ORIGINAL RESEARCH

Multivariate uncertain regression model with imprecise observations

Tingqing Ye[1] · **Yuhan Liu[2]**

Abstract

The multivariate regression model is a mathematical tool for estimating the relationships among some explanatory variables and some response variables. In some cases, observed data are imprecise. In order to model those imprecise data, we can employ uncertainty theory to design the uncertain regression model by regarding those data as uncertain variables. Parameters estimation is an important topic in the uncertain regression model. In this paper, we explore a method of parameters estimation by the principle of least squares in the multivariate uncertain regression model containing more than one response variables and assuming both explanatory variables and response variables as uncertain variables. Besides, when the new explanatory variables are given, we propose an approach to obtain the forecast value and the confidence interval of the response variables. At last, a numerical example of the multivariate uncertain regression model is showed.

Keywords Multivariate uncertain regression · Uncertainty theory · Parameter estimation · Residual · Confidence interval

1 Introduction

In order to understand the relationships among lots of factors, people need to impose structure on those factors. Usually, we build a regression model to describe how the changes in some variables (explanatory variables) affect other variables (response variables). If the regression model contains only one response variable, we call the model multiple regression model. Furthermore, if we want to study the relationships among explanatory variables and more than one response variables, maybe we have two methods. One is that we can establish a multiple regression model for each response variable and all explanatory variables, and consider those models independently. The other is that we can design a multivariate regression model including all response variables and explanatory variables, and take the relationships among response variables into consider. Perhaps the latter one is better. Because we often meet with cases that the

correlation of the response variables is high. For example, we want to study the relationships among systolic blood pressure, diastolic blood pressure of a patient (response variables) and his gender, body temperature, heart rate (explanatory variables). Since systolic blood pressure is highly correlated with diastolic blood pressure, it is inconclusive to separate them. Thus, multivariate regression model is more reasonable. In fact, the very reason why we employ a multivariate model is to incorporate the relationships of response variables.

In statistical domain, the relationship among each response variable and explanatory variables is expressed by a function, thus called functional relationship. For example, in the multivariate linear regression model, all functions are linear. Generally speaking, in the multivariate regression model, the functional relationships should be determined in advance through people's experience although there are unknown parameters in those functions. An experienced simple model not only is easier to remember but also can inspire new idea. The process of modeling is the process of understanding the world. In statistics, the most widely used model is linear model. In Galton ([1886](#)), firstly proposed "regression" for a simple linear regression model to study the relationship between children's height and parents' height. Twelve years later, Yule ([1897](#)) introduced the regression into statistical domain. In Fama et al. ([1969](#)), used event study methodology in the multivariate regression

✉ Yuhan Liu
 liu3yh@mail.uc.edu

 Tingqing Ye
 yetq18@mails.tsinghua.edu.cn

1 Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China

2 Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH 45221-00025, USA

model to study the effect of new information on asset prices. In recent years, Ganesh (2018) presented individual regression method to predict the $PM_{2.5}$ concentration. Krishnamurthy (2019) using support vector regression to calculate the Lyapunov exponents of short time series.

In the multivariate regression model, it is vital for us to estimate unknown parameters based on given observations. The multivariate least squares estimation is the most widely used estimation method, generalized by Aitken (1935) and developed by Watson (1967). Besides, the multivariate least absolute estimation (Gentle 1977; Bilodeau and Brenner 1999), maximum likelihood estimation (Anderson 1951) and least distance estimation (Bai et al. 1990) are other common methods of point estimation. However, those methods do not consider the relationships among the response variables. In order to take the relationships into consider, Breiman and Friedman (1997) proposed the restrained multivariate least squares estimation by canonical analysis, and Jhun and Choi (2009) presented the bootstrapping least distance estimation in the multivariate regression model.

Note that explanatory variables and the response variables in traditional regression model are assumed to be observed precisely. However, the observations are unable to be precise in some cases. For example, the data of the factories' carbon emission or the social benefit of factories during some time are collected in an imprecise way. How do we model those imprecise data? Liu (2012) suggested to employ uncertainty theory to model the imprecisely observed data given by the domain experts. Uncertainty theory was founded by Liu (2007) and developed by Liu (2009) based on normality, duality, subadditivity, and product axioms in order to deal with the belief degree with human uncertainty. The regression model based on uncertainty theory is called uncertain regression model. And in uncertain regression model, those imprecisely observed data are regarded as uncertain variables. It is an important topic to estimate the unknown parameters in the uncertain regression model. On the one hand, for the uncertain multiple regression model including only one response variable, many scholars have proposed lots of methods such as the least squares estimation (Yao and Liu 2018), the least absolute deviations estimation (Liu and Yang 2019), and the maximum likelihood estimation (Lio and Liu 2019). In addition, Lio and Liu (2018) explored the interval estimation to predict response variables. On the other hand, if the uncertain regression model includes more than one response variable, we call it multivariate uncertain regression model. Song and Fu (2018) applied the least squares estimation in multivariate uncertain regression model where only the observed data of response variables are imprecise. In this paper, we aim to study multivariate uncertain regression model where the observed data of both explanatory variables and response variables are imprecise.

Our work mainly includes parameters estimation, residual analysis, forecast value and confidence interval.

The rest of the paper is organized as follows: In Sect. 2, we propose the multivariate uncertain regression model and estimate the parameters in the model. In Sect. 3, we analyze the residual based on those estimations. In Sect. 4, confidence interval is suggested to forecast the response variables when new explanatory variables are given. In Sect. 5, we provide an example to show the application of the multivariate uncertain regression model. At last, some conclusions are made in Sect. 6.

2 Multivariate uncertain regression model

Assume $(x_1, x_2, \ldots, x_p)$ is a vector of explanatory variables and $(y_1, y_2, \ldots, y_q)$ is a vector of response variables. The functional relationships between y_j and $x_1, x_2, \ldots, x_p$ are assumed to be expressed by the multivariate regression model

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_q \end{pmatrix} = \begin{pmatrix} f_1(x_1, x_2, \ldots, x_p | \boldsymbol{\beta_1}) \\ f_2(x_1, x_2, \ldots, x_p | \boldsymbol{\beta_2}) \\ \vdots \\ f_p(x_1, x_2, \ldots, x_q | \boldsymbol{\beta_q}) \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_q \end{pmatrix} \quad (1)$$

where $\boldsymbol{\beta_j} = (\beta_{0j}, \beta_{1j}, \ldots, \beta_{pj})^T$ are vectors of unknown parameters and ε_j are disturbance terms for $j = 1, 2, \ldots, q$.

In traditional model, we assume that both $(x_1, x_2, \ldots, x_p)$ and $(y_1, y_2, \ldots, y_q)$ are precisely observational. However, the observations we can obtain are imprecise in some cases. And thus, those observations should be characterized as uncertain variables. Assume that we have the observed data of $x_1, x_2, \ldots, x_p$ and $y_1, y_2, \ldots, y_q$ as follows,

$$\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}$$

where $\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}$ are uncertain variables with uncertainty distributions $\Phi_{i1}, \Phi_{i2}, \ldots, \Phi_{ip}, \Psi_{i1}, \Psi_{i2}, \ldots, \Psi_{iq}$ for $i = 1, 2, \ldots, n$, respectively. For simplicity, denote

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta_1} & \boldsymbol{\beta_2} & \cdots & \boldsymbol{\beta_q} \end{pmatrix} = \begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0q} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1q} \\ \vdots & \vdots & & \vdots \\ \beta_{p1} & \beta_{p2} & \cdots & \beta_{pq} \end{pmatrix}.$$

The solution of the minimization problem

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \sum_{j=1}^{q} E\left[(\widetilde{y}_{ij} - f_j(\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip} | \boldsymbol{\beta_j}))^2 \right] \quad (2)$$

is the least squares estimate of $\boldsymbol{\beta}$ in the multivariate regression model (1). Denote the optimal solution by

$$\boldsymbol{\beta}^* = \left(\boldsymbol{\beta}_1^* \ \boldsymbol{\beta}_2^* \ \cdots \ \boldsymbol{\beta}_q^*\right) = \begin{pmatrix} \beta_{01}^* & \beta_{02}^* & \cdots & \beta_{0q}^* \\ \beta_{11}^* & \beta_{12}^* & \cdots & \beta_{1q}^* \\ \vdots & \vdots & & \vdots \\ \beta_{p1}^* & \beta_{p2}^* & \cdots & \beta_{pq}^* \end{pmatrix}.$$

Then the fitted regression model is

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_q \end{pmatrix} = \begin{pmatrix} f_1(x_1, x_2, \ldots, x_p | \boldsymbol{\beta}_1^*) \\ f_2(x_1, x_2, \ldots, x_p | \boldsymbol{\beta}_2^*) \\ \vdots \\ f_p(x_1, x_2, \ldots, x_q | \boldsymbol{\beta}_q^*) \end{pmatrix}. \tag{3}$$

Theorem 1 *Suppose the imprecisely observed data $\widetilde{x}_{i1}$, $\widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}, i = 1, 2, \ldots, n$ are independent uncertain variables. And assume those uncertain variables $\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}$ have regular uncertainty distributions $\Phi_{i1}, \Phi_{i2}, \ldots, \Phi_{ip}, \Psi_{i1}, \Psi_{i2}, \ldots, \Psi_{iq}, i = 1, 2, \ldots, n$, respectively. Then the least squares estimate of*

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0q} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1q} \\ \vdots & \vdots & & \vdots \\ \beta_{p1} & \beta_{p2} & \cdots & \beta_{pq} \end{pmatrix}$$

in the multivariate linear regression model

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_q \end{pmatrix} = \begin{pmatrix} \beta_{01} + \sum_{k=1}^{p} \beta_{k1} x_k \\ \beta_{02} + \sum_{k=1}^{p} \beta_{k2} x_k \\ \vdots \\ \beta_{0q} + \sum_{k=1}^{p} \beta_{kq} x_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_q \end{pmatrix}$$

is the solution of the following problem:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \sum_{j=1}^{q} \int_0^1 \left(\Psi_{ij}^{-1}(\alpha) - \beta_{0j} \right.$$

$$\left. - \sum_{k=1}^{p} \beta_{kj} Y_{ik}^{-1}(\alpha, \beta_{kj}) \right)^2 d\alpha$$

where

$$Y_{ik}^{-1}(\alpha, \beta_{kj}) = \begin{cases} \Phi_{ik}^{-1}(1-\alpha), & \text{if } \beta_{kj} \geq 0 \\ \Phi_{ik}^{-1}(\alpha), & \text{if } \beta_{kj} < 0 \end{cases}$$

for $i = 1, 2, \ldots, n, j = 1, 2, \ldots, q$ and $k = 1, 2, \ldots, p$.

Proof The least squares estimate of $\boldsymbol{\beta}$ in the linear regression model is the optimal solution of the following problem,

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \sum_{j=1}^{q} E\left[\left(\widetilde{y}_{i1} - \beta_{0j} - \sum_{k=1}^{p} \beta_{kj} \widetilde{x}_{ik} \right)^2 \right].$$

Since $\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}, i = 1, 2, \ldots, n$ are independent, we can obtain that $\widetilde{y}_{ij} - \beta_{0j} - \sum_{k=1}^{p} \beta_{kj} \widetilde{x}_{ik}$ have the inverse uncertainty distributions

$$\Psi_{ij}^{-1}(\alpha) - \beta_{0j} - \sum_{k=1}^{p} \beta_{kj} Y_{ik}^{-1}(\alpha, \beta_{kj})$$

for $i = 1, 2, \ldots, n, j = 1, 2, \ldots, q$, respectively. Thus,

$$E\left[\left(\widetilde{y}_{i1} - \beta_{0j} - \sum_{k=1}^{p} \beta_{kj} \widetilde{x}_{ik} \right)^2 \right]$$

$$= \int_0^1 \left(\Psi_{ij}^{-1}(\alpha) - \beta_{0j} - \sum_{k=1}^{p} \beta_{kj} Y_{ik}^{-1}(\alpha, \beta_{kj}) \right)^2 d\alpha,$$

$i = 1, 2, \ldots, n, j = 1, 2, \ldots, q$. Therefore, the least squares estimate of $\boldsymbol{\beta}$ in the multivariate linear regression model is the solution of the minimization problem,

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \sum_{j=1}^{q} \int_0^1 \left(\Psi_{ij}^{-1}(\alpha) - \beta_{0j} \right.$$

$$\left. - \sum_{k=1}^{p} \beta_{kj} Y_{ik}^{-1}(\alpha, \beta_{kj}) \right)^2 d\alpha.$$

The theorem is proved. □

Theorem 2 *Suppose the imprecisely observed data $\widetilde{x}_{i1}$, $\widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}, i = 1, 2, \ldots, n$ are independent uncertain variables. And assume those uncertain variables $\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}$ have regular uncertainty distributions $\Phi_{i1}, \Phi_{i2}, \ldots, \Phi_{ip}, \Psi_{i1}, \Psi_{i2}, \ldots, \Psi_{iq}, i = 1, 2, \ldots, n$, respectively. Then the least squares estimate of*

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0q} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1q} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2q} \end{pmatrix}$$

in the multivariate asymptotic regression model

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_q \end{pmatrix} = \begin{pmatrix} \beta_{01} - \beta_{11} \exp(-\beta_{21}x) \\ \beta_{02} - \beta_{12} \exp(-\beta_{22}x) \\ \vdots \\ \beta_{0q} - \beta_{1q} \exp(-\beta_{2q}x) \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_q \end{pmatrix},$$

$$\beta_{0j} > 0, \beta_{1j} > 0, \beta_{2j} > 0, j = 1, 2, \ldots, q$$

is the optimal solution of the following problem:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n \sum_{j=1}^q \int_0^1 \left(\Psi_{ij}^{-1}(\alpha) - \beta_{0j} \right. $$
$$+ \beta_{1j} \exp(-\beta_{2j} \Phi_i^{-1}(1-\alpha)) \right)^2 d\alpha.$$

Proof The least squares estimate of $\boldsymbol{\beta}$ in the asymptotic regression model is actually the optimal solution of the minimization problem,

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n \sum_{j=1}^q E\left[\left(\widetilde{y}_{ij} - \beta_{0j} + \beta_{1j} \exp(-\beta_{2j}\widetilde{x}_i) \right)^2 \right].$$

Since $\widetilde{x}_i, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}, i = 1, 2, \ldots, n$ are independent, we can obtain that $\widetilde{y}_{ij} - \beta_{0j} + \beta_{1j} \exp(-\beta_{2j}\widetilde{x}_i)$ have the inverse uncertainty distributions

$$\Psi_{ij}^{-1}(\alpha) - \beta_{0j} + \beta_{1j} \exp(-\beta_{2j}\Phi_i^{-1}(1-\alpha))$$

for $i = 1, 2, \ldots, n, j = 1, 2, \ldots, q$, respectively. Thus,

$$E\left[\left(\widetilde{y}_{ij} - \beta_{0j} + \beta_{1j} \exp(-\beta_{2j}\widetilde{x}_i) \right)^2 \right]$$
$$= \int_0^1 \left(\Psi_{ij}^{-1}(\alpha) - \beta_{0j} + \beta_{1j} \exp(-\beta_{2j}\Phi_{ik}^{-1}(1-\alpha)) \right)^2 d\alpha,$$

$i = 1, 2, \ldots, n, j = 1, 2, \ldots, q$. Therefore, the least squares estimate of $\boldsymbol{\beta}$ in the multivariate asymptotic regression model is the solution of the minimization problem,

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n \sum_{j=1}^q \int_0^1 \left(\Psi_{ij}^{-1}(\alpha) - \beta_{0j} \right.$$
$$+ \beta_{1j} \exp(-\beta_{2j}\Phi_i^{-1}(1-\alpha)) \right)^2 d\alpha.$$

The theorem is proved. □

Theorem 3 *Suppose the imprecisely observed data* $\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}, i = 1, 2, \ldots, n$ *are independent uncertain variables. And assume those uncertain variables* $\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}$ *have regular uncertainty distributions* $\Phi_{i1}, \Phi_{i2}, \ldots, \Phi_{ip}, \Psi_{i1}, \Psi_{i2}, \ldots, \Psi_{iq}, i = 1, 2, \ldots, n,$ *respectively. Then the least squares estimate of*

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1q} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2q} \end{pmatrix}$$

in the multivariate Michaelis-Menten regression model

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_q \end{pmatrix} = \begin{pmatrix} \dfrac{\beta_{11}x}{\beta_{21}+x} \\ \dfrac{\beta_{12}x}{\beta_{22}+x} \\ \vdots \\ \dfrac{\beta_{1q}x}{\beta_{2q}+x} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_q \end{pmatrix},$$

$$\beta_{1j} > 0, \beta_{2j} > 0, j = 1, 2, \ldots, q$$

is the optimal solution of the following problem:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n \sum_{j=1}^q \int_0^1 \left(\Psi_{ij}^{-1}(\alpha) - \frac{\beta_{1j}\Phi_i^{-1}(1-\alpha)}{\beta_{2j}+\Phi_i^{-1}(1-\alpha)} \right)^2 d\alpha.$$

Proof The least squares estimate of $\boldsymbol{\beta}$ in the Michaelis-Menten regression model is the optimal solution of the minimization problem,

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n \sum_{j=1}^q E\left[\left(\widetilde{y}_{ij} - \frac{\beta_{1j}\widetilde{x}_i}{\beta_{2j}+\widetilde{x}_i} \right)^2 \right].$$

Since $\widetilde{x}_i, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}$ are independent, we can obtain that

$\widetilde{y}_{ij} - \dfrac{\beta_{1j}\widetilde{x}_i}{\beta_{2j}+\widetilde{x}_i}$ have the inverse uncertainty distributions

$$\Psi_{ij}^{-1}(\alpha) - \frac{\beta_{1j}\Phi_i^{-1}(1-\alpha)}{\beta_{2j}+\Phi_i^{-1}(1-\alpha)}$$

for $i = 1, 2, \ldots, n, j = 1, 2, \ldots, q$, respectively. Thus,

$$E\left[\left(\widetilde{y}_{ij} - \frac{\beta_{1j}\widetilde{x}_i}{\beta_{2j}+\widetilde{x}_i} \right)^2 \right]$$
$$= \int_0^1 \left(\Psi_{ij}^{-1}(\alpha) - \frac{\beta_{1j}\Phi_i^{-1}(1-\alpha)}{\beta_{2j}+\Phi_i^{-1}(1-\alpha)} \right)^2 d\alpha,$$

$i = 1, 2, \ldots, n, j = 1, 2, \ldots, q$. Therefore, the least squares estimate of $\boldsymbol{\beta}$ in the multivariate Michaelis-Menten regression model is the solution of the minimization problem,

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n \sum_{j=1}^q \int_0^1 \left(\Psi_{ij}^{-1}(\alpha) - \frac{\beta_{1j}\Phi_i^{-1}(1-\alpha)}{\beta_{2j}+\Phi_i^{-1}(1-\alpha)} \right)^2 d\alpha.$$

The theorem is proved. □

3 Multivariate residual analysis

In the regression model (1), there is a disturbance term $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_q)^{\mathrm{T}}$. It is difficult to discover the disturbance term $\boldsymbol{\varepsilon}$ exactly since the term changes for each

observation. Thus, we are concerned about how to estimate ε based on imprecisely observed data, $\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots,$ $\widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}, i = 1, 2, \ldots, n.$

Definition 1 Assume the fitted regression model is

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_q \end{pmatrix} = \begin{pmatrix} f_1(x_1, x_2, \ldots, x_p | \boldsymbol{\beta_1^*}) \\ f_2(x_1, x_2, \ldots, x_p | \boldsymbol{\beta_2^*}) \\ \vdots \\ f_q(x_1, x_2, \ldots, x_p | \boldsymbol{\beta_q^*}) \end{pmatrix}$$

and $\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}, i = 1, 2, \ldots, n$ are imprecisely observed data. Then we call

$$\hat{\boldsymbol{\varepsilon}}_i = \begin{pmatrix} \hat{\varepsilon}_{i1} \\ \hat{\varepsilon}_{i2} \\ \vdots \\ \hat{\varepsilon}_{iq} \end{pmatrix} = \begin{pmatrix} \widetilde{y}_{i1} - f_1(\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip} | \boldsymbol{\beta_1^*}) \\ \widetilde{y}_{i2} - f_2(\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip} | \boldsymbol{\beta_2^*}) \\ \vdots \\ \widetilde{y}_{iq} - f_q(\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip} | \boldsymbol{\beta_q^*}) \end{pmatrix}$$

the i-th residual for each i $(i = 1, 2, \ldots, n)$.

Suppose that the disturbance term $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_q)^{\mathrm{T}}$ is an uncertain vector. Then, for each j $(j = 1, 2, \ldots, q)$, we use the average of the expected values of residuals, i.e.,

$$\hat{e}_j = \frac{1}{n} \sum_{i=1}^{n} E[\hat{\varepsilon}_{ij}]$$

to estimate the expected values of the disturbance term ε_j, and

$$\hat{\sigma}_j^2 = \frac{1}{n} \sum_{i=1}^{n} E[(\hat{\varepsilon}_{ij} - \hat{e}_j)^2]$$

to estimate the variances. Then, we call

$$\hat{e} = \begin{pmatrix} \hat{e}_1 \\ \hat{e}_2 \\ \vdots \\ \hat{e}_q \end{pmatrix}, \quad \hat{\boldsymbol{\sigma}}^2 = \begin{pmatrix} \hat{\sigma}_1^2 \\ \hat{\sigma}_2^2 \\ \vdots \\ \hat{\sigma}_q^2 \end{pmatrix}$$

the vectors of the estimated expected values and estimated variances of disturbance term $\boldsymbol{\varepsilon}$, respectively.

Theorem 4 *Suppose the imprecisely observed data* $\widetilde{x}_{i1},$ $\widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}, i = 1, 2, \ldots, n$ *are independent uncertain variables. And assume those uncertain variables* $\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}, i = 1, 2, \ldots, n,$ *have regular uncertainty distributions* $\Phi_{i1}, \Phi_{i2}, \ldots, \Phi_{ip}, \Psi_{i1}, \Psi_{i2}, \ldots, \Psi_{iq}, i = 1, 2, \ldots, n,$ *respectively. Let the fitted multivariate linear regression model be*

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_q \end{pmatrix} = \begin{pmatrix} \beta_{01}^* + \sum_{k=1}^{p} \beta_{k1}^* x_k \\ \beta_{02}^* + \sum_{k=1}^{p} \beta_{k2}^* x_k \\ \vdots \\ \beta_{0q}^* + \sum_{k=1}^{p} \beta_{kq}^* x_k \end{pmatrix}.$$

Then the vector of estimated expected values of the disturbance term $\boldsymbol{\varepsilon}$ is

$$\hat{e} = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \left(\Psi_{i1}^{-1}(\alpha) - \beta_{01}^* \\ - \sum_{k=1}^{p} \beta_{k1}^* Y_{ik}^{-1}(\alpha, \beta_{k1}^*) \right) \mathrm{d}x \\ \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \left(\Psi_{i2}^{-1}(\alpha) - \beta_{02}^* \\ - \sum_{k=1}^{p} \beta_{k2}^* Y_{ik}^{-1}(\alpha, \beta_{k2}^*) \right) \mathrm{d}x \\ \vdots \\ \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \left(\Psi_{iq}^{-1}(\alpha) - \beta_{0q}^* \\ - \sum_{k=1}^{p} \beta_{kq}^* Y_{ik}^{-1}(\alpha, \beta_{kq}^*) \right) \mathrm{d}x \end{pmatrix}$$

and the vector of estimated variances is

$$\hat{\boldsymbol{\sigma}}^2 = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \left(\Psi_{i1}^{-1}(\alpha) - \beta_{01}^* \\ - \sum_{k=1}^{p} \beta_{k1}^* Y_{ik}^{-1}(\alpha, \beta_{k1}^*) - \hat{e}_1 \right)^2 \mathrm{d}x \\ \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \left(\Psi_{i2}^{-1}(\alpha) - \beta_{02}^* \\ - \sum_{k=1}^{p} \beta_{k2}^* Y_{ik}^{-1}(\alpha, \beta_{k2}^*) - \hat{e}_2 \right)^2 \mathrm{d}x \\ \vdots \\ \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \left(\Psi_{iq}^{-1}(\alpha) - \beta_{0q}^* \\ - \sum_{k=1}^{p} \beta_{kq}^* Y_{ik}^{-1}(\alpha, \beta_{kq}^*) - \hat{e}_q \right)^2 \mathrm{d}x \end{pmatrix}$$

where

$$Y_{ik}^{-1}(\alpha, \beta_{kj}^*) = \begin{cases} \Phi_{ik}^{-1}(1 - \alpha), & \text{if } \beta_{kj}^* \geq 0 \\ \Phi_{ik}^{-1}(\alpha), & \text{if } \beta_{kj}^* < 0 \end{cases}$$

for $i = 1, 2, \ldots, n, j = 1, 2, \ldots, q$ and $k = 1, 2, \ldots, p$.

Proof Since the inverse uncertainty distributions of $\widetilde{y}_{ij} - \beta_{0j}^* - \sum_{k=1}^{p} \beta_{kj}^* \widetilde{x}_{ik}$ are

$$\Psi_{ij}^{-1}(\alpha) - \beta_{0j}^* - \sum_{k=1}^{p} \beta_{kj}^* Y_{ik}^{-1}(\alpha, \beta_{kj}^*)$$

for $i = 1, 2, \ldots, n, j = 1, 2, \ldots, q$, respectively, Theorem 4 holds immediately. $\square$

Theorem 5 *Suppose the imprecisely observed data* $\widetilde{x}_{i1},$ $\widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}, i = 1, 2, \ldots, n$ *are independent*

uncertain variables. And assume those uncertain variables $\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}$ *have regular uncertainty distributions* $\Phi_{i1}, \Phi_{i2}, \ldots, \Phi_{ip}, \Psi_{i1}, \Psi_{i2}, \ldots, \Psi_{iq}, i = 1, 2, \ldots, n,$ *respectively. Let the fitted multivariate asymptotic regression model be*

$$
\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_q \end{pmatrix} = \begin{pmatrix} \beta_{01}^* - \beta_{11} \exp\left(-\beta_{21}^* x\right) \\ \beta_{02}^* - \beta_{12}^* \exp\left(-\beta_{22}^* x\right) \\ \vdots \\ \beta_{0q}^* - \beta_{1q}^* \exp\left(-\beta_{2q}^* x\right) \end{pmatrix},
$$
$$
\beta_{0j}^* > 0, \beta_{1j}^* > 0, \beta_{2j}^* > 0, j = 1, 2, \ldots, q.
$$

Then the vector of estimated expected values of the disturbance term $\boldsymbol{\varepsilon}$ *is*

$$
\hat{\boldsymbol{e}} = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n \int_0^1 \left(\Psi_{i1}^{-1}(\alpha) - \beta_{01}^* \\ + \beta_{11}^* \exp\left(-\beta_{21}^* \Phi_i^{-1}(1-\alpha)\right) \right) \mathrm{d}x \\ \frac{1}{n} \sum_{i=1}^n \int_0^1 \left(\Psi_{i2}^{-1}(\alpha) - \beta_{02}^* \\ + \beta_{12}^* \exp\left(-\beta_{22}^* \Phi_i^{-1}(1-\alpha)\right) \right) \mathrm{d}x \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n \int_0^1 \left(\Psi_{iq}^{-1}(\alpha) - \beta_{0q}^* \\ + \beta_{1q}^* \exp\left(-\beta_{2q}^* \Phi_i^{-1}(1-\alpha)\right) \right) \mathrm{d}x \end{pmatrix}
$$

and the vector of estimated variances is

$$
\hat{\boldsymbol{\sigma}}^2 = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n \int_0^1 \left(\Psi_{i1}^{-1}(\alpha) - \beta_{01}^* \\ + \beta_{11}^* \exp\left(-\beta_{21}^* \Phi_i^{-1}(1-\alpha)\right) - \hat{e}_1 \right)^2 \mathrm{d}x \\ \frac{1}{n} \sum_{i=1}^n \int_0^1 \left(\Psi_{i2}^{-1}(\alpha) - \beta_{02}^* \\ + \beta_{12}^* \exp\left(-\beta_{22}^* \Phi_i^{-1}(1-\alpha)\right) - \hat{e}_2 \right)^2 \mathrm{d}x \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n \int_0^1 \left(\Psi_{iq}^{-1}(\alpha) - \beta_{0q}^* \\ + \beta_{1q}^* \exp\left(-\beta_{2q}^* \Phi_i^{-1}(1-\alpha)\right) - \hat{e}_q \right)^2 \mathrm{d}x \end{pmatrix}.
$$

Proof Since the inverse uncertainty distributions of $\widetilde{y}_{ij} - \beta_{0j}^* + \beta_{1j}^* \exp\left(-\beta_{2j}^* \widetilde{x}_i\right)$ are

$$
\Psi_{ij}^{-1}(\alpha) - \beta_{0j}^* + \beta_{1j}^* \exp\left(-\beta_{2j}^* \Phi_i^{-1}(1-\alpha)\right)
$$

for $i = 1, 2, \ldots, n, j = 1, 2, \ldots, q,$ respectively, Theorem 5 holds immediately. □

Theorem 6 *Suppose the imprecisely observed data* $\widetilde{x}_{i1},$ $\widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}, i = 1, 2, \ldots, n$ *are independent uncertain variables. And assume those uncertain variables* $\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}, i = 1, 2, \ldots, n,$ *have regular uncertainty distributions* $\Phi_{i1}, \Phi_{i2}, \ldots, \Phi_{ip}, \Psi_{i1}, \Psi_{i2}, \ldots, \Psi_{iq}, i = 1, 2, \ldots, n,$

respectively. Let the fitted multivariate Michaelis–enten regression model be

$$
\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_q \end{pmatrix} = \begin{pmatrix} \dfrac{\beta_{11}^* x}{\beta_{21}^* + x} \\ \dfrac{\beta_{12}^* x}{\beta_{22}^* + x} \\ \vdots \\ \dfrac{\beta_{1q}^* x}{\beta_{2q}^* + x} \end{pmatrix}, \beta_{1j}^* > 0, \beta_{2j}^* > 0, j = 1, 2, \ldots, q.
$$

Then the vector of estimated expected values of the disturbance term $\boldsymbol{\varepsilon}$ *is*

$$
\hat{\boldsymbol{e}} = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n \int_0^1 \left(\Psi_{i1}^{-1}(\alpha) - \frac{\beta_{11}^* \Phi_i^{-1}(1-\alpha)}{\beta_{21}^* + \Phi_i^{-1}(1-\alpha)} \right) \mathrm{d}x \\ \frac{1}{n} \sum_{i=1}^n \int_0^1 \left(\Psi_{i2}^{-1}(\alpha) - \frac{\beta_{12}^* \Phi_i^{-1}(1-\alpha)}{\beta_{22}^* + \Phi_i^{-1}(1-\alpha)} \right) \mathrm{d}x \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n \int_0^1 \left(\Psi_{iq}^{-1}(\alpha) - \frac{\beta_{1q}^* \Phi_i^{-1}(1-\alpha)}{\beta_{2q}^* + \Phi_i^{-1}(1-\alpha)} \right) \mathrm{d}x \end{pmatrix}
$$

and the vector of estimated variances is

$$
\hat{\boldsymbol{\sigma}}^2 = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n \int_0^1 \left(\Psi_{i1}^{-1}(\alpha) - \frac{\beta_{11}^* \Phi_i^{-1}(1-\alpha)}{\beta_{21}^* + \Phi_i^{-1}(1-\alpha)} - \hat{e}_1 \right)^2 \mathrm{d}x \\ \frac{1}{n} \sum_{i=1}^n \int_0^1 \left(\Psi_{i2}^{-1}(\alpha) - \frac{\beta_{12}^* \Phi_i^{-1}(1-\alpha)}{\beta_{22}^* + \Phi_i^{-1}(1-\alpha)} - \hat{e}_2 \right)^2 \mathrm{d}x \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n \int_0^1 \left(\Psi_{iq}^{-1}(\alpha) - \frac{\beta_{1q}^* \Phi_i^{-1}(1-\alpha)}{\beta_{2q}^* + \Phi_i^{-1}(1-\alpha)} - \hat{e}_q \right)^2 \mathrm{d}x \end{pmatrix}.
$$

Proof Since the inverse uncertainty distributions of $\widetilde{y}_{ij} - \dfrac{\beta_{1j} \widetilde{x}_i}{\beta_{2j} + \widetilde{x}_i}$ are

$$
\Psi_{ij}^{-1}(\alpha) - \frac{\beta_{1j} \Phi_i^{-1}(1-\alpha)}{\beta_{2j} + \Phi_i^{-1}(1-\alpha)}
$$

for $i = 1, 2, \ldots, n, j = 1, 2, \ldots, q,$ respectively, Theorem 6 holds immediately. □

4 Forecast value and confidence interval

In Sects. 2 and 3, we obtain the least squares estimate $\boldsymbol{\beta}^*$ and the estimations of the expected value $\hat{\boldsymbol{e}}$ and the variance $\hat{\boldsymbol{\sigma}}^2$ of disturbance term $\boldsymbol{\varepsilon}$ based on imprecisely observed data $(\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{ip}, \widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{iq}), i = 1, 2, \ldots, n.$ Based on the work, we are interested in forecasting the response vector for the new explanatory vector. Assume that $\widetilde{\boldsymbol{x}} = (\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_p)^{\mathrm{T}}$ is a vector of new explanatory variables where $\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_p$ are independent uncertain variables. Suppose those uncertain variables $\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_p$ have regular uncertainty

distributions $\Phi_1, \Phi_2, \ldots, \Phi_p$, respectively. Although the relationship between uncertain explanatory vector and the uncertain response vector may be complicated, it is still valuable and useful to apply linear regression model for the data. Suppose that the fitted linear regression model is

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_q \end{pmatrix} = \begin{pmatrix} \beta_{01}^* + \sum_{k=1}^p \beta_{k1}^* x_k \\ \beta_{02}^* + \sum_{k=1}^p \beta_{k2}^* x_k \\ \vdots \\ \beta_{0q}^* + \sum_{k=1}^p \beta_{kq}^* x_k \end{pmatrix}$$

and the disturbance term ε has the estimated expected value $\hat{e}$ and variance $\hat{\sigma}^2$, and is independent of $\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_p$. Then the forecast uncertain vector of $\boldsymbol{y} = (y_1, y_2, \ldots, y_q)^{\mathrm{T}}$ with respect to $(\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_p)$ is determined by

$$\hat{\boldsymbol{y}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_q \end{pmatrix} = \begin{pmatrix} \beta_{01}^* + \sum_{k=1}^p \beta_{k1}^* \widetilde{x}_k \\ \beta_{02}^* + \sum_{k=1}^p \beta_{k2}^* \widetilde{x}_k \\ \vdots \\ \beta_{0q}^* + \sum_{k=1}^p \beta_{kq}^* \widetilde{x}_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_q \end{pmatrix}. \tag{4}$$

For each j ($j = 1, 2, \ldots, q$), a single value of y_j should be estimated from the forecast uncertain vector, and it is natural to define the forecast value of y_j as

$$\mu_j = \beta_{0j}^* + \sum_{k=1}^p \beta_{kj}^* E[\widetilde{x}_k] + \hat{e}_j$$

which is the expected value of the forecast uncertain variable $\hat{y}_j$. Then we write

$$\boldsymbol{\mu} = \begin{pmatrix} \beta_{01}^* + \sum_{k=1}^p \beta_{k1}^* E[\widetilde{x}_k] + \hat{e}_1 \\ \beta_{02}^* + \sum_{k=1}^p \beta_{k2}^* E[\widetilde{x}_k] + \hat{e}_2 \\ \vdots \\ \beta_{01}^* + \sum_{k=1}^p \beta_{k1}^* E[\widetilde{x}_k] + \hat{e}_q \end{pmatrix} \tag{5}$$

as the forecast value of $\boldsymbol{y}$.

Furthermore, in Eq. (4), if we assume that the disturbances $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_q$ are identically distributed, but their expected values and variances differ across equation. Especially, if $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_q$ are normal uncertain variables $\mathcal{N}(\hat{e}_1, \hat{\sigma}_1), \mathcal{N}(\hat{e}_2, \hat{\sigma}_2), \ldots, \mathcal{N}(\hat{e}_q, \hat{\sigma}_q)$, respectively, then the inverse uncertainty distributions of $\hat{y}_j$ are determined by

$$\hat{\Psi}_j^{-1}(\alpha) = \beta_{0j}^* + \sum_{k=1}^p \beta_{kj}^* \Upsilon_k^{-1}(\alpha, \beta_{kj}) + \phi_j^{-1}(\alpha)$$

where

$$\Upsilon_k^{-1}(\alpha, \beta_{kj}) = \begin{cases} \Phi_k^{-1}(\alpha), & \text{if } \beta_{kj} \geq 0 \\ \Phi_k^{-1}(1 - \alpha), & \text{if } \beta_{kj} < 0 \end{cases}$$

and $\phi_j^{-1}(\alpha)$ are the inverse uncertainty distributions of ε_j, i.e.,

$$\phi_j^{-1}(\alpha) = \hat{e}_j + \frac{\hat{\sigma}_j}{\pi}$$

for $j = 1, 2, \ldots, q, k = 1, 2, \ldots, p$, respectively. Then the uncertainty distributions $\hat{\Psi}_j$ of $\hat{y}_j$ can be obtained by $\hat{\Psi}_j^{-1}$, $j = 1, 2, \ldots, q$, respectively.

For each j ($j = 1, 2, \ldots, q$), the forecast value μ_j is a point estimation of y_j. However, it is not convincing to claim that the value of each component of uncertain vector $\boldsymbol{y}$ is always a precise value. In fact, the point estimates are hard to convince people that they are accurate. If the estimation is a range, like $3 \sim 4$, it is more convincing. Although the value range is wider, the reliability is obviously higher. Thus, we propose the confidence interval to estimate $\boldsymbol{y}$.

Taking α (e.g., 95%) as a confidence level, we are interested in finding the minimum values b_j such that

$$\hat{\Psi}_j(\mu_j + b_j) - \hat{\Psi}_j(\mu_j - b_j) \geq \alpha, \tag{6}$$

$j = 1, 2, \ldots, q$, respectively. Since

$$\mathcal{M}\{\mu_j - b_j \leq \hat{y}_j \leq \mu_j + b_j\} \geq \hat{\Psi}_j(\mu_j + b_j) - \hat{\Psi}_j(\mu_j - b_j) \geq \alpha,$$

the α confidence intervals of y_j are suggested as $[\mu_j - b_j, \mu_j + b_j]$, which can be abbreviated as $\mu_j \pm b_j$, $j = 1, 2, \ldots, q$, respectively. Denote

$$\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_q \end{pmatrix}.$$

Then, the α confidence interval of $\boldsymbol{y}$ is written as

$$\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{pmatrix} \pm \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_q \end{pmatrix}$$

which represents the set

$$[\boldsymbol{\mu} - \boldsymbol{b}, \boldsymbol{\mu} + \boldsymbol{b}] = \{\boldsymbol{x} = (x_1, x_2, \ldots, x_q) \in \mathfrak{R}^q :$$
$$\mu_j - b_j \leq x_j \leq \mu_j + b_j, j = 1, 2, \ldots, q\}$$

and we have a chance of α to cover $\boldsymbol{y}$ with our confidence interval.

5 Numerical example

In this section, we design a numerical example to show the estimation of unknown parameters, residual analysis, forecast value and confidence interval in multivariate uncertain regression model.

Consider the linear regression model

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \beta_{01} + \beta_{11}x_1 + \beta_{21}x_2 + \beta_{31}x_3 + \beta_{41}x_4 \\ \beta_{02} + \beta_{12}x_1 + \beta_{22}x_2 + \beta_{32}x_3 + \beta_{42}x_4 \\ \beta_{03} + \beta_{13}x_1 + \beta_{23}x_2 + \beta_{33}x_3 + \beta_{43}x_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}.$$

Denote

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{01} & \beta_{02} & \beta_{03} \\ \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \\ \beta_{41} & \beta_{42} & \beta_{43} \end{pmatrix}.$$

Suppose imprecisely observed data $\tilde{x}_{i1}, \tilde{x}_{i2}, \tilde{x}_{i3}, \tilde{x}_{i4}, \tilde{y}_{i1}, \tilde{y}_{i2}, \tilde{y}_{i3}$, $i = 1, 2, \ldots, 21$ are independent uncertain variables, where $\tilde{x}_{i1}, \tilde{x}_{i2}, \tilde{x}_{i3}, \tilde{x}_{i4}, \tilde{y}_{i1}, \tilde{y}_{i2}, \tilde{y}_{i3}$ have linear uncertainty distributions, $\boldsymbol{\Phi}_{i1}, \boldsymbol{\Phi}_{i2}, \boldsymbol{\Phi}_{i3}, \boldsymbol{\Phi}_{i4}, \boldsymbol{\Psi}_{i1}, \boldsymbol{\Psi}_{i2}, \boldsymbol{\Psi}_{i3}$, respectively. See the data in Table 1.

First, we estimate the unknown parameters. That is, we should solve the following problem,

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{21} \sum_{j=1}^{3} E\big[\tilde{y}_{ij} - \beta_{0j} - \beta_{1j}\tilde{x}_{i1} - \beta_{2j}\tilde{x}_{i2} - \beta_{3j}\tilde{x}_{i3} \\ - \beta_{4j}\tilde{x}_{i4})^2 \big]. \tag{7}$$

It follows from Theorem 1 that the equation (7) can be transformed to an equivalent form

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{21} \sum_{j=1}^{3} \int_0^1 \Big(\Psi_{ij}^{-1}(\alpha) - \beta_{0j} - \beta_{1j}Y_{i1}^{-1}(\alpha, \beta_{1j}) \\ - \beta_{2j}Y_{i2}^{-1}(\alpha, \beta_{2j}) - \beta_{3j}Y_{i3}^{-1}(\alpha, \beta_{3j}) - \beta_{4j}Y_{i4}^{-1}(\alpha, \beta_{4j}) \Big)^2 d\alpha$$

where

$$Y_{ik}^{-1}(\alpha, \beta_{kj}) = \begin{cases} \Phi_{ik}^{-1}(1-\alpha), & \text{if } \beta_{kj} \geq 0 \\ \Phi_{ik}^{-1}(\alpha), & \text{if } \beta_{kj} < 0 \end{cases}$$

for $i = 1, 2, \ldots, 21, j = 1, 2, \ldots, 3$ and $k = 1, 2, \ldots, 4$. Then we can obtain the least squares estimate

$$\boldsymbol{\beta}^* = \begin{pmatrix} \beta_{01}^* & \beta_{02}^* & \beta_{03}^* \\ \beta_{11}^* & \beta_{12}^* & \beta_{13}^* \\ \beta_{21}^* & \beta_{22}^* & \beta_{23}^* \\ \beta_{31}^* & \beta_{32}^* & \beta_{33}^* \\ \beta_{41}^* & \beta_{42}^* & \beta_{43}^* \end{pmatrix} = \begin{pmatrix} 20.7786 & 15.8565 & 14.4941 \\ 1.1164 & 0.9178 & 1.2615 \\ 0.3080 & 0.5033 & 0.6114 \\ 0.9210 & 1.2423 & 1.4588 \\ 0.2032 & 0.2775 & 0.3950 \end{pmatrix}.$$

Thus, the fitted linear multivariate regression model is

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 20.7786 & 15.8565 & 14.4941 \\ 1.1164 & 0.9178 & 1.2615 \\ 0.3080 & 0.5033 & 0.6114 \\ 0.9210 & 1.2423 & 1.4588 \\ 0.2032 & 0.2775 & 0.3950 \end{pmatrix}^T \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}.$$

Table 1 Imprecisely observed data where $\mathcal{L}(a, b)$ represents linear uncertain variable

i	$\tilde{x}_{i1}$	$\tilde{x}_{i2}$	$\tilde{x}_{i3}$	$\tilde{x}_{i4}$	$\tilde{y}_{i1}$	$\tilde{y}_{i2}$	$\tilde{y}_{i3}$
1	$\mathcal{L}(6,7)$	$\mathcal{L}(18,21)$	$\mathcal{L}(3,4)$	$\mathcal{L}(25,26)$	$\mathcal{L}(41,44)$	$\mathcal{L}(43,44)$	$\mathcal{L}(49,50)$
2	$\mathcal{L}(3,4)$	$\mathcal{L}(17,18)$	$\mathcal{L}(9,10)$	$\mathcal{L}(21,24)$	$\mathcal{L}(43,44)$	$\mathcal{L}(45,47)$	$\mathcal{L}(52,53)$
3	$\mathcal{L}(9,10)$	$\mathcal{L}(12,14)$	$\mathcal{L}(5,6)$	$\mathcal{L}(18,19)$	$\mathcal{L}(43,45)$	$\mathcal{L}(42,44)$	$\mathcal{L}(49,51)$
4	$\mathcal{L}(7,8)$	$\mathcal{L}(15,16)$	$\mathcal{L}(4,5)$	$\mathcal{L}(20,22)$	$\mathcal{L}(41,43)$	$\mathcal{L}(41,42)$	$\mathcal{L}(47,48)$
5	$\mathcal{L}(5,6)$	$\mathcal{L}(13,16)$	$\mathcal{L}(5,6)$	$\mathcal{L}(27,30)$	$\mathcal{L}(41,43)$	$\mathcal{L}(42,44)$	$\mathcal{L}(48,51)$
6	$\mathcal{L}(6,7)$	$\mathcal{L}(13,14)$	$\mathcal{L}(2,3)$	$\mathcal{L}(19,22)$	$\mathcal{L}(36,38)$	$\mathcal{L}(35,37)$	$\mathcal{L}(40,42)$
7	$\mathcal{L}(5,6)$	$\mathcal{L}(15,18)$	$\mathcal{L}(3,4)$	$\mathcal{L}(22,24)$	$\mathcal{L}(38,40)$	$\mathcal{L}(38,40)$	$\mathcal{L}(44,46)$
8	$\mathcal{L}(2,3)$	$\mathcal{L}(12,13)$	$\mathcal{L}(10,11)$	$\mathcal{L}(27,29)$	$\mathcal{L}(42,43)$	$\mathcal{L}(44,46)$	$\mathcal{L}(50,52)$
9	$\mathcal{L}(4,5)$	$\mathcal{L}(13,16)$	$\mathcal{L}(10,11)$	$\mathcal{L}(20,22)$	$\mathcal{L}(43,45)$	$\mathcal{L}(45,47)$	$\mathcal{L}(52,53)$
10	$\mathcal{L}(3,4)$	$\mathcal{L}(18,19)$	$\mathcal{L}(7,8)$	$\mathcal{L}(18,19)$	$\mathcal{L}(40,41)$	$\mathcal{L}(41,43)$	$\mathcal{L}(47,48)$
11	$\mathcal{L}(3,4)$	$\mathcal{L}(14,16)$	$\mathcal{L}(2,3)$	$\mathcal{L}(31,33)$	$\mathcal{L}(36,38)$	$\mathcal{L}(37,38)$	$\mathcal{L}(43,44)$
12	$\mathcal{L}(4,5)$	$\mathcal{L}(19,20)$	$\mathcal{L}(4,5)$	$\mathcal{L}(29,31)$	$\mathcal{L}(42,43)$	$\mathcal{L}(43,45)$	$\mathcal{L}(50,51)$
13	$\mathcal{L}(5,6)$	$\mathcal{L}(19,22)$	$\mathcal{L}(5,6)$	$\mathcal{L}(24,25)$	$\mathcal{L}(43,45)$	$\mathcal{L}(45,46)$	$\mathcal{L}(52,53)$
14	$\mathcal{L}(2,3)$	$\mathcal{L}(20,22)$	$\mathcal{L}(9,10)$	$\mathcal{L}(23,26)$	$\mathcal{L}(43,45)$	$\mathcal{L}(47,48)$	$\mathcal{L}(54,55)$
15	$\mathcal{L}(10,11)$	$\mathcal{L}(16,19)$	$\mathcal{L}(2,3)$	$\mathcal{L}(24,25)$	$\mathcal{L}(45,47)$	$\mathcal{L}(45,46)$	$\mathcal{L}(52,54)$
16	$\mathcal{L}(10,11)$	$\mathcal{L}(12,13)$	$\mathcal{L}(2,3)$	$\mathcal{L}(21,22)$	$\mathcal{L}(43,44)$	$\mathcal{L}(40,41)$	$\mathcal{L}(47,48)$
17	$\mathcal{L}(6,7)$	$\mathcal{L}(14,15)$	$\mathcal{L}(3,4)$	$\mathcal{L}(25,26)$	$\mathcal{L}(40,41)$	$\mathcal{L}(39,41)$	$\mathcal{L}(45,46)$
18	$\mathcal{L}(6,7)$	$\mathcal{L}(18,19)$	$\mathcal{L}(7,8)$	$\mathcal{L}(25,26)$	$\mathcal{L}(46,47)$	$\mathcal{L}(48,50)$	$\mathcal{L}(55,57)$
19	$\mathcal{L}(5,6)$	$\mathcal{L}(13,14)$	$\mathcal{L}(8,9)$	$\mathcal{L}(30,32)$	$\mathcal{L}(45,47)$	$\mathcal{L}(47,49)$	$\mathcal{L}(55,56)$
20	$\mathcal{L}(7,8)$	$\mathcal{L}(18,21)$	$\mathcal{L}(4,5)$	$\mathcal{L}(26,27)$	$\mathcal{L}(44,47)$	$\mathcal{L}(46,47)$	$\mathcal{L}(54,55)$
21	$\mathcal{L}(5,6)$	$\mathcal{L}(17,19)$	$\mathcal{L}(6,7)$	$\mathcal{L}(27,30)$	$\mathcal{L}(44,46)$	$\mathcal{L}(44,47)$	$\mathcal{L}(53,55)$

It follows from Theorem 4 that we obtain the vectors of the estimated expected values and estimated variances of disturbance term, i.e.,

$$\hat{e} = \begin{pmatrix} 0.0000 \\ 0.0000 \\ 0.0000 \end{pmatrix}, \ \hat{\sigma}^2 = \begin{pmatrix} 2.3898 \\ 2.9500 \\ 3.8313 \end{pmatrix},$$

respectively. Now assume

$$(\widetilde{x}_1, \widetilde{x}_2, \widetilde{x}_3, \widetilde{x}_4) \sim (\mathcal{L}(5, 7), \mathcal{L}(16, 17), \mathcal{L}(5, 6), \mathcal{L}(24, 26))$$

is a new uncertain explanatory vector and $\widetilde{x}_1, \widetilde{x}_2, \widetilde{x}_3, \widetilde{x}_4$, $\varepsilon_1, \varepsilon_2, \varepsilon_3$ are independent. Then the forecast uncertain vector of $\boldsymbol{y} = (y_1, y_2, y_3)^{\mathrm{T}}$ is

$$\begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{pmatrix} = \begin{pmatrix} 20.7786 & 15.8565 & 14.4941 \\ 1.1164 & 0.9178 & 1.2615 \\ 0.3080 & 0.5033 & 0.6114 \\ 0.9210 & 1.2423 & 1.4588 \\ 0.2032 & 0.2775 & 0.3950 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 1 \\ \widetilde{x}_1 \\ \widetilde{x}_2 \\ \widetilde{x}_3 \\ \widetilde{x}_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}.$$

Hence, it follows from equation (5) that the forecast value of $\boldsymbol{y}$ is

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} 42.7051 \\ 43.4379 \\ 50.0501 \end{pmatrix}.$$

In order to obtain confidence interval of $\boldsymbol{y}$, we take a confidence level $\alpha = 95\%$ and suppose $\varepsilon_1, \varepsilon_2, \varepsilon_3$ are normal uncertain variables $\mathcal{N}(\hat{e}_1, \hat{\sigma}_1), \mathcal{N}(\hat{e}_2, \hat{\sigma}_2), \mathcal{N}(\hat{e}_3, \hat{\sigma}_3)$, respectively. It follows from equation (6) that we can calculate

$$\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 4.9540 \\ 5.4339 \\ 6.5106 \end{pmatrix}$$

such that for each j $(j = 1, 2, 3)$, b_j is the minimum value satisfying

$$\hat{\Psi}_j(\mu_j + b_j) - \hat{\Psi}_j(\mu_j - b_j) \geq \alpha$$

where $\hat{\Psi}_j$ is the uncertainty distribution of $\hat{y}_j$. Thus, the 95% confidence interval of the vector of the response variables $\boldsymbol{y}$ is

$$\begin{pmatrix} 42.7051 \\ 43.4379 \\ 50.0501 \end{pmatrix} \pm \begin{pmatrix} 4.9540 \\ 5.4339 \\ 6.5106 \end{pmatrix}.$$

6 Conclusions

This paper is aimed at studying the multivariate uncertain regression model that contains more than one response variables and assumes both explanatory variables and response variables as uncertain variables since the observed data are imprecise in some cases. Based on those data, we estimate unknown parameters by the principle of least squares in the different multivariate regression model, such as multivariate linear regression model, multivariate asymptotic regression model and multivariate Michaelis-Menten regression model. In order to analyze the disturbance terms in the model, we propose the concepts of residuals, and design the vectors of estimated expected values and variances of disturbance terms. Furthermore, it is significant to forecast the response variables when a set of new explanatory variables is given. In the future, we will try to take the relationships among the response variables into consider.

References

Aitken A (1935) On least squares and linear combinations of observations. Proc R Soc Edinb 55:42–48

Anderson T (1951) Estimating linear restrictions on regression coefficients for multivariate normal distributions. Ann Math Stat 22(3):327–351

Bai Z, Chen X, Miao B, Radhakrishna R (1990) Asymptotic theory of least distances estimate in multivariate linear models. Statistics 21(4):503–519

Bilodeau M, Brenner D (1999) Theory of multivariate statistics. Springer, New York

Breiman L, Friedman J (1997) Predicting multivariate responses in multiple linear regression. J R Stat Soc 59(1):3–54

Fama E, Fisher L, Jensen M, Roll R (1969) The adjustment of stock prices to new information. International Economic Review 10(1):1–21

Galton F (1886) Regression towards mediocrity in hereditary stature. J Anthropol Inst GB Ireland 15:246–263

Ganesh S, Arulmozhivarman P, Tatavarti V (2018) Prediction of $PM_{2.5}$ using an ensemble of artificial neural networks and regression models. J Ambient Intell Humaniz Comput 8:1–11

Gentle J (1977) Least absolute values estimation: an introduction. Commun Stat Simul Comput 6(4):313–328

Jhun M, Choi I (2009) Bootstrapping least distance estimator in the multivariate regression model. Comput Stat Data Anal 53(12):4221–4227

Krishnamurthy K, Manoharan S, Swaminathan R A Jacobian approach for calculating the Lyapunov exponents of short time series using support vector regression. J Ambient Intell Hum Comput. https://doi.org/10.1007/s12652-019-01525-6

Lio W, Liu B (2018) Residual and confidence interval for uncertain regression model with imprecise observations. J Intell Fuzzy Syst 35(2):2573–2583

Lio W, Liu B (2019) Maximum likelihood estimation for uncertain regression analysis. Technical Report

Liu B (2007) Uncertainty theory, 2nd edn. Springer, Berlin

Liu B (2009) Some research problems in uncertainty theory. J Uncert Syst 3(1):3–10

Liu B (2012) Why is there a need for uncertainty theory. J Uncert Syst 6(1):3–10

Liu Z, Yang Y (2019) Least absolute deviations estimation for uncertain regression with imprecise observations. Fuzzy Optim Decis Mak. https://doi.org/10.1007/s10700-019-09312-w

Song Y, Fu Z (2018) Uncertain multivariable regression model. Soft Comput 22(17):5861–5866

Watson G (1967) Linear least squares regression. Ann Math Stat 38(6):1679–1699

Yao K, Liu B (2018) Uncertain regression analysis: an approach for imprecise observations. Soft Comput 22(17):5579–5582

Yule G (1897) On the theory of correlation. J R Stat Soc 60(4):812–854