set $Ex_1 = (a_1 + b_1) / 2$, $En_1 = |a_1 - b_1| / 6$, $He_1 = En_1 / 6$, the fuzzy simulation data is converted into the sample data of normal cloud variables x_1 (Liu,

2022), and the sample data of the variable x_2 is the sample data of the variable x_1 plus a disturbance term, to build a multicollinearity variable. The linear model of the normal cloud is as follows: $y = 2x_1 + 3x_2 + 1$. In this way, the sample data of the dependent variable y is obtained, as shown in Table 1 (all data are reserved for three decimal places), which is taken from the following fitting $m = 20$.

Table 1 Simulation data I

	x_1	x_2	y
1	NC(19.405, 1.028, 0.171)	NC(19.638, 1.066, 0.178)	NC(98.722, 0.855, 0.142)
2	NC(17.576, 0.589, 0.098)	NC(17.747, 0.577, 0.096)	NC(89.393, 0.476, 0.079)
3	NC(21.612, 0.736, 0.123)	NC(22.188, 0.616, 0.103)	NC(110.788, 0.554, 0.092)
4	NC(16.287, 0.688, 0.115)	NC(16.770, 0.761, 0.127)	NC(83.882, 0.592, 0.099)
5	NC(18.614, 0.190, 0.032)	NC(19.029, 0.263, 0.044)	NC(95.316, 0.187, 0.031)
6	NC(22.892, 0.429, 0.071)	NC(23.712, 0.466, 0.078)	NC(117.919, 0.366, 0.061)
7	NC(17.466, 0.492, 0.082)	NC(17.996, 0.591, 0.099)	NC(89.921, 0.444, 0.074)
8	NC(16.423, 0.143, 0.024)	NC(17.146, 0.211, 0.035)	NC(85.285, 0.147, 0.025)
9	NC(22.920, 0.893, 0.149)	NC(23.439, 0.875, 0.146)	NC(117.156, 0.722, 0.120)
10	NC(16.547, 0.973, 0.162)	NC(17.209, 0.980, 0.163)	NC(85.720, 0.797, 0.133)

Table 2 Simulation data II

	x_1	x_2	y
1	NC(21.275, 0.501, 0.083)	NC(21.806, 0.400, 0.067)	NC(108.969, 0.370, 0.062)
2	NC(21.526, 1.041, 0.174)	NC(22.225, 1.046, 0.174)	NC(110.727, 0.852, 0.142)
3	NC(22.103, 0.159, 0.026)	NC(22.741, 0.188, 0.031)	NC(113.429, 0.142, 0.024)
4	NC(21.273, 0.863, 0.144)	NC(21.576, 0.831, 0.139)	NC(108.276, 0.692, 0.115)
5	NC(18.093, 1.165, 0.194)	NC(18.450, 1.250, 0.208)	NC(92.534, 0.987, 0.164)
6	NC(19.704, 0.252, 0.042)	NC(20.461, 0.296, 0.049)	NC(101.790, 0.225, 0.037)
7	NC(17.509, 0.094, 0.016)	NC(17.909, 0.211, 0.035)	NC(89.744, 0.133, 0.022)
8	NC(19.668, 0.335, 0.056)	NC(20.266, 0.230, 0.038)	NC(101.133, 0.235, 0.039)
9	NC(15.211, 0.831, 0.138)	NC(15.788, 0.837, 0.139)	NC(78.787, 0.681, 0.113)
10	NC(19.711, 3.049, 0.508)	NC(20.565, 3.002, 0.500)	NC(102.118, 2.473, 0.412)

Cloud least squares fitting was performed on the simulation sample data 1, and the results were obtained:

$$\hat{A} = [0.252, 1.401, 3.622]^T, MSE(\hat{A}) = 1.305, E_\alpha(\hat{Y}) = 0.202 \quad (8)$$

Normal cloud ridge estimation and fitting are carried out on simulated sample data 1. When the ridge parameter k is set to 9, the fitting result is obtained:

$$\begin{aligned} \hat{A}(k=9) &= [0.206, 1.767, 3.266]^T \\ MSE(\hat{A}(k=9)) &= 0.755, E_\alpha(\hat{Y}(k=9)) = 0.128 \end{aligned} \quad (9)$$

By comparing the results of equations (8) and (9), it can be found that the mean square error of NCR is smaller than cloud least squares, and the coefficient results are closer to the true value.

The sample data is generated again using the same method, as shown in Table 2:

Cloud least squares fitting was performed on the simulated sample data II, and the results were obtained:

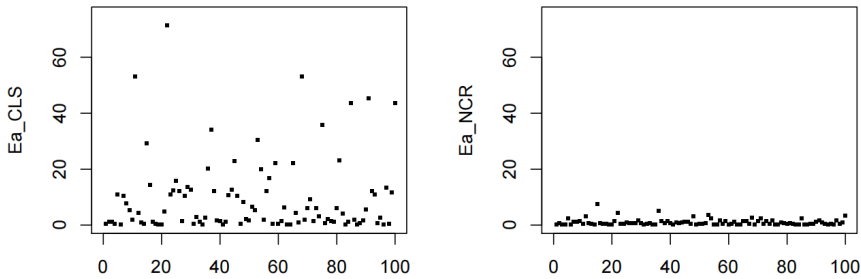
$$\hat{A} = [0.580, -8.727, 13.444]^T, \text{MSE}(\hat{A}) = 224.322, E_\alpha(\hat{Y}) = 30.808 \quad (10)$$

Normal cloud ridge estimation fitting was performed on simulated sample data 2. When the ridge parameter k was 77, the fitting result was obtained:

$$\begin{aligned} \hat{A}(k=77) &= [0.220, 0.477, 4.498]^T, \\ \text{MSE}(\hat{A}(k=77)) &= 5.172, \\ E_\alpha(\hat{Y}(k=77)) &= 2.406 \end{aligned} \quad (11)$$

By comparing the results of equation (10) and equation (11), it can be found that the mean square error of normal cloud ridge estimation is much smaller than that of cloud least squares. Moreover, in the coefficient fitting results, the cloud least squares even have a negative value, which is much different from the real value. In contrast, the NCR simulation results are closer to the real value.

Figure 1 Comparison of simulation results between CLS and NCR



The same method continues to be used to generate 100 sets of sample data many times and carry out fitting. As shown in Figure 1, it is found that the cloud least squares fitting result coefficient changes greatly, and mean square error fluctuates between $[0, 80]$, while NCR fitting result coefficient changes little. The mean square error is smaller than the mean square error of cloud least squares. It also basically changes between $[0, 8]$ and has strong stability.

The simulation in Figure 1 results show that:

- On the one hand, from the perspective of data simulation, when MSE the coefficient estimator decreases, the mean square error E_α of the model also decreases, and vice versa; Moreover, for the simulated sample data with strong multicollinearity, the mean square error of the NCR fitting coefficient estimator and the mean square error of the model are smaller than the cloud least squares method.
- On the other hand, when there is a serious multicollinearity between variables, the cloud least squares method leads to the instability of the fitting results to pursue the

minimum mean square error, while the NCR can make the model more stable in the fitting results and the model mean square error by adjusting the values of the corresponding ridge parameters k .

4.3 NCP model

4.3.1 Model construction

For some practical problems of normal cloud regression, the negative regression coefficient may have no practical physical and economic significance, or even contrary to the facts. For example, using cloud least squares fitting between the stock index and the corresponding stock variables often get some negative coefficients, which means that the stock index falls when the stock rises, which is inconsistent with the fact. In this case, when applying the normal cloud linear regression model to practical problems, it is necessary to combine the problem background and add some non-negative constraints to the coefficients.

This section introduces the NCP, a method for constructing a linear regression model with non-negative coefficients under normal cloud data. It is based on the construction of a traditional regression model with non-negative coefficients, which was proposed by scholar Xie Zhongqiu (Xie, 2006).

The idea is to multiply the negative coefficients (without considering constant coefficients) in the cloud least squares method by a negative correction factor to make it positive. The steps for constructing this model are as follows:

- 1 Find the estimated coefficient of the cloud least squares method $a_0^*, a_1^*, \dots, a_p^*$, where, $a_t^* \geq 0$ $a_s^* \geq 0$ $1 < s, t < p$ and the number of such a_s^* and a_t^* is p .
- 2 For the negative coefficient a_s^* obtained in step 1, multiply it by a negative correction factor $Q < 0$, making the corrected coefficient $a_s = Qa_s^* > 0$. a_t^* changes along with a_s^* . Therefore, when a_t^* multiplied for the same positive correction factor, it can enable $\sum a_s^* x_{si}$ and $\sum a_t^* x_{ti}$ undergo the same degree of change processing;
- 3 The non-negative coefficient linear regression model becomes:

$$\begin{aligned} & \min E_a \\ & \text{s.t.} \begin{cases} a_s = Qa_s^* \\ a_t = -Qa_t^* \\ Q < 0 \end{cases} \end{aligned} \quad (12)$$

- 4 Use Lagrange's multiplier method to solve equation (12), i.e.:

$$\min F = E_a - \sum \mu_s (a_s - Qa_s^*) - \sum \mu_t (a_t + Qa_t^*) \quad (13)$$

Take the partial derivative of F over a_0 and make it equal to zero to obtain:

$$\frac{\partial F}{\partial a_0} = \frac{2}{4m} \sum_{k=1}^m \sum_{i=1}^n \left\{ \begin{aligned} & \left[a_0 + \sum a_s (x_{si})_{r1} (k/m) + \sum a_t (x_{ti})_{r1} (k/m) - y_{i,r1} (k/m) \right] \\ & + \left[a_0 + \sum a_s (x_{si})_{r2} (k/m) + \sum a_t (x_{ti})_{r2} (k/m) - y_{i,r2} (k/m) \right] \\ & + \left[a_0 + \sum a_s (x_{si})_{l1} (k/m) + \sum a_t (x_{ti})_{l1} (k/m) - y_{i,l1} (k/m) \right] \\ & + \left[a_0 + \sum a_s (x_{si})_{l2} (k/m) + \sum a_t (x_{ti})_{l2} (k/m) - y_{i,l2} (k/m) \right] \end{aligned} \right\} \quad (14)$$

$$= 0$$

Substitute $a_s = Qa_s^*$ and $a_t = -Qa_t^*$ into equation (14) to obtain:

$$a_0 = -\frac{1}{n} \sum_{i=1}^n \left(\sum a_s Ex_{si} + \sum a_t Ex_{ti} \right) Q + \frac{1}{n} \sum_{i=1}^n Ex_{yi} \quad (15)$$

- 5 Combine with equation (15), obtain the negative correction factor Q that can minimise the mean square error, that is:

min E_α

$$= \frac{1}{4m} \sum_{k=1}^m \sum_{i=1}^n \left\{ \begin{aligned} & \left[a_0 + \sum a_s (x_{si})_{r1} (k/m) + \sum a_t (x_{ti})_{r1} (k/m) - y_{i,r1} (k/m) \right]^2 \\ & + \left[a_0 + \sum a_s (x_{si})_{r2} (k/m) + \sum a_t (x_{ti})_{r2} (k/m) - y_{i,r2} (k/m) \right]^2 \\ & + \left[a_0 + \sum a_s (x_{si})_{l1} (k/m) + \sum a_t (x_{ti})_{l1} (k/m) - y_{i,l1} (k/m) \right]^2 \\ & + \left[a_0 + \sum a_s (x_{si})_{l2} (k/m) + \sum a_t (x_{ti})_{l2} (k/m) - y_{i,l2} (k/m) \right]^2 \end{aligned} \right\}$$

$$= \frac{1}{4m} \sum_{k=1}^m \sum_{i=1}^n \left[\begin{aligned} & \left[Q \left(\sum a_s (x_{si})_{r1} (k/m) + \sum a_t (x_{ti})_{r1} (k/m) - \frac{1}{n} \sum_{i=1}^n (\sum a_s Ex_{si} + \sum a_t Ex_{ti}) \right) \right. \\ & \left. + \left(\frac{1}{n} \sum_{i=1}^n Ex_{yi} - y_{i,r1} (k/m) \right) \right]^2 + \left[Q \left(\sum a_s (x_{si})_{r2} (k/m) + \sum a_t (x_{ti})_{r2} (k/m) \right. \right. \\ & \left. \left. - \frac{1}{n} \sum_{i=1}^n (\sum a_s Ex_{si} + \sum a_t Ex_{ti}) \right) + \left(\frac{1}{n} \sum_{i=1}^n Ex_{yi} - y_{i,r2} (k/m) \right) \right]^2 \\ & + \left[Q \left(\sum a_s (x_{si})_{l1} (k/m) + \sum a_t (x_{ti})_{l1} (k/m) - \frac{1}{n} \sum_{i=1}^n (\sum a_s Ex_{si} + \sum a_t Ex_{ti}) \right) \right. \\ & \left. + \left(\frac{1}{n} \sum_{i=1}^n Ex_{yi} - y_{i,l1} (k/m) \right) \right]^2 + \left[Q \left(\sum a_s (x_{si})_{l2} (k/m) + \sum a_t (x_{ti})_{l2} (k/m) \right. \right. \\ & \left. \left. - \frac{1}{n} \sum_{i=1}^n (\sum a_s Ex_{si} + \sum a_t Ex_{ti}) \right) + \left(\frac{1}{n} \sum_{i=1}^n Ex_{yi} - y_{i,l2} (k/m) \right) \right]^2 \end{aligned} \right]$$

It can be found that E_α is a univariate quadratic function of Q , and when $Q = -K/2J$, E_α can take the minimum value, where, K and J take the following values:

$$K = \frac{1}{4m} \sum_{k=1}^m \sum_{i=1}^n \left\{ \begin{aligned} & \left[\sum a_s (x_{si})_{r1} (k/m) + \sum a_t (x_{ti})_{r1} (k/m) - \frac{1}{n} \sum_{i=1}^n (\sum a_s Ex_{si} + \sum a_t Ex_{ti}) \right]^2 \\ & + \left[\sum a_s (x_{si})_{r2} (k/m) + \sum a_t (x_{ti})_{r2} (k/m) - \frac{1}{n} \sum_{i=1}^n (\sum a_s Ex_{si} + \sum a_t Ex_{ti}) \right]^2 \\ & + \left[\sum a_s (x_{si})_{l1} (k/m) + \sum a_t (x_{ti})_{l1} (k/m) - \frac{1}{n} \sum_{i=1}^n (\sum a_s Ex_{si} + \sum a_t Ex_{ti}) \right]^2 \\ & + \left[\sum a_s (x_{si})_{l2} (k/m) + \sum a_t (x_{ti})_{l2} (k/m) - \frac{1}{n} \sum_{i=1}^n (\sum a_s Ex_{si} + \sum a_t Ex_{ti}) \right]^2 \end{aligned} \right\}$$

$$J = \frac{2}{4m} \sum_{k=1}^m \sum_{i=1}^n \left\{ \begin{aligned} & \left[\sum a_s (x_{si})_{r1} (k/m) + \sum a_t (x_{ti})_{r1} (k/m) - \frac{1}{n} \sum_{i=1}^n (\sum a_s Ex_{si} + \sum a_t Ex_{ti}) \right] \\ & \left[\frac{1}{n} \sum_{i=1}^n Ex_{yi} - y_{i,r1} (k/m) \right] + \left[\sum a_s (x_{si})_{r2} (k/m) + \sum a_t (x_{ti})_{r2} (k/m) \right. \\ & \left. - \frac{1}{n} \sum_{i=1}^n (\sum a_s Ex_{si} + \sum a_t Ex_{ti}) \right] \left[\frac{1}{n} \sum_{i=1}^n Ex_{yi} - y_{i,r2} (k/m) \right] \\ & + \left[\sum a_s (x_{si})_{l1} (k/m) + \sum a_t (x_{ti})_{l1} (k/m) - \frac{1}{n} \sum_{i=1}^n (\sum a_s Ex_{si} + \sum a_t Ex_{ti}) \right] \\ & \left[\frac{1}{n} \sum_{i=1}^n Ex_{yi} - y_{i,l1} (k/m) \right] + \left[\sum a_s (x_{si})_{l2} (k/m) + \sum a_t (x_{ti})_{l2} (k/m) \right. \\ & \left. - \frac{1}{n} \sum_{i=1}^n (\sum a_s Ex_{si} + \sum a_t Ex_{ti}) \right] \left[\frac{1}{n} \sum_{i=1}^n Ex_{yi} - y_{i,l2} (k/m) \right] \end{aligned} \right\}$$

Here, K is positive. If $J > 0$, take $Q = -K/2J$. If $J < 0$, simply take Q as a negative number sufficiently close to zero to obtain a sufficiently small E_α , that is, the value of Q is:

$$Q = \begin{cases} -\frac{K}{2J} & J > 0 \\ \text{sufficiently close to zero, negative numbers} & J < 0 \end{cases} \quad (16)$$

- 6 Substitute equation (16) into equation (15) to obtain a_0 , and calculate $a_s = Qa_s^*$ and $a_t = -Qa_t^*$ to obtain the normal cloud positive regression coefficient $a_0, a_1, \dots, a_p$.

Compared with the coefficients obtained by cloud least squares, the normal cloud regression coefficients $a_1, \dots, a_p$ obtained through steps 1 to 6 are all non-negative coefficients, which also appear more reasonable in terms of economic and practical significance.

4.3.2 Simulated data and analysis

For the data in Table 2, according to the cloud least square fitting result (16), it can be found that $a_1^* < 0$, $a_2^* \geq 0$, by substituting the data $s = 1$, $t = 2$ in step 1, what can be got as follows:

$$Q = -0.201, \hat{A}^+ = [11.477, 1.754, 2.702]^T, E_\alpha(\hat{Y}^+) = 13.741 \quad (17)$$

According to the results of equation (17), for the data in Table 2, the normal cloud positive regression not only obtains a positive coefficient of a_1 a_2 and, but also makes the final mean square error less than that obtained by the cloud least square method.

The normal cloud positive regression does not necessarily make the mean square error less than the mean square error of the cloud least square method, and sometimes the mean square error obtained is larger or even much larger than the cloud least square method, but it is worthwhile to sacrifice a part of the mean square error when the coefficient is guaranteed to be non-negative and certain physical or economic rationality is taken into account.

5 Application of normal cloud regression model

With the study of the linear relationship between the Shanghai 50 stock indexes and its corresponding 50 stocks, it was found that there was a strong multicollinearity among seven stocks, including CONCH, SANY, CPIC, WHCHEM, SINOPEC, CHINA SECURITIES, and GWM. In particular, CONCH had a significant positive linear correlation coefficient with the other six stocks, as shown in Table 3. Daily data from seven stocks from 3 January 2023 to 18 April 2023 (sourced from <https://www.tdx.com.cn/>) were collected to construct normal cloud data. The normal cloud sample data from 3 January 2023 to 20 March 2023, is used as the training set, with a total of 50 sets of sample data. The normal cloud sample data from 21 March 2023 to 18 April 2023, was used as the testing set, with a total of 25 sets of sample data. Cloud least squares and normal cloud regression models were applied to explore the linear relationship between the normal cloud-dependent variable CONCH and the normal cloud-independent variables SANY, CPIC, WHCHEM, SINOPEC, CHINA SECURITIES, and GWM.

Software R is used for calculation. Take $m = 20$, and use $x_1 \sim x_6$ to represent the respective prices of stocks SANY, CPIC, WHCHEM, SINOPEC, CHINA SECURITIES, and GWM. Use y to represent the stock price of CONCH. The linear cloud model estimation results obtained by the cloud least squares method and NCR are shown in Table 4.

Table 3 Linear correlation coefficients between seven stocks

	CONCH	SANY	CPIC	WHCHEM	SINOPEC	CHINA SECURITIES	GWM
CONCH	1	0.965	0.947	0.909	0.509	0.557	0.887
SANY	0.965	1	0.931	0.859	0.551	0.473	0.849
CPIC	0.947	0.931	1	0.905	0.545	0.591	0.884
WHCHEM	0.909	0.859	0.905	1	0.595	0.603	0.918
SINOPEC	0.509	0.551	0.545	0.595	1	0.235	0.608
CHINA SECURITIES	0.557	0.473	0.591	0.603	0.235	1	0.557
GWM	0.887	0.849	0.884	0.918	0.608	0.557	1

Table 4 Comparison of cloud least squares and positive regression model estimation results

	a_0	a_1	a_2	a_3	a_4	a_5	a_6	E_α
CLS: $\hat{y}$	0.008876	0.532046	0.206601	0.124767	-0.262605	0.194445	-0.043080	6.068798
NCP: $\hat{y}^+$	4.820403	0.379330	0.147299	0.088954	0.187228	0.138632	0.030714	4.816230

Comparing the results, it is seen that the prices of SANY (x_1) have a significant impact on the price of CONCH. In addition, it can be found that the prices of CPIC (x_2) and WHCHEM (x_3) also have a large positive impact on the price of CONCH, which is also consistent with the correlation coefficient among stocks in Table 3.

Comparing and analysing the results of the two methods, it can be seen that: on the one hand, the improved cloud model, that is, the NCR model is better than the cloud least squares method under the criterion of mean square error; On the other hand, by comparing the coefficient estimators of the two models, NCR weakens the influence of the price of SANY (x_1) on the price of CONCH in the model, and increases the influence of the price of CPIC(x_2), which also has a very high linear correlation coefficient, on the price of CONCH stock. At the same time, the positive linear correlation coefficient is not too high, and the influence of the price of CHINA SECURITIES (x_5) on the price of CONCH is also weakened, to reduce the negative impact of the price of SINOPEC(x_4) and GWM (x_6) on the price of CONCH.

Therefore, the results of NCR in this application are better than the cloud least squares method.

In these two models, the price coefficients of SINOPEC(x_4) and GWM (x_6) are negative, but according to the linear correlation coefficients in Table 3, we can see that both of them should have a positive impact on the price of conch cement respectively. Therefore, the results of the cloud least squares method are further improved by considering normal cloud positive regression.

After calculating, $J = 110.3589$ $K = 157.3638$ and correcting factors $Q = -K/2J = -0.7129634$, the normal cloud positive regression linear cloud model is obtained as follows:

$$\begin{aligned}\hat{y}^+ = & 4.820403 + 0.379330x_1 + 0.147300x_2 + 0.088954x_3 \\ & + 0.187228x_4 + 0.138632x_5 + 0.030714x_6\end{aligned}$$

Its mean square error is $E_\alpha(\hat{y}^+) = 4.816230$.

Compare this with the cloud least squares results, as shown in Table 5.

Table 5 Comparison of results between cloud least squares and positive regression models on the test set

	<i>Cloud least square</i>	<i>Normal cloud positive regression</i>
E_α	25.342410	1.650681

According to the results of the normal cloud positive regression, the price of SANY (x_1) and the price of CPIC(x_2) has a great impact on the price of CONCH, which is consistent with the results of the linear correlation coefficient. However, the price of SINOPEC(x_4) also has a greater impact on the price of CONCH, which is not consistent with the linear correlation coefficient. However, it ensures that the influence of the six stock prices on the price of conch cement is positive, and its mean square error is smaller than that obtained by the cloud least square method.

Therefore, the results of normal cloud positive regression in this application are superior to the cloud least squares method.

6 Conclusions

The normal cloud model, as a kind of cloud model with universal adaptability, not only takes into account the randomness and fuzziness of uncertainties but also represents the model through three digital features, which are more in line with natural language features, making the normal cloud model more widely applied.

In this paper, under the theoretical support of the normal cloud model and cloud least squares method, combined with the improvement of traditional regression, the NCR model is proposed. Through theoretical proof and simulation data analysis, the value of the improved model is explained. Considering the physical significance and economic value of the model, the normal cloud positive regression is proposed, and the coefficients in the cloud least squares method are improved by multiplying a correction factor to get more reasonable results. Finally, a case of financial stock data shows that the improved model is valuable under certain evaluation criteria, so that cloud linear regression can also be applied to different types of actual data in different aspects.

As a model construction method of normal cloud linear regression with non-negative coefficients, normal cloud positive regression can effectively transform the negative coefficients of the model coefficients obtained by the cloud least squares method into positive coefficients so that the model can conform to the corresponding logic and significance under the specific problem background. In the last case of the stock price, the effect and significance of the normal cloud positive regression are further illustrated in the context of practical problems.

When using normal cloud to improve models, note that not all multicollinearity data perform better with NCR than with the cloud least squares method. By comparing the mean square error of the two models in Table 4, it can be found that the mean square error of the NCR model is less reduced than that of the cloud least squares method. This shows that due to the clever characteristics of the normal cloud itself, the cloud least squares method can obtain good results even in the face of severe multicollinearity such as this application. In fact, in many practical applications, due to the larger and higher sample size and dimension of the data, if a NCR model is to be adopted at this time, the

calculation amount in selecting a better ridge parameter is very large, time-consuming, and laborious. What's more, the results obtained may be only a little better or even worse than the cloud least squares method, in which the loss outweighs the gain.

References

- Bo, L. and Fang, Y. (2015) 'A portfolio selection model based on fuzzy regression analysis', *System Engineering Theory and Practice*, Vol. 35, No. 7, pp.1770–1776.
- Chen, L.H., Hsueh, C.C. and Chang, C.J. (2013) 'A two-stage approach for formulating fuzzy regression models', *Knowledge-Based Systems*, November, Vol. 52, pp.302–310.
- Chhatra, G.N. and Tripathy, A.K. (2023) 'Riccati technique and oscillation of third order difference equations with impulse', *International Journal of Dynamical Systems and Differential Equations (IJDSE)*, Vol. 13, No. 3, pp.250–269.
- Choi, S.H., Jung, H.Y. and Kim, H. (2013) 'Ridge fuzzy regression model', *International Journal of Fuzzy Systems*, Vol. 21, No. 7, pp.2077–2090.
- Gong, Y.B., Dai, L.L. and Liu, G.G. (2017) 'Normal cloud linear regression model and its least square parameter estimation method', *Statistics and Decision Making*, Vol. 23, No. 1, pp.5–9.
- Gong, Y.B., Yang, S.X., Ma, H.L. and Ge, M. (2018) 'Fuzzy regression model based on geometric coordinate points distance and application to performance evaluation', *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, Vol. 34, No. 1, pp.395–404.
- Guo, M., Zhao, J.W., Pan, D., Sun, M.X., Zhou, Y.Q. and Yan, B.N. (2022) 'Normal cloud model theory-based comprehensive fuzzy assessment of wooden pagoda safety', *Journal of Cultural Heritage*, Vol. 2022, No. 55, pp.1–10.
- Guo, Z. and Guo, X. (2024) 'Weak convergence of truncation error of differential order partial derivative. Equations under mathematical chaos theory', *Engineering Intelligent Systems*, Vol. 32, No. 3, pp.277–284.
- Hoerl, A.E. and Kennard, R.W. (1970) 'Ridge regression – applications to nonorthogonal problems', *Technometrics*, Vol. 12, No. 1, pp.69–82.
- Li, D.Y. and Du, Y. (2014) *Artificial Intelligence with Uncertainty*, National Defense Industry Press, Beijing.
- Li, D.Y. and Liu, C.Y. (2004) 'Study on the universality of the normal cloud model', *China Engineering Science*, Vol. 6, No. 8, pp.28–34.
- Li, D.Y., Liu, C.Y. and Gan, W.Y. (2009) 'A new cognitive model: cloud mode', *International Journal of Intelligent Systems*, Vol. 24, No. 3, pp.357–375.
- Li, Y.F., He, X.X., Meng, D. and Qin, K.Y. (2023) 'An improved method to estimate the similarity between LR-type fuzzy numbers', *Fuzzy Sets and Systems*, Vol. 476, No. 1, pp.1–27.
- Liu, Z.J. (2022) 'Optimization model for linear regression of normal cloud coefficients of ordinary variables', *Statistics and Decision Making*, Vol. 38, No. 8, pp.43–46.
- Ma, H., Huang, Z.X., Zhu, H.B., Tang, W.S., Zhang, H.Y. and Li, K.Q. (2023) 'Predicting examinee performance based on a fuzzy cloud cognitive diagnosis framework in an e-learning environment', *Soft Computing*, Vol. 27, No. 24, pp.18949–18969.
- Min, H.J., Shim, J.W., Han, H.J., Park, C.H. and Jung, H.Y. (2022) 'Fuzzy transform and least-squares fuzzy transform: comparison and application', *International Journal of Fuzzy Systems*, Vol. 24, No. 6, pp.2740–2752.
- Nishat, R.I., Bahoo, Y., Georgiou, K., Hedrick, R. and Urbanic, R.J. (2024) 'Improved collision-free multi-axis tool-path for additive manufacturing', *Computer-Aided Design and Applications*, Vol. 21, No. 6, pp.1103–1115.
- Peng, L.N. (2023) 'The selection of non-negative mode regression variables and its application in stock index tracking', *Journal of Leshan Normal University*, Vol. 38, No. 4, pp.13–19.

- Sirbiladze, G., Khutsishvili, I. and Ghvaberidze, B. (2014) 'Multistage decision-making fuzzy methodology for optimal investments based on experts' evaluations', *European Journal of Operational Research*, Vol. 232, No. 1, pp.169–177.
- Song, W. and Zhu, J.J. (2019) 'A multistage risk decision-making method for normal cloud model considering behavior characteristics', *Applied Soft Computing Journal*, Vol. 78, No. 1, pp.393–406.
- Xie, Z.Q. (2006) 'Construction of a non-negative coefficient linear regression model', *Statistical Education*, Vol. 2006, No. 1, pp.22–26.
- Xu, C.L., Wang, G.Y. and Zhang, Q.H. (2014) 'A new multi-step backward cloud transformation algorithm based on a normal cloud model', *Fundamenta Informaticae*, Vol. 33, No. 1, pp.55–85.
- Xu, Z. (2022) 'Computational intelligence based sustainable computing with classification model for big data visualization on map reduce environment', *Discov. Internet Things*, Vol. 2, No. 1, p.2.
- Yang, J., Wang, G.Y., Liu, Q., Guo, Y.K., Liu, Y., Gan, W.Y. and Liu, Y.C. (2018) 'Review and prospect of research on normal cloud models', *Journal of Computer Science*, Vol. 41, No. 3, pp.724–744.
- Zhang, Z. (2024) 'Improvement of computer adaptive multistage testing algorithm based on adaptive genetic algorithm', *International Journal of Intelligent Information Technologies (IJIT)*, Vol. 20, No. 1, pp.1–19.