


Article

Relative Error Linear Combination Forecasting Model Based on Uncertainty Theory

Hongmei Shi ¹, Lin Wei ¹, Cui Wang ¹, Shuai Wang ^{2,3,4,*} and Yufu Ning ^{2,3,4}

¹ School of Information Science and Engineering, Shandong Agriculture and Engineering University, Jinan 250100, China; z2013454@sdaeu.edu.cn (H.S.); z2013567@sdaeu.edu.cn (L.W.); z2013450@sdaeu.edu.cn (C.W.)

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, Universities of Shandong, Jinan 250103, China

⁴ Smart Healthcare Big Data Engineering and Ubiquitous Computing Characteristic Laboratory, Universities of Shandong, Jinan 250103, China

* Correspondence: wangshuai@sdyu.edu.cn

Abstract: The traditional combination forecasting model has good forecasting effect, but it needs precise historical data. In fact, many random events are uncertain, and much of the data are imprecise; sometimes, historical data are lacking. We need to study combination forecasting problems by means of uncertainty theory. Uncertain least squares estimation is an important technique of uncertain statistics, an important way to deal with imprecise data, and one of the best methods to solve the unknown parameters of uncertain linear regression equations. On the basis of the traditional combination forecasting method and uncertain least squares estimation, this paper proposes two kinds of uncertain combination forecasting models, which are the unary uncertain linear combination forecasting model and the uncertain relative error combination forecasting model, respectively. We set up several piecewise linear regression models according to the data of different periods and, according to certain weights, These piecewise linear regression models are combined into a unary uncertain linear combination forecasting model with a better forecasting effect. The uncertain relative error combination forecasting model is a new forecasting model that combines the traditional relative error linear forecasting model and the uncertain least squares estimation. Compared with the traditional forecasting model, the model can better deal with the forecasting problem of imprecise data. We verify the feasibility of the uncertain combination forecasting model through a numerical example. According to the data analysis, compared with the existing model, the forecasting effect of the proposed model is better.

Keywords: combination forecasting model; relative error; least squares estimation; uncertainty theory; linear regression model



Citation: Shi, H.; Wei, L.; Wang, C.; Wang, S.; Ning, Y. Relative Error Linear Combination Forecasting Model Based on Uncertainty Theory. *Symmetry* **2023**, *15*, 1379. <https://doi.org/10.3390/sym15071379>

Academic Editors: Jian Zhou, Ke Wang, Yuanyuan Liu and Marek T. Malinowski

Received: 7 June 2023

Revised: 29 June 2023

Accepted: 4 July 2023

Published: 7 July 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/>).

1. Introduction

Regression analysis is a forecasting method for data analysis based on the causal relationship of changes in things; that is, according to the actual statistical data, through mathematical calculation, the interdependent quantitative relationship between variables is determined, and a reasonable mathematical model is established to calculate the future value of variables. Linear regression is a statistical analysis method that uses regression analysis in mathematical statistics to determine the interdependent quantitative relationship between two or more variables, which is widely used. Linear regression analysis is mainly used to analyze the observed values and fit a reasonable model. When a new value appears, it can be forecast using this model. The least squares method is a mathematical optimization technique, which is one of the most commonly used methods to solve the

unknown parameters of linear regression. By minimizing the sum of squares of errors, it can find the best-matching data function and obtain a better linear regression fitting equation. The combination forecasting model adopts different single-item forecasting models for the same forecasting object, making full use of the information provided by various single item forecasting methods, and assigning appropriate weighting coefficients to improve the forecasting accuracy. There are many kinds of combination forecasting models, including the linear regression model, exponential model, power function model, logistic model, and neural network. Each model has its own characteristics and application scope. The idea of combining various models to achieve a better forecasting effect is basis of the combination forecasting. Many experts and scholars have conducted in-depth research on the linear combination forecasting model, deduced some forecasting models, achieved good results and carried out practical applications [1–5]. We know that precision and imprecision are symmetrical, precise data are relative, and imprecision data are absolute. Many of the observed data are imprecise. In other words, in practice, the obtained observation is often not a definite value, and may even show an approximate range. At this time, the traditional combination forecasting model cannot solve these problems. However, the uncertainty theory proposed by Liu [6] can solve this problem.

The relation between certainty and uncertainty is symmetrical, and any random event is uncertain. We need to study these problems by means of uncertainty theory. Liu [7] founded the uncertainty theory and gradually improved it [6,8–10]. Uncertainty theory is a branch of mathematics concerned with the analysis of degree of belief. Its main theories include uncertain measure, uncertain variable, uncertain distribution, uncertain inverse operation and expected value. Uncertainty theory has become an important branch of axiomatic mathematics to deal with uncertainty problems in reality. It has been widely used in uncertain planning, uncertain statistics, comprehensive evaluation and production planning [11–13], and has achieved fruitful results, which has aroused great attention. In 2010, Liu [6] began his research on uncertainty statistics, which is a methodology for collecting and interpreting expert experience data through uncertainty theory. Uncertainty statistics mainly include uncertain regression equation, uncertain estimation and uncertain hypothesis testing. Based on the keen interest in uncertain regression equations, many uncertain regression models have been proposed by experts and scholars [14–18]. Yao and Liu [19] proposed the least squares estimation to solve the unknown parameters of the uncertain regression equation. Wang et al. [20–22] proposed two new uncertain linear regression models. Shi et al. [23] proposed total least squares estimation model based on uncertainty theory. Uncertainty statistics also have real applications; when COVID-19 was spreading rapidly in most countries around the world, Liu Z. [24] proposed an uncertain growth model for the cumulative number of COVID-19 infections in China.

It is not easy to build a scientific forecasting model, because whether the forecasting model is scientific depends on the accuracy of the forecasting results on the one hand, and on the simplicity of the model itself on the other. However, these two aspects are contradictory: when the model is simple, the forecasting results are often not too accurate; when the forecasting is relatively accurate, the model is not too simple. On the basis of the previous research [3–5] and uncertainty theory, this paper puts forward two kinds of uncertain combination forecasting models, which are the unary uncertain linear combination forecasting model and the uncertain relative error combination forecasting model. In general, the newer the data information, the greater the impact of the given data on the model, but the historical data are also a factor affecting the accuracy of the model. According to the principle of minimum error, the unary uncertain linear combination forecasting model combines the piecewise linear regression of the data corresponding to different periods into a prediction model with higher accuracy. The uncertain relative error combination forecasting model is based on the least squares principle and relative error, combined with the uncertainty theory, which can better deal with the regression and forecasting of imprecision data. The two kinds of uncertain linear combination forecasting

models can be used for both imprecise data and precise data, and the forecasting effect of the models is very good.

In this paper, we propose the unary uncertain linear combination forecasting model and the uncertain relative error combination forecasting model. Both of these models can solve the regression equation of imprecise observation data better, and the forecasting effect is better. The main organizational structure is as follows: In Section 2, we propose the unary uncertain linear combination forecasting model. This model aims to establish several piecewise linear regression models according to the data of different periods, and combine the piecewise linear regression into an uncertain combination forecasting model. In Section 3, we propose the uncertain relative error combination forecasting model. This model is a new model, which combines relative error and uncertainty theory together and has a good forecasting effect. In the Section 4, the feasibility of the uncertain linear regression combination forecasting model is verified by numerical example. The forecasting effect of the model is good. Finally, we summarize the proposed model and point out the future research direction.

2. Uncertain Regression Model

Certainty and uncertainty are symmetrical, and precision and imprecision are also symmetrical. In order to resolve uncertainty problems such as imprecise data, Liu [7] founded uncertainty theory. The main content of uncertainty theory includes the basic theory of uncertainty variable, uncertainty measure and uncertainty distribution, as well as the calculation methods of uncertainty operational laws and expected value. If you are interested in uncertainty theory and uncertainty statistics, please study Reference [10]. In this section, we mainly introduce the uncertain least squares estimation method of the uncertain regression equation.

Assume that (x_1, x_2, \dots, x_n) is an independent variables vector, and y is a dependent variable. If the functional relationship is between (x_1, x_2, \dots, x_n) , then y can be expressed by a regression model

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

where β is an unknown vector of parameters, ε is a disturbance term and ε is an uncertain variable. If the regression equation fits well, its expected value $E[\varepsilon]$ should be 0 [10].

In particular, Liu [10] call

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

a linear regression model.

Assume that we have a set of imprecisely observed data,

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

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.

Yao-Liu [19] proposed the least squares estimate of unknown parameter β of linear regression model. The parameter β is the solution of the following minimization problem

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

If the minimization solution is β^* , then the fitted regression equation is determined by $y = f(x_1, x_2, \dots, x_n | \beta^*)$. Then, 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 is uncertain variable, its 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 [25].

Let (x_1, x_2, \dots, x_n) be a new independent variables vector; the forecast uncertain variable of dependent variable \hat{y} is

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

Lio-Liu [25] suggested that the forecast value is defined as the expected value of the uncertain variable \hat{y} , i.e.,

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

Taking α (e.g., 95%) as the confidence level, the confidence interval of dependent variable \hat{y} is

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

3. Unary Uncertain Linear Regression Combination Forecasting Model

In this section, we derive the unary uncertain linear combination forecasting model, which is abbreviated as UULCFM. We all know that timely updated data have a greater effect on the forecasting model, so when establishing the forecasting model scientifically, we should fully consider the changes in time and conditions. The data information of different periods has different influences on the model and recent data information is obviously more valuable than long-term data information. The idea of UULCFM is to establish m regression models by discarding a certain amount of previous historical data according to the existing data, and then assemble m regression models into a forecasting model according to the principle of minimum error.

Assume that $(\tilde{x}_i, \tilde{y}_i)$ ($i = 1, 2, \dots, n$) be a set of imprecise data, where \tilde{x}_i, \tilde{y}_i are independent uncertain variables with regular uncertainty distributions Φ_i, Ψ_i ($i = 1, 2, \dots, n$), respectively. We always assumed that there is a linear relationship between \tilde{x}_i and \tilde{y}_i ($i = 1, 2, \dots, n$), and y can be expressed by an uncertain regression model $y = \alpha + \beta x + \varepsilon$, where α, β are unknown parameters, and ε is an uncertain disturbance term.

The main steps of the UULCFM are as follows.

Step 1. For the original n sets of data, we obtained the following unary linear regression model using uncertain least squares estimation.

$$y_1 = \alpha_1 + \beta_1 x, \quad (11)$$

where α_1 and β_1 are unknown parameters.

Step 2. Discarding the first N_1 sets of data, we can obtain the following unary linear regression model for the remaining $n - N_1$ sets of data through least squares estimation.

$$y_2 = \alpha_2 + \beta_2 x. \quad (12)$$

where α_2 and β_2 are unknown parameters, N_1 is positive integer and $N_1 < n$.

Step 3. Discarding the first N_2 sets of data, we can obtain the following unary linear regression model for the remaining $n - N_2$ sets of data through least squares estimation.

$$y_3 = \alpha_3 + \beta_3 x. \quad (13)$$

where α_3 and β_3 are unknown parameters. Both N_1, N_2 are positive integers, and $N_1 < N_2 < n$.

By analogy, we can obtain the m th unary linear regression equation.

Step m . Discarding the first N_{m-1} sets of data, we can obtain the following unary linear regression model for the remaining $n - N_{m-1}$ sets of data through least squares estimation.

$$y_m = \alpha_m + \beta_m x. \tag{14}$$

where α_m and β_m are unknown parameters. Both N_1, N_2, \dots, N_{m-1} are positive integers, and $N_1 < N_2 < \dots < N_{m-1} < n$.

In this way, m unary linear regression models are obtained as follows,

$$y_i = \alpha_i + \beta_i x, i = 1, 2, 3, \dots, m. \tag{15}$$

where α_i, β_i are unknown parameters.

Each regression equation of Equation (5) is fitted to the remaining $n - N_{m-1}$ sets of data, and the generated errors are, respectively,

$$\begin{aligned} \varepsilon_{ij} &= y_{ij} - \tilde{y}_j, \\ i &= 1, 2, 3, \dots, m, j = N_m, N_{m+1}, \dots, n. \end{aligned} \tag{16}$$

Since $\tilde{y}_i, i = 1, 2, \dots, n$ is a set of imprecise data, Equation (6) is deformed into the following form according to the uncertain expected value formula [10].

$$\varepsilon_{ij} = y_{ij} - E[\tilde{y}_j] = y_{ij} - \int_0^1 \Psi_j^{-1}(\alpha) d\alpha, \quad i = 1, 2, 3, \dots, m, \quad j = N_m, N_{m+1}, \dots, n. \tag{17}$$

The purpose of this paper is to find m numbers $k_1, k_2, k_3, \dots, k_m$ that satisfy $k_1 + k_2 + k_3 + \dots + k_m = 1$. Then, we construct the composite model

$$y = \sum_{i=1}^m k_i y_i. \tag{18}$$

This minimizes the sum of the squares $R = \sum_{j=N_m}^n \varepsilon_j$ of error $\varepsilon_{ij} = y_{ij} - E[\tilde{y}_j]$, $i = 1, 2, 3, \dots, m, j = N_m, N_{m+1}, \dots, n$. This model is called the linear regression combinatorial model.

We know from the above derivation

$$\begin{aligned} \varepsilon_j &= y_j - E[\tilde{y}_j] = \sum_{i=1}^m k_i y_{ij} - \sum_{i=1}^m k_i E[\tilde{y}_j] = \sum_{i=1}^m k_i [y_{ij} - E[\tilde{y}_j]] = \sum_{i=1}^m k_i \varepsilon_{ij} \\ &= [\varepsilon_{1j}, \varepsilon_{2j}, \dots, \varepsilon_{mj}] [k_1, k_2, k_3, \dots, k_m]^T = [k_1, k_2, k_3, \dots, k_m] [\varepsilon_{1j}, \varepsilon_{2j}, \dots, \varepsilon_{mj}]^T. \end{aligned} \tag{19}$$

So,

$$\begin{aligned} \varepsilon_j^2 &= [k_1, k_2, k_3, \dots, k_m] [\varepsilon_{1j}, \varepsilon_{2j}, \dots, \varepsilon_{mj}]^T [\varepsilon_{1j}, \varepsilon_{2j}, \dots, \varepsilon_{mj}] [k_1, k_2, k_3, \dots, k_m]^T \\ &= [k_1, k_2, k_3, \dots, k_m] \begin{bmatrix} \varepsilon_{1j}^2 & \varepsilon_{1j}\varepsilon_{2j} & \dots & \varepsilon_{1j}\varepsilon_{mj} \\ \varepsilon_{1j}\varepsilon_{2j} & \varepsilon_{2j}^2 & \dots & \varepsilon_{2j}\varepsilon_{mj} \\ \dots & \dots & \dots & \dots \\ \varepsilon_{1j}\varepsilon_{mj} & \varepsilon_{2j}\varepsilon_{mj} & \dots & \varepsilon_{mj}^2 \end{bmatrix} [k_1, k_2, k_3, \dots, k_m]^T. \end{aligned} \tag{20}$$

Therefore, we obtain

$$R = \sum_{i=N_m}^n \varepsilon_j^2 = [k_1, k_2, k_3, \dots, k_m]$$

$$\begin{bmatrix} \sum_{i=N_m}^n \varepsilon_{1j}^2 & \sum_{i=N_m}^n \varepsilon_{1j}\varepsilon_{2j} & \cdots & \sum_{i=N_m}^n \varepsilon_{1j}\varepsilon_{mj} \\ \sum_{i=N_m}^n \varepsilon_{1j}\varepsilon_{2j} & \sum_{i=N_m}^n \varepsilon_{2j}^2 & \cdots & \sum_{i=N_m}^n \varepsilon_{2j}\varepsilon_{mj} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=N_m}^n \varepsilon_{1j}\varepsilon_{mj} & \sum_{i=N_m}^n \varepsilon_{2j}\varepsilon_{mj} & \cdots & \sum_{i=N_m}^n \varepsilon_{mj}^2 \end{bmatrix} \quad (21)$$

$$[k_1, k_2, k_3, \dots, k_m]^T.$$

Denoted as

$$K = [k_1, k_2, k_3, \dots, k_m],$$

$$E = \begin{bmatrix} \sum_{i=N_m}^n \varepsilon_{1j}^2 & \sum_{i=N_m}^n \varepsilon_{1j}\varepsilon_{2j} & \cdots & \sum_{i=N_m}^n \varepsilon_{1j}\varepsilon_{mj} \\ \sum_{i=N_m}^n \varepsilon_{1j}\varepsilon_{2j} & \sum_{i=N_m}^n \varepsilon_{2j}^2 & \cdots & \sum_{i=N_m}^n \varepsilon_{2j}\varepsilon_{mj} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=N_m}^n \varepsilon_{1j}\varepsilon_{mj} & \sum_{i=N_m}^n \varepsilon_{2j}\varepsilon_{mj} & \cdots & \sum_{i=N_m}^n \varepsilon_{mj}^2 \end{bmatrix}. \quad (22)$$

There are

$$R = K \cdot E \cdot K^T. \quad (23)$$

Assuming that

$$U = [1, 1, \dots, 1]^T. \quad (24)$$

So, the linear regression combination model becomes the problem of finding the minimum value of $R = K \cdot E \cdot K^T$ under the constraint condition $K \cdot U = 1$.

We use the Lagrange multiplier method to solve the conditional extremum. We construct the Lagrange function as follows

$$L = K \cdot E \cdot K^T + \lambda(KU - 1). \quad (25)$$

The Lagrange function L is a binary elementary function and the minimum point is the stagnation point of the function. If we take the first partial derivative of L w.r.t K , then we obtain

$$\frac{\partial L}{\partial K} = 2E \cdot K^T + \lambda U = 0. \quad (26)$$

If we solve the Equation (26), we get

$$K^T = -\frac{1}{2}\lambda E^{-1}U. \quad (27)$$

According to constraint $K \cdot U = 1$, we can solve Equation (27) and obtain

$$K^T = \frac{E^{-1}U}{U^T E^{-1}U}, \lambda = \frac{-2}{U^T E^{-1}U}. \quad (28)$$

For m numbers m , numbers $k_1, k_2, k_3, \dots, k_m$ satisfy $k_1 + k_2 + k_3 + \dots + k_m = 1$. Thus, the linear regression combination model $y = \sum_{i=1}^m k_i y_i$ was obtained.

The derivation process of the UULCFM involves the matrix inverse and matrix elementary transformation, which requires readers to have a certain matrix foundation and linear algebra foundation.

4. Uncertain Relative Error Linear Combination Forecasting Model

In this section, we derive the uncertain relative error linear combination forecasting model, which is abbreviated as UURELCFM. Suppose that we have a set of imprecise data $X = (x_1, x_2, \dots, x_n)^T$, where $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n, \tilde{y}$ are independent uncertain variables with regular uncertainty distributions $\Phi_1, \Phi_2, \dots, \Phi_n, \Psi$, respectively.

The basic principles of UURELCFM are as follows. The forecasting result of the i th ($i = 1, 2, \dots, m$) forecasting method is $X_i = (x_{1i}, x_{2i}, \dots, x_{ni})^T$. The linear combination of the m forecasting result is

$$Y = (y_1, y_2, \dots, y_n)^T = \omega_1 X_1 + \omega_2 X_2 + \dots + \omega_m X_m. \quad (29)$$

The relative error between the forecasting value and the original data can be defined as

$$E = (e_1, e_2, \dots, e_n)^T. \quad (30)$$

Among them,

$$e_j = \frac{|y_j - \tilde{x}_j|}{|\tilde{x}_j|} = \frac{|\sum_{i=1}^m \omega_i x_{ji} - \tilde{x}_j|}{|\tilde{x}_j|} = \left| \sum_{i=1}^m \omega_i \frac{x_{ji}}{\tilde{x}_j} - 1 \right|, \quad j = 1, 2, \dots, n. \quad (31)$$

Since $(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n, \tilde{y})^T$ leads to imprecise data, we have to solve Equation (31) by means of uncertain expectations [10].

$$e_j = \left| \sum_{i=1}^m \omega_i E\left[\frac{x_{ji}}{\tilde{x}_j}\right] - 1 \right| = \left| \sum_{i=1}^m \omega_i x_{ji} \int_0^1 \frac{1}{\Phi_j^{-1}(1-\alpha)} d\alpha - 1 \right|, \quad j = 1, 2, \dots, n. \quad (32)$$

The uncertain relative error linear combination forecasting model I (URELCFM I) with the minimum sum of squares of relative errors is

$$\min Q = \sum_{j=1}^n e_j^2, \quad (33)$$

and the constraint of the model is

$$\sum_{i=1}^m \omega_i = 1. \quad (34)$$

Denoted as

$$\begin{aligned} Y_i &= (y_{1i}, y_{2i}, \dots, y_{ni})^T = (E\left[\frac{x_{1i}}{\tilde{x}_1}\right] - 1, E\left[\frac{x_{2i}}{\tilde{x}_2}\right] - 1, \dots, E\left[\frac{x_{ni}}{\tilde{x}_n}\right] - 1)^T \\ &= (x_{1i} \int_0^1 \frac{1}{\Phi_1^{-1}(1-\alpha)} d\alpha - 1, x_{2i} \int_0^1 \frac{1}{\Phi_2^{-1}(1-\alpha)} d\alpha - 1, \\ &\dots, x_{ni} \int_0^1 \frac{1}{\Phi_n^{-1}(1-\alpha)} d\alpha - 1)^T, \end{aligned} \quad (35)$$

and

$$R = [1, 1, \dots, 1]^T, W = [\omega_1, \omega_2, \dots, \omega_n]^T,$$

$$Y = \begin{bmatrix} Y_1^T Y_1 & Y_1^T Y_2 & Y_1^T Y_3 & \dots & Y_1^T Y_m \\ Y_2^T Y_1 & Y_2^T Y_2 & Y_2^T Y_3 & \dots & Y_2^T Y_m \\ Y_3^T Y_1 & Y_3^T Y_2 & Y_3^T Y_3 & \dots & Y_3^T Y_m \\ \dots & \dots & \dots & \dots & \dots \\ Y_m^T Y_1 & Y_m^T Y_2 & Y_m^T Y_3 & \dots & Y_m^T Y_m \end{bmatrix}. \quad (36)$$

The sum of squares of relative errors is

$$Q = \sum_{j=1}^n e_j^2 = \sum_{j=1}^n \left(\sum_{i=1}^m \omega_i E \left[\frac{x_{ji}}{\bar{x}_j} \right] - 1 \right)^2 = \sum_{j=1}^n \left(\sum_{i=1}^m \omega_i E \left[\frac{x_{ji}}{\bar{x}_j} \right] - \sum_{i=1}^m \omega_i \right)^2$$

$$= \sum_{j=1}^n \left[\sum_{i=1}^m \omega_i \left(E \left[\frac{x_{ji}}{\bar{x}_j} \right] - 1 \right) \right]^2 = \sum_{j=1}^n \left[\sum_{i=1}^m \omega_i y_{ji} \right]^2 = \sum_{j=1}^n [W^T (y_{j1}, y_{j2}, \dots, y_{jm})^T]^2 \quad (37)$$

$$= \sum_{j=1}^n W^T (y_{j1}, y_{j2}, \dots, y_{jm})^T (y_{j1}, y_{j2}, \dots, y_{jm}) W = W^T Y W.$$

Equation (33) is transformed into

$$\min Q = W^T Y W, \quad (38)$$

and the constraint is transformed into

$$R^T W = 1. \quad (39)$$

According to the Lagrange multiplier method, the optimal coefficient W^* is

$$W^* = \frac{Y^{-1} R}{R^T Y^{-1} R}. \quad (40)$$

The solution of model I sometimes has a negative component, which does not achieve the expected effect of the linear combination forecasting model. In order to overcome the limitations of URELCFM I, we put forward an uncertain combination forecasting model with the minimum sum of squares of relative errors of non-negative weights, namely, the uncertain relative error linear combination forecasting model II (URELCFM II)

$$\min Q = \sum_{j=1}^n e_j^2, \quad (41)$$

and the constraint of the model is

$$\sum_{i=1}^m \omega_i = 1, \quad \omega_i \geq 0, \quad i = 1, 2, \dots, m. \quad (42)$$

According to the derivation of URELCFM I, we can obtain

$$\min Q = W^T Y W, \quad (43)$$

and the constraint is transformed into

$$R^T W = 1, W \geq 0. \quad (44)$$

URELCFM II belongs to quadratic convex programming and can be solved by the simplex algorithm of quadratic convex programming. This method needs to be solved by the linear programming method of finite number or can also be solved by MATLAB optimization toolbox.

Both URELCFM I and URELCFM II require that the sum of the weighting coefficients is 1. In fact, there is no need for this limitation. The weight can also be negative, and the goal is to minimize the sum of squares of the combined forecasting errors. Although it is controversial that the weight is negative, it is also common from a mathematical perspective; for example, multiple regression often has negative coefficients. By removing the limitation of the weighting coefficient, we can obtain the uncertain relative error linear combination forecasting model III (URELCFM III) with the minimum sum of squares of relative errors

$$\min Q = \sum_{j=1}^n e_j^2. \tag{45}$$

Define

$$\begin{aligned} Z_i &= (z_{1i}, z_{2i}, \dots, z_{ni})^T = (E[\frac{x_{1i}}{\tilde{x}_1}], E[\frac{x_{2i}}{\tilde{x}_2}], \dots, E[\frac{x_{ni}}{\tilde{x}_n}])^T \\ &= (x_{1i} \int_0^1 \frac{1}{\Phi_1^{-1}(1-\alpha)} d\alpha, x_{2i} \int_0^1 \frac{1}{\Phi_2^{-1}(1-\alpha)} d\alpha, \\ &\dots, x_{ni} \int_0^1 \frac{1}{\Phi_n^{-1}(1-\alpha)} d\alpha)^T, i = 1, 2, \dots, m. \end{aligned} \tag{46}$$

The sum of squares of relative errors is

$$Q = \sum_{j=1}^n e_j^2 = \sum_{j=1}^n (\sum_{i=1}^m \omega_i E[\frac{x_{ji}}{\tilde{x}_j}] - 1)^2 = \sum_{j=1}^n (\sum_{i=1}^m \omega_i z_{ji} - 1)^2. \tag{47}$$

Q in Equation (47) is an elementary function. In order to solve the minimum value of Q, we take the partial derivative of $\omega_i (i = 1, 2, \dots, m$ and equal to 0, respectively, to obtain the following equations

$$\begin{aligned} \omega_1 \sum_{i=1}^n Z_{i1}^2 + \omega_2 \sum_{i=1}^n Z_{i1}Z_{i2} + \omega_3 \sum_{i=1}^n Z_{i1}Z_{i3} + \dots + \omega_m \sum_{i=1}^n Z_{i1}Z_{im} &= \sum_{i=1}^n Z_{i1}, \\ \omega_1 \sum_{i=1}^n Z_{i2}Z_{i1} + \omega_2 \sum_{i=1}^n Z_{i2}^2 + \omega_3 \sum_{i=1}^n Z_{i2}Z_{i3} + \dots + \omega_m \sum_{i=1}^n Z_{i2}Z_{im} &= \sum_{i=1}^n Z_{i2}, \\ \omega_1 \sum_{i=1}^n Z_{i3}Z_{i1} + \omega_2 \sum_{i=1}^n Z_{i3}Z_{i2} + \omega_3 \sum_{i=1}^n Z_{i3}^2 + \dots + \omega_m \sum_{i=1}^n Z_{i3}Z_{im} &= \sum_{i=1}^n Z_{i3}, \\ \dots & \\ \omega_1 \sum_{i=1}^n Z_{im}Z_{i1} + \omega_2 \sum_{i=1}^n Z_{im}Z_{i2} + \omega_3 \sum_{i=1}^n Z_{im}Z_{i3} + \dots + \omega_m \sum_{i=1}^n Z_{im}^2 &= \sum_{i=1}^n Z_{im}. \end{aligned} \tag{48}$$

Denoted as

$$\begin{aligned} Z &= \begin{bmatrix} Z_1^T Z_1 & Z_1^T Z_2 & Z_1^T Z_3 & \dots & Z_1^T Z_m \\ Z_2^T Z_1 & Z_2^T Z_2 & Z_2^T Z_3 & \dots & Z_2^T Z_m \\ Z_3^T Z_1 & Z_3^T Z_2 & Z_3^T Z_3 & \dots & Z_3^T Z_m \\ \dots & \dots & \dots & \dots & \dots \\ Z_m^T Z_1 & Z_m^T Z_2 & Z_m^T Z_3 & \dots & Z_m^T Z_m \end{bmatrix}, \\ M &= \begin{bmatrix} Z_1^T R \\ Z_2^T R \\ Z_3^T R \\ \dots \\ Z_m^T R \end{bmatrix}. \end{aligned} \tag{49}$$

We express the above equations by matrix equations, and obtain

$$ZW = M. \tag{50}$$

The matrix Z is invertible and the solution W is

$$W = Z^{-1}M. \quad (51)$$

URELCFM I has more constraints than URELCFM III, and the accuracy of URELCFM I is lower than that of URELCFM III, while URELCFM II has more constraints than URELCFM I. Therefore, URELCFM III has the highest accuracy, that is, the sum of relative error squares $Q_{III} \leq Q_I \leq Q_{II}$.

5. Numerical Example

To verify the feasibility and effectiveness of the model proposed in this paper, we provide a numerical example of imprecise data. Moreover, we followed the numerical analysis method for the disturbance term in Reference [25], calculated the expected values and variance of the disturbance term, and forecasting and solved the confidence interval. The numerical analysis results show that the model proposed in this paper can lead to better forecasting data.

Assuming that $(\tilde{x}_i, \tilde{y}_i)$, $i = 1, 2, \dots, 8$ are imprecise data provided in Table 1, where \tilde{x}_i, \tilde{y}_i , $i = 1, 2, \dots, 8$ are independent linear uncertain variables with regular uncertainty distributions Φ_i and Ψ_i , $i = 1, 2, \dots, 8$, respectively.

Table 1. Imprecise data (Linear uncertainty distribution).

i	1	2	3	4
\tilde{x}_i	$\mathcal{L}(1, 3)$	$\mathcal{L}(3, 5)$	$\mathcal{L}(5, 7)$	$\mathcal{L}(7, 9)$
\tilde{y}_i	$\mathcal{L}(4, 6)$	$\mathcal{L}(5, 6)$	$\mathcal{L}(7, 9)$	$\mathcal{L}(10, 12)$
i	5	6	7	8
\tilde{x}_i	$\mathcal{L}(9, 11)$	$\mathcal{L}(11, 13)$	$\mathcal{L}(13, 15)$	$\mathcal{L}(15, 17)$
\tilde{y}_i	$\mathcal{L}(12, 14)$	$\mathcal{L}(15, 16)$	$\mathcal{L}(20, 22)$	$\mathcal{L}(18, 20)$

We carried out linear regression using the uncertain uncertain slope mean method (USMM) [20] and uncertain equation deformation method (UEDM) [21] respectively, and then solved the linear regression equations according to the combination forecasting model proposed in this paper. The results are shown in Table 2.

Table 2. The linear regression equations.

Model	Linear Regression Equations
UEDM	$y = 1.6258 + 1.1805x$
USMM	$y = 1.5406 + 1.1901x$
UULCFM	$y = 1.5336 + 1.1906x$
URELCFM I	$y = 1.5346 + 1.1906x$
URELCFM II	$y = 1.5321 + 1.1907x$
URELCFM III	$y = 1.5355 + 1.1905x$

It can be seen from Table 2 that there are some differences in the coefficients of the linear regression equation of USMM and UEDM. The coefficients of the linear regression equation of the model proposed in this paper are almost the same, the stability of the model is strong, and the difference in fitting effect is small, which can be ignored.

The estimated expected values and estimated variances of the each model disturbance term are shown in Table 3.

As can be seen from Table 3, the estimated expected values of the disturbance terms of the URELCFM I, URELCFM II and URELCFM III are all 0.0000, and the variance is relatively small, indicating that the three models have a better fitting effect and better forecast effect, and URELCFM III has the best performance.

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

Model	Expected Values	Variances
UEDM	−0.2124	12.3274
USMM	−0.0120	5.6523
UULCFM	0.0080	4.4235
URELCFM I	0.0000	2.2146
URELCFM II	0.0000	2.4678
URELCFM III	0.0000	1.6131

We forecast the data according to the URELFCM III and obtained the confidence interval. We assumed that $\tilde{x} \sim \mathcal{L}(17, 19)$ is a new imprecise form of data, and we take the confidence level to be $\alpha = 95\%$. According to References [25], the forecast value and confidence interval were obtained as shown in Table 4.

Table 4. The forecast value and confidence interval.

Model	Forecast Value	Confidence Interval
URELCFM III	22.9645	22.9645 ± 2.5666

From the perspective of numerical examples, all four models proposed in this paper are feasible. From the perspective of data analysis and comparison with existing models, the prediction effect of the four models proposed in this paper is better.

6. Conclusions

Traditional forecasting models all require data to be precise. In fact, statistics can be imprecise. For example, after the college entrance examination, we invited a teacher to estimate the score of a certain candidate. If the teacher believes that the candidate's score is bound to exceed 500, we would obtain an expert's experience data (500, 0), if the teacher thinks the candidate's score is less than 520 is 0.3, we obtain an expert's experience data (520, 0.3), if the teacher thinks the candidate's score is less than 550 is 0.6, we obtain an expert's experience data (550, 0.6), if the teacher thinks the candidate's score is less than 580 is 0.8, we obtain an expert experience data (580, 0.8), and the teacher believes that the candidate will score no higher than 600, we obtain an expert experience data (600, 1). This gives us five pieces of expert experience data (500, 0), (520, 0.3), (550, 0.6), (580, 0.8), (600, 1), all of which are imprecise data.

Based on traditional combination forecasting methods and uncertainty theory, this paper proposes two kinds of uncertain combination forecasting models. The forecasting models proposed in this paper are all aimed at imprecise data, and they rely on uncertainty theory when solving. Univariate uncertain linear combination forecasting model is a relatively basic linear model. It establishes several piecewise linear regression models based on data in different periods and combines them into an uncertain combination forecasting model with high accuracy. The uncertain relative error combination forecasting model is based on the principle of minimizing the sum of squares of relative errors, setting weight restrictions, and obtaining three kinds of uncertain relative error combination forecasting models with good forecasting results. The four models proposed in this paper are all feasible, and the forecasting effect of the models proposed in this paper is better than the existing models obtained through data analysis.

The numerical example in this paper is a univariate linear forecasting problem, and the model solution and data analysis are not too complicated. The derivation and calculation of multivariable uncertain linear combination forecasting model are relatively complex, and can only be realized with the help of computer programs or MATLAB programming.

Author Contributions: Conceptualization, Y.N.; methodology, S.W.; validation, C.W.; data curation, L.W.; 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.

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

References

1. Yao, C.; Wang, J. A comparative study on forecast models of apple output in China. *J. Fruit Sci.* **2007**, *24*, 682–684.
2. Meng, Y.; Wang, J.; Kang, G.; Chen, C. Analysis of apple production status in China. *China Fruit Trees* **2007**, *1*, 43–44.
3. Hou, L.; Sun, C. A neural network forecast model for apple yield. *J. China Agric. Univ. Soc. Sci. Ed.* **2001**, *42*, 51–53.
4. Gao, S.; Zhang, S.; Mei, L. Linear combination forecast based on relative error criterion. *J. Syst. Eng. Electron.* **2008**, *30*, 481–484.
5. Hao, L.; Wu, D. Linear regression combined forecasting model—Take the prediction of China’s aging population as an example. *J. Shenyang Univ. Soc. Sci.* **2016**, *18*, 290–293.
6. Liu, B. *Uncertainty Theory: A Branch of Mathematics for Modeling Human Uncertainty*; Springer: Berlin/Heidelberg, Germany, 2010.
7. Liu, B. *Uncertainty Theory*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2007.
8. Liu, B. Some research problems in uncertainty theory. *J. Uncertain Syst.* **2009**, *3*, 3–10.
9. Liu, B. Why is there a need for uncertainty theory? *J. Uncertain Syst.* **2012**, *6*, 3–10.
10. Liu, B. *Uncertainty Theory*, 5th ed.; Springer: Berlin/Heidelberg, Germany, 2021.
11. Wang, X.; Ning, Y. An Uncertain currency model with floating interest rates. *Soft Comput.* **2017**, *21*, 6739–6754. [[CrossRef](#)]
12. Ning, Y.; Pang, N.; Wang, X. An Uncertain aggregate production planning model considering investment in vegetable preservation technology. *Math. Probl. Eng.* **2019**. [[CrossRef](#)]
13. Guo, H.; Wang, X.; Gao, Z. Uncertain linear regression model and its application. *J. Intell. Manuf.* **2014**, *28*, 559–564. [[CrossRef](#)]
14. Wang, X.; Peng, Z. Method of moments for estimating uncertainty distributions. *J. Uncertain. Anal. Appl.* **2014**, *2*, 5. [[CrossRef](#)]
15. Chen, D. Tukey’s biweight estimation for uncertain regression model with imprecise observations. *Soft Comput.* **2020**, *24*, 16803–16809. [[CrossRef](#)]
16. Song, Y.; Fu, Z. Uncertain multivariable regression model. *Soft Comput.* **2018**, *22*, 5861–5866. [[CrossRef](#)]
17. Wang, X.; Li, H.; Guo, H. A new Uncertain regression model and its application. *Soft Comput.* **2020**, *24*, 6297–6305. [[CrossRef](#)]
18. Liu, Z. Least absolute deviations estimation for uncertain regression with imprecise observations. *Fuzzy Optim. Decis. Mak.* **2020**, *19*, 33–52. [[CrossRef](#)]
19. Yao, K.; Liu, B. Uncertain regression analysis: An approach for imprecise observations. *Soft Comput.* **2018**, *22*, 5579–5582. [[CrossRef](#)]
20. 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](#)]
21. Wang, S.; Ning, Y.; Shi, H. A new uncertain linear regression model based on equation deformation. *Soft Comput.* **2021**, *25*, 12817–12824. [[CrossRef](#)]
22. Wang, S.; Ning, Y.; Huang, H. Uncertain least squares estimation model based on relative error. *J. Intell. Fuzzy Syst.* **2023**, *44*, 8281–8290. [[CrossRef](#)]
23. Shi, H.; 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](#)]
24. Liu, Z. Uncertain growth model for the cumulative number of COVID-19 infections in China. *Fuzzy Optim. Decis. Mak.* **2021**, *20*, 229–242. [[CrossRef](#)]
25. Lio, W.; Liu, B. Residual and confidence interval for uncertain regression model with imprecise observations. *J. Intell. Fuzzy Syst.* **2018**, *35*, 2573–2583. [[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.