J Syst Sci Complex (2020) 33: 2048–2066

A Constrained Interval-Valued Linear Regression Model: A New Heteroscedasticity Estimation Method*

ZHONG Yu · ZHANG Zhongzhan · LI Shoumei

DOI: 10.1007/s11424-020-9075-2
Received: 7 March 2019 / Revised: 28 August 2019

Abstract Linear regression models for interval-valued data have been widely studied. Most literatures are to split an interval into two real numbers, i.e., the left- and right-endpoints or the center and radius of this interval, and fit two separate real-valued or two dimension linear regression models. This paper is focused on the bias-corrected and heteroscedasticity-adjusted modeling by imposing order constraint to the endpoints of the response interval and weighted linear least squares with estimated covariance matrix, based on a generalized linear model for interval-valued data. A three step estimation method is proposed. Theoretical conclusions and numerical evaluations show that the proposed estimator has higher efficiency than previous estimators.

Keywords Conditional maximum likelihood estimation, interval-valued data, order constraint, truncated normal distribution, weighted least squares estimation.

1 Introduction

The classical linear regression model is a widely used statistical model, which is usually used to quantify the interdependent relationship between two or more real-valued variables. With the development of technology, data become so complicated that the classical linear model cannot be employed directly. There are growing literatures on modeling complicated data. In this paper, we consider a constrained linear regression model for interval-valued data.

The presence of interval-valued variables is quite common in practice. For instance, a weather forecast provides the lowest and highest temperature of the following day, which is an interval giving the range of temperatures varies during the whole day (Wang, et al.[1]). The blood pressures of people are consisted of systolic blood pressures and diastolic blood pressures, which form interval-valued observations naturally. Many authors analyzed this kind of data,

ZHONG Yu · ZHANG Zhongzhan · LI Shoumei (Corresponding author)
Beijing University of Technology, College of Applied Sciences, Beijing 100020, *China.*
Email: lishoumei2011@163.com.
*This research was supported by the National Nature Science Foundation of China under Grant Nos. 11571024 and 11771032, and the Humanities and Social Science Foundation of Ministry of Education of China under Grant No. 20YJCZH245.
◇ *This paper was recommended for publication by Editor XU Jin.*

see Gil, et al.[2, 3] and Fagundes, et al.[4]. Besides, the range of interval reflects the variation. Some econometric researches have already noticed that modeling the range indeed can improve the estimation of volatility, e.g., Engle and Gallo[5] and its references. González-Rivera and Lin[6], and Wang, et al.[7] analyzed the interval-valued stock index data where interval-valued data not only includes the closing index, but also reflects the daily variation.

On the other hand, as a special case of set-valued data, the study of set-valued random variables has a long history, cf. e.g., [8–16]. As is well known, the space of set-value random variables is not linear with respect to its addition and scalar multiplication, thus various different approaches from the classical modeling of real-valued data have been developed. Among them are linear regression models for interval-valued data.

There are a lot of literatures on linear regression models for interval-valued data. Generally speaking, there are three kinds of modeling methods. The first is to split an interval into two real numbers, i.e., the left- and right-endpoints, or the center and radius of this interval. Through establishing a two-dimensional linear regression model with two endpoints, or centers and radii, the linear relationships between interval-valued random variables can be studied. There are many works using such kind of methods, including MinMax method[17], center method (CM)[18], center and range method (CRM)[19], constrained center and range method (CCRM)[20−23]. The second is to divide an interval into grid points, then establish statistical models through those grid points, e.g., [24–26]. The third is to view the interval-valued data as data units, then use the theory of set-valued random variables to study the statistical inference problem, e.g., [1, 7] and [27].

In this paper, we combine the idea of the third kind methods into the first, and consider a constrained model, in which the left- and right-endpoints of response variable are modeled simultaneously. However, noticing the interval-valued data is a whole, and the whole interval is decided by two endpoints, in the models, we will relate the end points of the response variable not only to the left-endpoints but also to the right-endpoints of explanatory variables. Thus, our model is more general than the cases of [17–23].

Notice that the constrained conditions make the CCRM is not universally suitable, it may be weakly fitted when the response variable and explanatory variables are negatively correlated, because of the constraint condition to the regression coefficients of the interval range. To solve this problem, González-Rivera and Lin[6] combined the ideas of classical truncated regression model (cf. [28, 29]) and Heckman two step estimation (cf. [30–32]), and proposed a new constrained model and estimation method. In [33], Li, et al. built a constrained interval-valued linear regression, in which a two step estimation was given and the second step was the interval-valued least squares estimation. However, heteroscedasticity problems exist in the estimation procedures of [6], [33], and it reduces the asymptotic efficiency of estimators, even though the estimators are consistent.

We will propose a three step estimation procedure in this paper, which is adjusted to the heteroscedasticity based on the two step estimation in [6]. The proposed method is shown having higher efficiency than the two step estimators. The rest of the paper is arranged as follows. Section 2 will list some preliminaries. Section 3 will give the three step estimation procedure in detail. The covariance of errors is also calculated in order to show the heteroscedasticity. At last, we will illustrate that our proposed method is effective through simulation studies and a real data analysis in Section 4.

2 Preliminaries

We will first briefly introduce the formulation. Next, we will present some existing interval-valued linear regression models for easy reference, namely MinMax method, CM, CRM, and CCRM.

2.1 The Space of Closed Intervals

Let $\mathbb{R}$ be the 1-dimensional Euclidean space, and $\boldsymbol{K}_{kc}(\mathbb{R})$ be the family of all non-empty bounded closed intervals in $\mathbb{R}$, i.e., $\boldsymbol{K}_{kc}(\mathbb{R}) = \{A = [\underline{a}, \overline{a}] : -\infty < \underline{a} \le \overline{a} < \infty, \underline{a}, \overline{a} \in \mathbb{R}\}$. We also write the interval $A = [\underline{a}, \overline{a}] = (a^c; a^r)$, where

$$a^c = (\overline{a} + \underline{a})/2, \quad a^r = (\overline{a} - \underline{a})/2$$

are the center and radius of interval A respectively.

The addition and scalar multiplication for intervals $A = [\underline{a}, \overline{a}]$, $B = [\underline{b}, \overline{b}]$ are defined as

$$A + B = [\underline{a} + \underline{b}, \overline{a} + \overline{b}], \quad kA = \begin{cases} [k\underline{a}, k\overline{a}], & k \ge 0, \\ [k\overline{a}, k\underline{a}], & k < 0. \end{cases}$$

Notice that if A does not degenerate to a single point, then $A - A = A + (-A) \ne \{0\}$. Thus, $\boldsymbol{K}_{kc}(\mathbb{R})$ is not a linear space with respect to the addition and scalar multiplication.

Vitale[10] defined the d_p metric for sets, where p is an arbitrary positive integer. The d_p metric between $A = [\underline{a}, \overline{a}]$ and $B = [\underline{b}, \overline{b}]$ is

$$d_p(A, B) = [|\underline{b} - \underline{a}|^p + |\overline{b} - \overline{a}|^p]^{\frac{1}{p}}.$$

Then $(\boldsymbol{K}_{kc}(\mathbb{R}), d_p)$ is a complete separable metric space (see [15]). We will use the averaged d_2^2 to evaluate the goodness of fit of the models in Section 4.

2.2 Interval-Valued Random Variables

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a complete probability space. Interval-valued mapping $X(\omega) : \Omega \to \boldsymbol{K}_{kc}(\mathbb{R})$ is called an interval-valued random variable, if for every closed interval $C \in \boldsymbol{K}_{kc}(\mathbb{R})$, $X^{-1}(C) = \{\omega \in \Omega : X(\omega) \cap C \ne \emptyset\} \in \mathcal{A}$. Using the equivalent definitions of set-valued random variables (e.g., Theorem 1.2.2 in [15]), we can infer that $X = [\underline{x}, \overline{x}] = (x^c; x^r)$ is an interval-valued random variable if $\underline{x}$ and $\overline{x}$, or x^c and x^r are real-valued random variables.

From the definition of Aumann integral in [8], the expectation of interval-valued random variable $X = [\underline{x}, \overline{x}]$ can be written as $\mathrm{E}[X] = [\mathrm{E}\underline{x}, \mathrm{E}\overline{x}]$, see also [15].

Furthermore, call $X = [\underline{x}, \overline{x}]$ a constrained interval-valued Gaussian random variable, if random vector $(\underline{x}, \overline{x})^\tau$ follows the conditional distribution $f(\underline{x}, \overline{x})/\mathrm{P}(\underline{x} \le \overline{x})$, where $f(\cdot, \cdot)$ is a bivariate normal probability density function.

2.3 Previous Researches on Interval-Valued Linear Regression

Denote the interval-valued explanatory variables $X_1 = [\underline{x}_1, \overline{x}_1] = (x_1^c; x_1^r)$, $\cdots$, $X_p = [\underline{x}_p, \overline{x}_p] = (x_p^c; x_p^r)$, the interval-valued response variable $Y = [\underline{y}, \overline{y}] = (y^c; y^r)$, and corresponding observations $\{X_{i,1}, X_{i,2}, \cdots, X_{i,p}, Y_i\}_{i=1,2,\cdots,n}$. Let $\boldsymbol{y}^c = (y_1^c, y_2^c, \cdots, y_n^c)^\tau$, $\boldsymbol{y}^r = (y_1^r, y_2^r, \cdots, y_n^r)^\tau$, $\boldsymbol{X}^c = \left(x_{i,j}^c\right)_{i=1,2,\cdots,n, j=1,2,\cdots,p}$, $\boldsymbol{X}^r = \left(x_{i,j}^r\right)_{i=1,2,\cdots,n, j=1,2,\cdots,p}$. Assume $\boldsymbol{X}^c$ and $\boldsymbol{X}^r$ are of full rank, i.e., $\operatorname{rank}(\boldsymbol{X}^c) = \operatorname{rank}(\boldsymbol{X}^r) = p$.

1) Billard and Diday[17] proposed the MinMax method. They considered the model

$$\begin{cases} \underline{y} = \beta_{0,1} + \displaystyle\sum_{j=1}^{p} \beta_{j,1} \underline{x}_j + \varepsilon_1, \\ \overline{y} = \beta_{0,2} + \displaystyle\sum_{j=1}^{p} \beta_{j,2} \overline{x}_j + \varepsilon_2, \end{cases}$$

where $\beta_{0,1}, \beta_{1,1}, \cdots, \beta_{p,1}, \beta_{0,2}, \beta_{1,2} \cdots, \beta_{p,2} \in \mathbb{R}$ are the coefficients, $\varepsilon_1, \varepsilon_2$ are real-valued model errors. Minimizing the criterion function

$$\sum_{i=1}^{n} \left[\left(\underline{y}_i - \beta_{0,1} - \sum_{j=1}^{p} \beta_{j,1} \underline{x}_{i,j} \right)^2 + \left(\overline{y}_i - \beta_{0,2} - \sum_{j=1}^{p} \beta_{j,2} \overline{x}_{i,j} \right)^2 \right]$$

leads to the estimators $\widehat{\beta}_{0,1}, \widehat{\beta}_{1,1}, \cdots, \widehat{\beta}_{p,1}, \widehat{\beta}_{0,2}, \widehat{\beta}_{1,2} \cdots, \widehat{\beta}_{p,2}$. Furthermore, for any given interval-valued observations $X_{i,1}, X_{i,2}, \cdots, X_{i,p}$, the prediction is $\widehat{Y}_i = [\widehat{\underline{y}}_i, \widehat{\overline{y}}_i]$, where

$$\widehat{\underline{y}}_i = \widehat{\beta}_{0,1} + \sum_{j=1}^{p} \widehat{\beta}_{j,1} \underline{x}_{i,j}, \quad \widehat{\overline{y}}_i = \widehat{\beta}_{0,2} + \sum_{j=1}^{p} \widehat{\beta}_{j,2} \overline{x}_{i,j}, \quad i = 1, 2, \cdots, n.$$

In above model, the minimum of the response interval is assumed to depend only the minimums of the explanatory variables, and similar to the maximums. These assumptions might be too restrictive in applications. Thus, we consider a more general model in next section, in which the minimum (the maximum) of the response may depend not only the minimums but also the maximums of the explanatory variables.

2) Billard and Diday[18] proposed the CM method. They modeled the center as

$$y^c = \beta_0^c + \sum_{j=1}^{p} \beta_j^c x_j^c + \varepsilon^c,$$

where $\beta_0^c, \beta_1^c, \cdots, \beta_p^c \in \mathbb{R}$ are the coefficients, $\varepsilon^c \in \mathbb{R}$ is the model error. The least squares estimator of the coefficients is

$$\widehat{\boldsymbol{\beta}}^c = (\widehat{\beta}_0^c, \widehat{\beta}_1^c, \cdots, \widehat{\beta}_p^c)^\tau = ((\boldsymbol{X}^c)^\tau \boldsymbol{X}^c)^{-1} (\boldsymbol{X}^c)^\tau \boldsymbol{y}^c.$$

For any observations $X_{i,1}, X_{i,2}, \cdots, X_{i,p}$ of the explanatory variables, the prediction $\widehat{Y}_i = [\widehat{\underline{y}}_i, \widehat{\overline{y}}_i]$ is given by

$$\widehat{\underline{y}}_i = \widehat{\beta}_0^c + \sum_{j=1}^{p} \widehat{\beta}_j^c \underline{x}_{i,j}, \quad \widehat{\overline{y}}_i = \widehat{\beta}_0^c + \sum_{j=1}^{p} \widehat{\beta}_j^c \overline{x}_{i,j}, \quad i = 1, 2, \cdots, n.$$

3) Lima and Carvalho[19] proposed the CRM method, by which they assumed that

$$\begin{cases} y^c = \beta_0^c + \sum_{j=1}^{p} \beta_j^c x_j^c + \varepsilon^c, \\ y^r = \beta_0^r + \sum_{j=1}^{p} \beta_j^r x_j^r + \varepsilon^r, \end{cases}$$

where $\beta_0^c, \beta_1^c, \cdots, \beta_p^c, \beta_0^r, \beta_1^r, \cdots, \beta_p^r \in \mathbb{R}$ are the coefficients, $\varepsilon^c, \varepsilon^r \in \mathbb{R}$ are the model errors. The estimators are obtained as

$$\widehat{\boldsymbol{\beta}}^c = (\widehat{\beta}_0^c, \widehat{\beta}_1^c, \cdots, \widehat{\beta}_p^c)^\tau = ((\boldsymbol{X}^c)^\tau \boldsymbol{X}^c)^{-1} (\boldsymbol{X}^c)^\tau \boldsymbol{y}^c,$$
$$\widehat{\boldsymbol{\beta}}^r = (\widehat{\beta}_0^r, \widehat{\beta}_1^r, \cdots, \widehat{\beta}_p^r)^\tau = ((\boldsymbol{X}^r)^\tau \boldsymbol{X}^r)^{-1} (\boldsymbol{X}^r)^\tau \boldsymbol{y}^r.$$

For given $X_{i,1}, X_{i,2}, \cdots, X_{i,p}$, the prediction $\widehat{Y}_i = [\widehat{\underline{y}}_i, \widehat{\overline{y}}_i]$ is obtained from

$$\widehat{\underline{y}}_i = \widehat{y}_i^c - \widehat{y}_i^r, \quad \widehat{\overline{y}}_i = \widehat{y}_i^c + \widehat{y}_i^r,$$

where $\widehat{y}_i^c = \widehat{\beta}_0^c + \sum_{j=1}^{p} \widehat{\beta}_j^c x_{i,j}^c$, $\widehat{y}_i^r = \widehat{\beta}_0^c + \sum_{j=1}^{p} \widehat{\beta}_j^r x_{i,j}^r$, $i = 1, 2, \cdots, n$.

4) Lima and Carvalho[20] further proposed the CCRM method by employing the same model as the CRM, and putting the constrain $\beta_j^r \geq 0$, $j = 0, 1, \cdots, p$. They used the Lawson-Hanson algorithm[34] to solve the constrained least squares and get coefficient estimates.

Besides, Blanco-Fernández, et al.[21] extended this CCRM method by assuming

$$\begin{cases} y^c = \beta_0^c + \sum_{j=1}^{p} \beta_j^c x_j^c + \varepsilon^c, \\ y^r = |\beta_0^r| + \sum_{j=1}^{p} |\beta_j^r| x_j^r + \varepsilon^r. \end{cases}$$

Sun[22] proposed the following constrained model

$$\begin{cases} y^c = \alpha x^c + \gamma x^r + \eta + \varepsilon^c, \\ y^r = \beta x^r + \theta + \varepsilon^r, \end{cases}$$

where coefficients $\alpha, \gamma, \eta \in \mathbb{R}$, and $\beta \geq 0, \theta \geq 0$.

Note that all the constraint conditions of [20–22] operate on the coefficients of radii regression. On the other hand, Guo and Hao[23] considered the constraint condition $\beta^r x^r \geq 0$, and optimization method was used in estimation procedure.

3 Constrained Interval-Valued Linear Regression Model

In this section, we discuss the following constrained interval-valued linear regression model

$$\begin{cases} \underline{y} = \beta_{0,1} + \sum_{j=1}^{p} (\beta_{j,11} \underline{x}_j + \beta_{j,12} \overline{x}_j) + \varepsilon_1, \\ \overline{y} = \beta_{0,2} + \sum_{j=1}^{p} (\beta_{j,21} \underline{x}_j + \beta_{j,22} \overline{x}_j) + \varepsilon_2, \end{cases} \tag{1}$$

where $\beta_{0,i}, \beta_{j,kl} \in \mathbb{R}$, $i, k, l = 1, 2, j = 1, 2, \cdots, p$ are coefficients, $(\varepsilon_1, \varepsilon_2)^\tau$ is the model error vector which is independent of $\underline{x}_j$, $\overline{x}_j$, $j = 1, 2, \cdots, p$, and $\mathrm{E}\varepsilon_1 = \mathrm{E}\varepsilon_2 = 0$, $\mathrm{E}\varepsilon_1^2 < \infty$, $\mathrm{E}\varepsilon_2^2 < \infty$. Note that when $\underline{y} > \overline{y}$, interval Y does not exist. Thus, we impose the order constraint condition of $\underline{y} \leq \overline{y}$ in the regression. This is related to the classical truncated model discussed in [28] and [29].

For the convenience, we always take $\{X_{i,1}, X_{i,2}, \cdots, X_{i,p}\}_{i=1,2,\cdots,n}$ as given when the distributions of model errors and the response endpoints are used hereinafter, except other declaration is stated.

To simplify notation, denote

$$g_1 \triangleq g_1(\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_p, \overline{x}_1, \overline{x}_2, \cdots, \overline{x}_p) = \beta_{0,1} + \sum_{j=1}^{p}(\beta_{j,11}\underline{x}_j + \beta_{j,12}\overline{x}_j), \tag{2}$$

$$g_2 \triangleq g_2(\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_p, \overline{x}_1, \overline{x}_2, \cdots, \overline{x}_p) = \beta_{0,2} + \sum_{j=1}^{p}(\beta_{j,21}\underline{x}_j + \beta_{j,22}\overline{x}_j). \tag{3}$$

Without loss of generality, we further assume error vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)^\tau$ follows $N((0,0)^\tau, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

i.e.,

$$f(\varepsilon_1, \varepsilon_2) = \frac{1}{2\pi|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\boldsymbol{\varepsilon}^\tau \boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}\right\}. \tag{4}$$

Notice that condition $\underline{y} \leq \overline{y}$ is equivalent to $\varepsilon_2 - \varepsilon_1 \geq g_1 - g_2$, thus leads to a truncated normal distribution. Then from Nath[35] we have the following conditional expectations.

$$\mathrm{E}(\varepsilon_1|\underline{y} \leq \overline{y}) = C_1\lambda\left(\frac{g_2 - g_1}{\sigma}\right), \quad \mathrm{E}(\varepsilon_2|\underline{y} \leq \overline{y}) = C_2\lambda\left(\frac{g_2 - g_1}{\sigma}\right), \tag{5}$$

where $\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$, $C_1 = \frac{-\sigma_1^2 + \rho\sigma_1\sigma_2}{\sigma}$, $C_2 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma}$, and

$$\lambda(t) = \varphi(t)/\Phi(t), \tag{6}$$

where $\varphi(t) = (2\pi)^{-\frac{1}{2}}\exp\{-\frac{t^2}{2}\}$ is the probability density function of the standard normal distribution, and $\Phi(t) = \int_{-\infty}^{t}\varphi(x)\mathrm{d}x$ is its cumulative distribution function. Furthermore, we have

$$\mathrm{E}(\underline{y}|\underline{y} \leq \overline{y}) = g_1 + \mathrm{E}(\varepsilon_1|\underline{y} \leq \overline{y}) = g_1 + C_1\lambda\left(\frac{g_2 - g_1}{\sigma}\right), \tag{7}$$

$$\mathrm{E}(\overline{y}|\underline{y} \leq \overline{y}) = g_2 + \mathrm{E}(\varepsilon_2|\underline{y} \leq \overline{y}) = g_2 + C_2\lambda\left(\frac{g_2 - g_1}{\sigma}\right). \tag{8}$$

$$\begin{aligned}\mathrm{Var}(\underline{y}|\underline{y}\le\overline{y}) &= \mathrm{E}(\underline{y}^2|\underline{y}\le\overline{y}) - \mathrm{E}(\underline{y}|\underline{y}\le\overline{y})^2\\ &= \sigma_1^2 + C_1^2\lambda\left(\frac{g_2-g_1}{\sigma}\right)\left[\frac{g_2-g_1}{\sigma} - \lambda\left(\frac{g_2-g_1}{\sigma}\right)\right], \qquad (9)\\ \mathrm{Var}(\overline{y}|\underline{y}\le\overline{y}) &= \mathrm{E}(\overline{y}^2|\underline{y}\le\overline{y}) - \mathrm{E}(\overline{y}|\underline{y}\le\overline{y})^2\\ &= \sigma_2^2 + C_2^2\lambda\left(\frac{g_2-g_1}{\sigma}\right)\left[\frac{g_2-g_1}{\sigma} - \lambda\left(\frac{g_2-g_1}{\sigma}\right)\right]. \qquad (10)\end{aligned}$$

From (7)–(8), it is easy to see that if we ignore the constraint condition and use the least squares estimation method directly, the coefficient estimators would be biased. Furthermore, (9)–(10) shows that under the constraint condition, the conditional variances of both $\underline{y}$ and $\overline{y}$ have an increment depending on the explanatory variables through g_1 and g_2, and thus the heteroscedasticity occurs. As is well known in the linear regression, neglecting the heteroscedasticity makes the ordinary least squares estimator inefficient.

González-Rivera and Lin[6] proposed a two step estimation procedure, in which the constraint condition is used to solve the biased problem. To improve the efficiency of the two step estimation, we further propose a three step estimation procedure based on the two step estimation. Let $\{X_{i,1}, X_{i,2}, \cdots, X_{i,p}, Y_i\}_{i=1,2,\cdots,n}$ be i.i.d. interval-valued observations of Model (1).

3.1 The First Step Estimation

From Model (1), it holds that

$$\frac{1}{\sigma}y^R = \alpha_0 + \sum_{j=1}^{p}(\alpha_{j,1}\underline{x}_j + \alpha_{j,2}\overline{x}_j) + \frac{1}{\sigma}\varepsilon = \frac{1}{\sigma}(g_2-g_1) + \frac{1}{\sigma}\varepsilon, \quad y^R \ge 0, \qquad (11)$$

where $y^R = \overline{y} - \underline{y}$ is the range of interval Y, coefficients $\alpha_0 = (\beta_{0,2} - \beta_{0,1})/\sigma$, $\alpha_{j,1} = (\beta_{j,21} - \beta_{j,11})/\sigma$, $\alpha_{j,2} = (\beta_{j,22} - \beta_{j,12})/\sigma$, $j = 1, 2, \cdots, p$, model error $\varepsilon = \varepsilon_2 - \varepsilon_1 \sim N(0, \sigma^2)$, $\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$.

Then we can estimate the coefficients of Model (11) through conditional maximum likelihood method. Denote the parameter vector as

$$\boldsymbol{v}_0 = (\alpha_0, \alpha_{1,1}, \alpha_{1,2}, \cdots, \alpha_{p,1}, \alpha_{p,2}, \sigma)^\tau.$$

For interval-valued observations $\{X_{i,1}, X_{i,2}, \cdots, X_{i,p}, Y_i\}_{i=1,2,\cdots,n}$, the conditional log likelihood function is written as

$$\begin{aligned}L(\boldsymbol{v}_0|y_i^R \ge 0, i=1,2,\cdots,n) &= \sum_{i=1}^{n}\log\left[\varphi\left(\frac{1}{\sigma}\varepsilon_i\right)\Big/\Phi\left(\alpha_0 + \sum_{j=1}^{p}(\alpha_{j,1}\underline{x}_j + \alpha_{j,2}\overline{x}_j)\right)\right]\\ &= -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{n}\left[\frac{1}{\sigma}y_i^R - \alpha_0 - \sum_{j=1}^{p}(\alpha_{j,1}\underline{x}_j + \alpha_{j,2}\overline{x}_j)\right]^2\\ &\quad - \sum_{i=1}^{n}\log\Phi\left(\alpha_0 + \sum_{j=1}^{p}(\alpha_{j,1}\underline{x}_j + \alpha_{j,2}\overline{x}_j)\right). \qquad (12)\end{aligned}$$

Through maximizing (12), we can get the estimator $\widehat{\boldsymbol{v}}_0$. Amemiya[28] discussed the asymptotic properties of the conditional maximum likelihood estimator of truncated linear regression model. The asymptotic properties of $\widehat{\boldsymbol{v}}_0$ are given in the following theorem.

Theorem 3.1 (see [28]) *If further assume matrix* $\mathrm{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^{\tau}$ *exists and is nonsingular, where* $\boldsymbol{x}_i=(1,\underline{x}_{i,1},\overline{x}_{i,1},\cdots,\underline{x}_{i,p},\overline{x}_{i,p})^{\tau}$*, then the estimator* $\widehat{\boldsymbol{v}}_0$ *has the following properties.*

(i) $\widehat{\boldsymbol{v}}_0$ *converges to* $\boldsymbol{v}_0$ *in probability, i.e.,* $\widehat{\boldsymbol{v}}_0\xrightarrow{p}\boldsymbol{v}_0$.

(ii) $\widehat{\boldsymbol{v}}_0$ *is asymptotically normal, i.e.,*

$$\sqrt{n}(\widehat{\boldsymbol{v}}_0-\boldsymbol{v}_0)\to N(0,\boldsymbol{\Upsilon}),$$

where $\boldsymbol{\Upsilon}=-\mathrm{plim}_{n\to\infty}\left[\frac{1}{n}\mathrm{E}\left(\frac{\partial^2 L}{\partial\boldsymbol{v}_0\partial\boldsymbol{v}_0^{\tau}}\right)\right]^{-1}$, plim *stands for the limit converged to in probability.*

Orme[36] proved that the unique maximum likelihood estimator can be obtained, even though (12) is not global convex. González-Rivera and Lin[6] further studied the asymptotic properties of the conditional maximum likelihood estimator when data are autocorrelated.

Denote $\boldsymbol{v}_0^*=(\alpha_0,\alpha_{1,1},\alpha_{1,2},\cdots,\alpha_{p,1},\alpha_{p,2})^{\tau}$, which is a part of $\boldsymbol{v}_0$. Let $\widehat{\boldsymbol{v}}_0^*$ be the corresponding part of $\widehat{\boldsymbol{v}}_0$. Substituting $\boldsymbol{x}^{\tau}\widehat{\boldsymbol{v}}_0^*=\widehat{\alpha}_0+\sum_{j=1}^{p}(\widehat{\alpha}_{j,1}\underline{x}_j+\widehat{\alpha}_{j,2}\overline{x}_j)$ into (6) can calculate the estimator $\widehat{\lambda}\triangleq\lambda(\boldsymbol{x}^{\tau}\widehat{\boldsymbol{v}}_0^*)$. Since $\lambda(\cdot)$ is a continuously differentiable function of $\boldsymbol{v}_0^*$, we can deduce the asymptotic properties of $\lambda(\boldsymbol{x}^{\tau}\widehat{\boldsymbol{v}}_0^*)$ through Theorem 3.1.

Corollary 3.2 *Under the assumptions of Theorem* 3.1*, estimator* $\lambda(\boldsymbol{x}^{\tau}\widehat{\boldsymbol{v}}_0^*)$ *has the following properties.*

(i) $\lambda(\boldsymbol{x}^{\tau}\widehat{\boldsymbol{v}}_0^*)$ *converges to* $\lambda(\boldsymbol{x}^{\tau}\boldsymbol{v}_0^*)$ *in probability, i.e.,* $\lambda(\boldsymbol{x}^{\tau}\widehat{\boldsymbol{v}}_0^*)\xrightarrow{p}\lambda(\boldsymbol{x}^{\tau}\boldsymbol{v}_0^*)$.

(ii) $\lambda(\boldsymbol{x}^{\tau}\widehat{\boldsymbol{v}}_0^*)$ *is asymptotically normal, i.e.,*

$$\sqrt{n}(\lambda(\boldsymbol{x}^{\tau}\widehat{\boldsymbol{v}}_0^*)-\lambda(\boldsymbol{x}^{\tau}\boldsymbol{v}_0^*))\to N\left(0,\frac{\partial\lambda(\boldsymbol{x}^{\tau}\boldsymbol{v}_0^*)}{\partial(\boldsymbol{v}_0^*)^{\tau}}\boldsymbol{\Upsilon}^*\frac{\partial\lambda(\boldsymbol{x}^{\tau}\boldsymbol{v}_0^*)}{\partial(\boldsymbol{v}_0^*)}\right),$$

where $\boldsymbol{\Upsilon}^*$ *is the asymptotic covariance matrix of* $\sqrt{n}(\widehat{\boldsymbol{v}}_0^*-\boldsymbol{v}_0^*)$*, which is also a part of matrix* $\boldsymbol{\Upsilon}$*,* $\partial\lambda(\boldsymbol{x}^{\tau}\boldsymbol{v}_0^*)/\partial\boldsymbol{v}_0^*=[\lambda(\boldsymbol{x}^{\tau}\boldsymbol{v}_0^*)-\boldsymbol{x}^{\tau}\boldsymbol{v}_0^*]\lambda(\boldsymbol{x}^{\tau}\boldsymbol{v}_0^*)\boldsymbol{x}$ *is the first order partial derivatives of* $\lambda(\boldsymbol{x}^{\tau}\boldsymbol{v}_0^*)$.

Using continuous mapping theorem, we can prove (i) in Corollary 3.2. Next, using the Taylor expansion we have $\lambda(\boldsymbol{x}^{\tau}\widehat{\boldsymbol{v}}_0^*)-\lambda(\boldsymbol{x}^{\tau}\widehat{\boldsymbol{v}}_0^*)\approx\frac{\partial\lambda(\boldsymbol{x}^{\tau}\boldsymbol{v}_0^*)}{\partial(\boldsymbol{v}_0^*)^{\tau}}(\widehat{\boldsymbol{v}}_0^*-\boldsymbol{v}_0^*)$. Combining Delta method (e.g., Theorem 3.1 in the book [37]) and Theorem 3.1 we can prove (ii) in Corollary 3.2.

3.2 The Second Step Estimation

Denote

$$\begin{aligned}&u_{1,i}=\underline{y}_i-\mathrm{E}(\underline{y}_i|\underline{y}_i\le\overline{y}_i),\quad u_{2,i}=\overline{y}_i-\mathrm{E}(\overline{y}_i|\underline{y}_i\le\overline{y}_i),\\&\lambda_i=\lambda(\boldsymbol{x}_i^{\tau}\boldsymbol{v}_0^*),\qquad\widehat{\lambda}_i=\lambda(\boldsymbol{x}_i^{\tau}\widehat{\boldsymbol{v}}_0^*),\quad i=1,2,\cdots,n,\end{aligned}$$

and

$$v_{1,i}=C_1(\lambda_i-\widehat{\lambda}_i),\qquad v_{2,i}=C_2(\lambda_i-\widehat{\lambda}_i),$$

then $\mathrm{E}(u_{1,i}|\underline{y}_i\le\overline{y}_i)=\mathrm{E}(u_{2,i}|\underline{y}_i\le\overline{y}_i)=0$, $\mathrm{E}v_{1,i}=\mathrm{E}v_{2,i}\approx 0$. Combining (7) and (8), Model (1) can be written as

$$\begin{cases}\underline{y}_i=\beta_{0,1}+\sum\limits_{j=1}^{p}(\beta_{j,11}\underline{x}_{i,j}+\beta_{j,12}\overline{x}_{i,j})+C_1\widehat{\lambda}_i+u_{1,i}+v_{1,i},\\\overline{y}_i=\beta_{0,2}+\sum\limits_{j=1}^{p}(\beta_{j,21}\underline{x}_{i,j}+\beta_{j,22}\overline{x}_{i,j})+C_2\widehat{\lambda}_i+u_{2,i}+v_{2,i},\end{cases}\quad i=1,2,\cdots,n.\tag{13}$$

Introduce the matrix notations:

$$\boldsymbol{Y}_{n\times 2}=(\underline{\boldsymbol{y}},\overline{\boldsymbol{y}})=\begin{pmatrix}\underline{y}_1 & \underline{y}_2 & \cdots & \underline{y}_n\\ \overline{y}_1 & \overline{y}_2 & \cdots & \overline{y}_n\end{pmatrix}^{\tau},$$

$$\boldsymbol{X}_{n\times(2p+1)}=\begin{pmatrix}\boldsymbol{x}_1\\ \vdots\\ \boldsymbol{x}_n\end{pmatrix}=\begin{pmatrix}1 & \underline{x}_{1,1} & \overline{x}_{1,1} & \underline{x}_{1,2} & \overline{x}_{1,2} & \cdots & \underline{x}_{1,p} & \overline{x}_{1,p}\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & \underline{x}_{n,1} & \overline{x}_{n,1} & \underline{x}_{n,2} & \overline{x}_{n,2} & \cdots & \underline{x}_{n,p} & \overline{x}_{n,p}\end{pmatrix},$$

$$\boldsymbol{\varLambda}=(\lambda_1,\lambda_2,\cdots,\lambda_n)^{\tau},\quad \widehat{\boldsymbol{\varLambda}}=(\widehat{\lambda}_1,\widehat{\lambda}_2,\cdots,\widehat{\lambda}_n)^{\tau},$$

$$\widehat{\boldsymbol{Z}}_{n\times 2(p+1)}=(\boldsymbol{X},\widehat{\boldsymbol{\varLambda}})\triangleq(\widehat{\boldsymbol{z}}_1,\widehat{\boldsymbol{z}}_2,\cdots,\widehat{\boldsymbol{z}}_n)^{\tau},\quad \boldsymbol{Z}_{n\times 2(p+1)}=(\boldsymbol{X},\boldsymbol{\varLambda})\triangleq(\boldsymbol{z}_1,\boldsymbol{z}_2,\cdots,\boldsymbol{z}_n)^{\tau},$$

$$\boldsymbol{B}_{2(p+1)\times 2}=(\boldsymbol{b}_1,\boldsymbol{b}_2)=\begin{pmatrix}\beta_{0,1} & \beta_{1,11} & \beta_{1,12} & \cdots & \beta_{p,11} & \beta_{p,12} & C_1\\ \beta_{0,2} & \beta_{1,21} & \beta_{1,22} & \cdots & \beta_{p,21} & \beta_{p,22} & C_2\end{pmatrix}^{\tau},$$

$$\boldsymbol{U}_{n\times 2}=(\boldsymbol{u}_1,\boldsymbol{u}_2)=\begin{pmatrix}u_{1,1} & u_{1,2} & \cdots & u_{1,n}\\ u_{2,1} & u_{2,2} & \cdots & u_{2,n}\end{pmatrix}^{\tau},\quad \boldsymbol{V}_{n\times 2}=(\boldsymbol{v}_1,\boldsymbol{v}_2)=\begin{pmatrix}v_{1,1} & v_{1,2} & \cdots & v_{1,n}\\ v_{2,1} & v_{2,2} & \cdots & v_{2,n}\end{pmatrix}^{\tau},$$

the matrix form of Model (13) can be written as

$$\mathrm{Vec}(\boldsymbol{Y})=(I_2\otimes\widehat{\boldsymbol{Z}})\mathrm{Vec}(\boldsymbol{B})+\mathrm{Vec}(\boldsymbol{U})+\mathrm{Vec}(\boldsymbol{V}),\tag{14}$$

where $\mathrm{Vec}(\cdot)$ is the vectorization of matrix, I_2 is the 2×2 identity matrix, $\otimes$ is the Kronecker product.

The least squares estimator of Model (14) is

$$\mathrm{Vec}(\widehat{\boldsymbol{B}})=[(I_2\otimes\widehat{\boldsymbol{Z}}^{\tau})(I_2\otimes\widehat{\boldsymbol{Z}})]^{-1}(I_2\otimes\widehat{\boldsymbol{Z}}^{\tau})\mathrm{Vec}(\boldsymbol{Y}),\tag{15}$$

Estimator (15) also can be written as

$$\widehat{\boldsymbol{B}}=(\widehat{\boldsymbol{Z}}^{\tau}\widehat{\boldsymbol{Z}})^{-1}\widehat{\boldsymbol{Z}}^{\tau}\boldsymbol{Y},\tag{16}$$

or $\widehat{\boldsymbol{b}}_1=(\widehat{\boldsymbol{Z}}^{\tau}\widehat{\boldsymbol{Z}})^{-1}\widehat{\boldsymbol{Z}}^{\tau}\underline{\boldsymbol{y}}$, $\widehat{\boldsymbol{b}}_2=(\widehat{\boldsymbol{Z}}^{\tau}\widehat{\boldsymbol{Z}})^{-1}\widehat{\boldsymbol{Z}}^{\tau}\overline{\boldsymbol{y}}$.

González-Rivera and Lin[6] proved the asymptotic properties of estimators $\widehat{\boldsymbol{b}}_1$, $\widehat{\boldsymbol{b}}_2$ for different models. We will present the asymptotic properties of the estimator $\mathrm{Vec}(\widehat{\boldsymbol{B}})$ through the theory of multivariate statistics.

Theorem 3.3 *Assume* $\mathrm{plim}_{n\to\infty}\frac{1}{n}I_2\otimes(\boldsymbol{Z}^{\tau}\boldsymbol{Z})=\boldsymbol{Q}^{-1}$ *is nonsingular, and* $\mathrm{plim}_{n\to\infty}\frac{1}{n}\big((I_2\otimes\boldsymbol{Z}^{\tau})\mathrm{Cov}(\mathrm{Vec}(\boldsymbol{U}+\boldsymbol{V}))(I_2\otimes\boldsymbol{Z})\big)=\varXi$. *Then the estimator* $\mathrm{Vec}(\widehat{\boldsymbol{B}})$ *has the following properties.*

(i) $\mathrm{Vec}(\widehat{\boldsymbol{B}})$ *converges to* $\mathrm{Vec}(\boldsymbol{B})$ *in probability, i.e.,* $\mathrm{Vec}(\widehat{\boldsymbol{B}})\xrightarrow{p}\mathrm{Vec}(\boldsymbol{B})$.

(ii) $\mathrm{Vec}(\widehat{\boldsymbol{B}})$ *is asymptotically normal, i.e.,*

$$\sqrt{n}(\mathrm{Vec}(\widehat{\boldsymbol{B}})-\mathrm{Vec}(\boldsymbol{B}))\to N\big(0,\boldsymbol{Q}\varXi\boldsymbol{Q}^{\tau}\big).$$

From Corollary 3.2 it is easy to see that $\frac{1}{n}\widehat{\boldsymbol{Z}}^{\tau}\widehat{\boldsymbol{Z}}-\frac{1}{n}\boldsymbol{Z}^{\tau}\boldsymbol{Z}\xrightarrow{p}0$. Then using

$$\sqrt{n}(\mathrm{Vec}(\widehat{\boldsymbol{B}})-\mathrm{Vec}(\boldsymbol{B}))=\left[\frac{1}{n}(I_2\otimes\widehat{\boldsymbol{Z}}^{\tau})(I_2\otimes\widehat{\boldsymbol{Z}})\right]^{-1}\left[\frac{1}{\sqrt{n}}(I_2\otimes\widehat{\boldsymbol{Z}}^{\tau})(\mathrm{Vec}(\boldsymbol{U})+\mathrm{Vec}(\boldsymbol{V}))\right],$$

Theorem 3.3 can be proved.

3.3 The Third Step Estimation

From last subsection, we see that the second step estimator utilizes the restriction-imposed model (13) or (14) and thus is consistent. As is shown by Equations (9) and (10), the restriction-imposed model is of heteroscedasticity. To improve the second step estimator, we first find a reasonable estimate of $\boldsymbol{\Sigma}_1 = \mathrm{Cov}(\mathrm{Vec}(\boldsymbol{U}) + \mathrm{Vec}(\boldsymbol{V}))$.

According to Corollary 3.2, we have

$$\mathrm{Cov}(\mathrm{Vec}(\boldsymbol{V})) = \begin{pmatrix} C_1^2 & C_1C_2 \\ C_1C_2 & C_2^2 \end{pmatrix} \otimes \mathrm{Cov}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}),$$

where the (i,j)-th element of $\mathrm{Cov}(\widehat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda})$ is approximately

$$\frac{1}{n}\lambda_i\lambda_j \cdot (\lambda_i - \boldsymbol{x}_i^\tau \boldsymbol{v}_0^*)(\lambda_j - \boldsymbol{x}_j^\tau \boldsymbol{v}_0^*)\boldsymbol{x}_i^\tau \boldsymbol{\Upsilon}^* \boldsymbol{x}_j,$$

which would vanish as n goes to infinity. On the other hand, from Nath[35] it holds that

$$\begin{aligned}
&\mathrm{E}(\varepsilon_{1,i}^2 | \underline{y}_i \leq \overline{y}_i) = \sigma_1^2 + C_1^2 \boldsymbol{x}_i^\tau \boldsymbol{v}_0^* \lambda_i, \\
&\mathrm{E}(\varepsilon_{2,i}^2 | \underline{y}_i \leq \overline{y}_i) = \sigma_2^2 + C_2^2 \boldsymbol{x}_i^\tau \boldsymbol{v}_0^* \lambda_i, \\
&\mathrm{E}(\varepsilon_{1,i}\varepsilon_{2,i} | \underline{y}_i \leq \overline{y}_i) = \rho\sigma_1\sigma_2 + C_1C_2 \boldsymbol{x}_i^\tau \boldsymbol{v}_0^* \lambda_i,
\end{aligned}$$

which leads to that, given $\underline{y}_i \leq \overline{y}_i$,

$$\begin{aligned}
&\mathrm{Var}(u_{1,i}) = \sigma_1^2 + C_1^2 \lambda_i \cdot (\boldsymbol{x}_i^\tau \boldsymbol{v}_0^* - \lambda_i), \\
&\mathrm{Var}(u_{2,i}) = \sigma_2^2 + C_2^2 \lambda_i \cdot (\boldsymbol{x}_i^\tau \boldsymbol{v}_0^* - \lambda_i), \\
&\mathrm{Cov}(u_{1,i}, u_{2,i}) = \rho\sigma_1\sigma_2 + C_1C_2 \lambda_i \cdot (\boldsymbol{x}_i^\tau \boldsymbol{v}_0^* - \lambda_i).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\boldsymbol{\Sigma}_1 &= \begin{pmatrix} \mathrm{Cov}(\boldsymbol{u}_1 + \boldsymbol{v}_1) & \mathrm{Cov}(\boldsymbol{u}_1 + \boldsymbol{v}_1, \boldsymbol{u}_2 + \boldsymbol{v}_2) \\ \mathrm{Cov}(\boldsymbol{u}_1 + \boldsymbol{v}_1, \boldsymbol{u}_2 + \boldsymbol{v}_2) & \mathrm{Cov}(\boldsymbol{u}_2 + \boldsymbol{v}_2) \end{pmatrix} \\
&\approx \begin{pmatrix} \mathrm{Cov}(\boldsymbol{u}_1) & \mathrm{Cov}(\boldsymbol{u}_1, \boldsymbol{u}_2) \\ \mathrm{Cov}(\boldsymbol{u}_1, \boldsymbol{u}_2) & \mathrm{Cov}(\boldsymbol{u}_2) \end{pmatrix},
\end{aligned} \tag{17}$$

where

$$\mathrm{Cov}(\boldsymbol{u}_j) = \mathrm{diag}\big(\mathrm{Var}(u_{j,i})\big), \quad j = 1, 2; \quad \mathrm{Cov}(\boldsymbol{u}_1, \boldsymbol{u}_2) = \mathrm{diag}\big(\mathrm{Cov}(u_{1,i}, u_{2,i})\big)$$

are diagonal matrices.

Denote the residual matrix by

$$\widehat{\boldsymbol{W}}_{n\times 2} = \begin{pmatrix} \widehat{w}_{1,1} & \widehat{w}_{1,2} & \cdots & \widehat{w}_{1,n} \\ \widehat{w}_{2,1} & \widehat{w}_{2,2} & \cdots & \widehat{w}_{2,n} \end{pmatrix}^\tau = \boldsymbol{Y} - \widehat{\boldsymbol{Z}}\widehat{\boldsymbol{B}}.$$

Intuitive estimates of σ_1, σ_2 and ρ are given by

$$\begin{aligned}
\widehat{\sigma}_1^2 &= \frac{1}{n}\sum_{i=1}^{n}\big[\widehat{w}_{1,i}^2 - \widehat{C}_1^2\widehat{\lambda}_i \cdot (\boldsymbol{x}_i^\tau \widehat{\boldsymbol{v}}_0^* - \widehat{\lambda}_i)\big],\\
\widehat{\sigma}_2^2 &= \frac{1}{n}\sum_{i=1}^{n}\big[\widehat{w}_{2,i}^2 - \widehat{C}_2^2\widehat{\lambda}_i \cdot (\boldsymbol{x}_i^\tau \widehat{\boldsymbol{v}}_0^* - \widehat{\lambda}_i)\big],\\
\widehat{\rho} &= \frac{1}{n}\sum_{i=1}^{n}\big[\widehat{w}_{1,i}\widehat{w}_{2,i} - \widehat{C}_1\widehat{C}_2\widehat{\lambda}_i \cdot (\boldsymbol{x}_i^\tau \widehat{\boldsymbol{v}}_0^* - \widehat{\lambda}_i)\big]/(\widehat{\sigma}_1 \cdot \widehat{\sigma}_2),
\end{aligned} \tag{18}$$

where $\widehat{C}_1$ and $\widehat{C}_2$ are the estimates included in estimator (16). Substituting estimates (18) into (17) results in an estimate of $\boldsymbol{\Sigma}_1$, denoted by $\widehat{\boldsymbol{\Sigma}}_1$.

The weighted least squares estimator of Model (14) is

$$\mathrm{Vec}(\widetilde{\boldsymbol{B}}) = [(I_2 \otimes \widehat{\boldsymbol{Z}}^\tau)\widehat{\boldsymbol{\Sigma}}_1^{-1}(I_2 \otimes \widehat{\boldsymbol{Z}})]^{-1}(I_2 \otimes \widehat{\boldsymbol{Z}}^\tau)\widehat{\boldsymbol{\Sigma}}_1^{-1}\mathrm{Vec}(\boldsymbol{Y}). \tag{19}$$

It has the following properties.

Theorem 3.4 *Under the assumptions of Theorem* 3.3*, estimator* $\mathrm{Vec}(\widetilde{\boldsymbol{B}})$ *has the following properties.*

(i) $\mathrm{Vec}(\widetilde{\boldsymbol{B}})$ *converges to* $\mathrm{Vec}(\boldsymbol{B})$ *in probability, i.e.,* $\mathrm{Vec}(\widetilde{\boldsymbol{B}}) \xrightarrow{p} \mathrm{Vec}(\boldsymbol{B})$.

(ii) $\mathrm{Vec}(\widetilde{\boldsymbol{B}})$ *is asymptotically normal, i.e.,*

$$\sqrt{n}(\mathrm{Vec}(\widetilde{\boldsymbol{B}}) - \mathrm{Vec}(\boldsymbol{B})) \to N(0, \Xi^{-1}).$$

Theorem 3.4 can be proved employing the equation

$$\sqrt{n}(\mathrm{Vec}(\widetilde{\boldsymbol{B}}) - \mathrm{Vec}(\boldsymbol{B})) = \left[\frac{1}{n}(I_2 \otimes \widehat{\boldsymbol{Z}}^\tau)\widehat{\boldsymbol{\Sigma}}_1^{-1}(I_2 \otimes \widehat{\boldsymbol{Z}})\right]^{-1}\left[\frac{1}{\sqrt{n}}(I_2 \otimes \widehat{\boldsymbol{Z}}^\tau)\widehat{\boldsymbol{\Sigma}}_1^{-1} \cdot (\mathrm{Vec}(\boldsymbol{U}) + \mathrm{Vec}(\boldsymbol{V}))\right],$$

Theorems 3.1 and 3.3. Furthermore, we have the following corollary.

Corollary 3.5 *The estimator* $\widetilde{\boldsymbol{B}}$ *is more efficient than* $\widehat{\boldsymbol{B}}$*, that is, the difference of the asymptotic covariance matrix of* $\mathrm{Vec}(\widehat{\boldsymbol{B}})$ *from that of* $\mathrm{Vec}(\widetilde{\boldsymbol{B}})$ *is a positive semi-definite matrix.*

The conclusion is from the following calculations

$$\begin{aligned}
&\mathrm{Cov}(\mathrm{Vec}(\widehat{\boldsymbol{B}})) - \mathrm{Cov}(\mathrm{Vec}(\widetilde{\boldsymbol{B}}))\\
\approx&(I_2 \otimes (\boldsymbol{Z}^\tau\boldsymbol{Z})^{-1})(I_2 \otimes \boldsymbol{Z}^\tau) \cdot \big[\boldsymbol{\Sigma}_1 - (I_2 \otimes \boldsymbol{Z})\big((I_2 \otimes \boldsymbol{Z}^\tau)\boldsymbol{\Sigma}_1^{-1}(I_2 \otimes \boldsymbol{Z})\big)^{-1}(I_2 \otimes \boldsymbol{Z}^\tau)\big]\\
&\cdot (I_2 \otimes \boldsymbol{Z})(I_2 \otimes (\boldsymbol{Z}^\tau\boldsymbol{Z})^{-1})\\
=&(I_2 \otimes (\boldsymbol{Z}^\tau\boldsymbol{Z})^{-1})(I_2 \otimes \boldsymbol{Z}^\tau)\boldsymbol{\Sigma}_1^{\frac{1}{2}}[I_{2n} - \widetilde{\boldsymbol{Z}}(\widetilde{\boldsymbol{Z}}^\tau\widetilde{\boldsymbol{Z}})^{-1}\widetilde{\boldsymbol{Z}}^\tau] \cdot \boldsymbol{\Sigma}_1^{\frac{1}{2}}(I_2 \otimes \boldsymbol{Z})(I_2 \otimes (\boldsymbol{Z}^\tau\boldsymbol{Z})^{-1}),
\end{aligned}$$

in which the right hand side is positive semi-definite, and so is the left hand side when n is large enough, where $\widetilde{\boldsymbol{Z}} \triangleq \boldsymbol{\Sigma}_1^{\frac{1}{2}}(I_2 \otimes \boldsymbol{Z})$.

Remark Although $\widetilde{\boldsymbol{B}}$ is an efficient estimator theoretically, there might be a numerical problem for small and moderate sample sizes. Note that $\lambda(t)$ is a decreasing function, and as $(g_2 - g_1)/\sigma$ is increasing, $\lambda(\frac{g_2-g_1}{\sigma}) \to 0$. Small $\widehat{\lambda}$s will lead to the design matrix $\widehat{\boldsymbol{Z}}_{n\times 2(p+1)} =$

$(\boldsymbol{X}, \widehat{\boldsymbol{\Lambda}})$ being not approximately full rank, and hence make the estimates of both βs and C_is unstable numerically.

To overcome above problem, we consider to substitute the estimate $\widehat{\sigma}$ obtained from the first step, $\widehat{\sigma}_1$, $\widehat{\sigma}_2$ and $\widehat{\rho}$ in (18) into expressions $C_1 = \frac{-\sigma_1^2+\rho\sigma_1\sigma_2}{\sigma}$, $C_2 = \frac{\sigma_2^2-\rho\sigma_1\sigma_2}{\sigma}$ to obtain new estimates of C_1 and C_2. Denote

$$\widetilde{C}_1 = \frac{-\widehat{\sigma}_1^2 + \widehat{\rho}\widehat{\sigma}_1\widehat{\sigma}_2}{\widehat{\sigma}}, \quad \widetilde{C}_2 = \frac{\widehat{\sigma}_2^2 - \widehat{\rho}\widehat{\sigma}_1\widehat{\sigma}_2}{\widehat{\sigma}}, \tag{20}$$

$$\widetilde{\boldsymbol{C}}_{n\times 2} = (\widetilde{C}_1\widehat{\boldsymbol{\Lambda}}, \widetilde{C}_2\widehat{\boldsymbol{\Lambda}}) = \begin{pmatrix} \widetilde{C}_1\widehat{\lambda}_1 & \widetilde{C}_1\widehat{\lambda}_2 & \cdots & \widetilde{C}_1\widehat{\lambda}_n \\ \widetilde{C}_2\widehat{\lambda}_1 & \widetilde{C}_2\widehat{\lambda}_2 & \cdots & \widetilde{C}_2\widehat{\lambda}_n \end{pmatrix}^\tau,$$

$$\boldsymbol{\beta}_{(2p+1)\times 2} = \begin{pmatrix} \beta_{0,1} & \beta_{1,11} & \beta_{1,12} & \cdots & \beta_{p,11} & \beta_{p,12} \\ \beta_{0,2} & \beta_{1,21} & \beta_{1,22} & \cdots & \beta_{p,21} & \beta_{p,22} \end{pmatrix}^\tau.$$

Then we adjust the bias by subtracting $\widetilde{\boldsymbol{C}}$ from $\boldsymbol{Y}$, and define an adjusted third step estimator of the regression coefficients by

$$\mathrm{Vec}(\widetilde{\boldsymbol{\beta}}) = [(I_2 \otimes \boldsymbol{X}^\tau)\widehat{\boldsymbol{\Sigma}}_1^{-1}(I_2 \otimes \boldsymbol{X})]^{-1}(I_2 \otimes \boldsymbol{X}^\tau)\widehat{\boldsymbol{\Sigma}}_1^{-1}\mathrm{Vec}(\boldsymbol{Y} - \widetilde{\boldsymbol{C}}). \tag{21}$$

Theoretically, the adjusted third step estimator shares same asymptotic properties of the corresponding part of $\widetilde{\boldsymbol{B}}$, but it performs much better according to our computing experience. In simulation study we will show the efficiency of estimators (20), (21).

4 Numerical Studies

4.1 Simulation

In this subsection, we verify the effectiveness of the proposed estimation method through simulation studies.

Data are generated as follows. Explanatory variable observation $X_i = (x_i^c; x_i^r)$, $x_i^c \sim N(1, 2^2)$, $x_i^r \sim Ga(0.8, 1)$, $i = 1, 2, \cdots, n$, where Ga stands for Gamma distribution. Model error vector $(\varepsilon_{1,i}, \varepsilon_{2,i})^\tau \sim N((0,0)^\tau, \Sigma)$, $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. Response variable observation $Y_i = [\underline{y}_i, \overline{y}_i]$ follows

$$\begin{cases} \underline{y}_i = \beta_{10} + \beta_{11}\underline{x}_i + \beta_{12}\overline{x}_i + \varepsilon_{1,i}, \\ \overline{y}_i = \beta_{20} + \beta_{21}\underline{x}_i + \beta_{22}\overline{x}_i + \varepsilon_{2,i}. \end{cases}$$

In the process of generating Y_i, $i = 1, 2, \cdots, n$, we can borrow the idea of truncation model. We generate $X_i = (x_i^c; x_i^r)$ and $(\varepsilon_{1,i}, \varepsilon_{2,i})^\tau$ as above. If the obtained $\underline{y}_i^*, \overline{y}_i^*$ satisfies $\underline{y}_i^* > \overline{y}_i^*$, we abandon this group of sample. Otherwise, the observation is reserved and write $Y_i = [\underline{y}_i^*, \overline{y}_i^*]$. Repeat this process until n observations are collected.

Consider the following cases:

Case 1 $\beta_{10} = -2$, $\beta_{11} = 0.9$, $\beta_{12} = 2$, $\beta_{20} = 0.5$, $\beta_{21} = 0.5$, $\beta_{2,2} = 0.9$, $\sigma_1 = 1.5$, $\sigma_2 = 1.5$, $\rho = 0.5$;

Case 2 $\beta_{10}=0$, $\beta_{11}=0.9$, $\beta_{12}=0.5$, $\beta_{20}=0$, $\beta_{21}=0.5$, $\beta_{2,2}=0.9$, $\sigma_1=1.5$, $\sigma_2=1.5$, $\rho=0.5$;

Case 3 $\beta_{10}=0$, $\beta_{11}=0.9$, $\beta_{12}=0.5$, $\beta_{20}=0$, $\beta_{21}=0.5$, $\beta_{2,2}=0.9$, $\sigma_1=2$, $\sigma_2=1.5$, $\rho=0.5$;

Case 4 $\beta_{10}=0$, $\beta_{11}=0.9$, $\beta_{12}=0.5$, $\beta_{20}=0$, $\beta_{21}=0.5$, $\beta_{2,2}=0.9$, $\sigma_1=2.5$, $\sigma_2=1.5$, $\rho=0.5$.

Realizations of above four cases are shown in Figure 1. Note that the shapes become more dynamic from Case 1 to Case 4. We introduce the criterion (see Cribari-Neto and Lima[38])

$$\varpi=\frac{\max_i\{\mathrm{Var}(u_{1,i})\}}{\min_i\{\mathrm{Var}(u_{1,i})\}}=\frac{\max_i\{\sigma_1^2+C_1^2\lambda_i[(g_{2,i}-g_{1,i})/\sigma-\lambda_i]\}}{\min_i\{\sigma_1^2+C_1^2\lambda_i[(g_{2,i}-g_{1,i})/\sigma-\lambda_i]\}},$$

to measure the heteroscedasticity. Under homoscedasticity $\varpi=1$, and under heteroscedasticity $\varpi>1$. Table 1 displays the means of λ and ϖ, in above cases $n=200,1000$ with 1000 repetitions. The values of ϖ show that the strength of heteroscedasticity is increasing from Case 1 to Case 4.

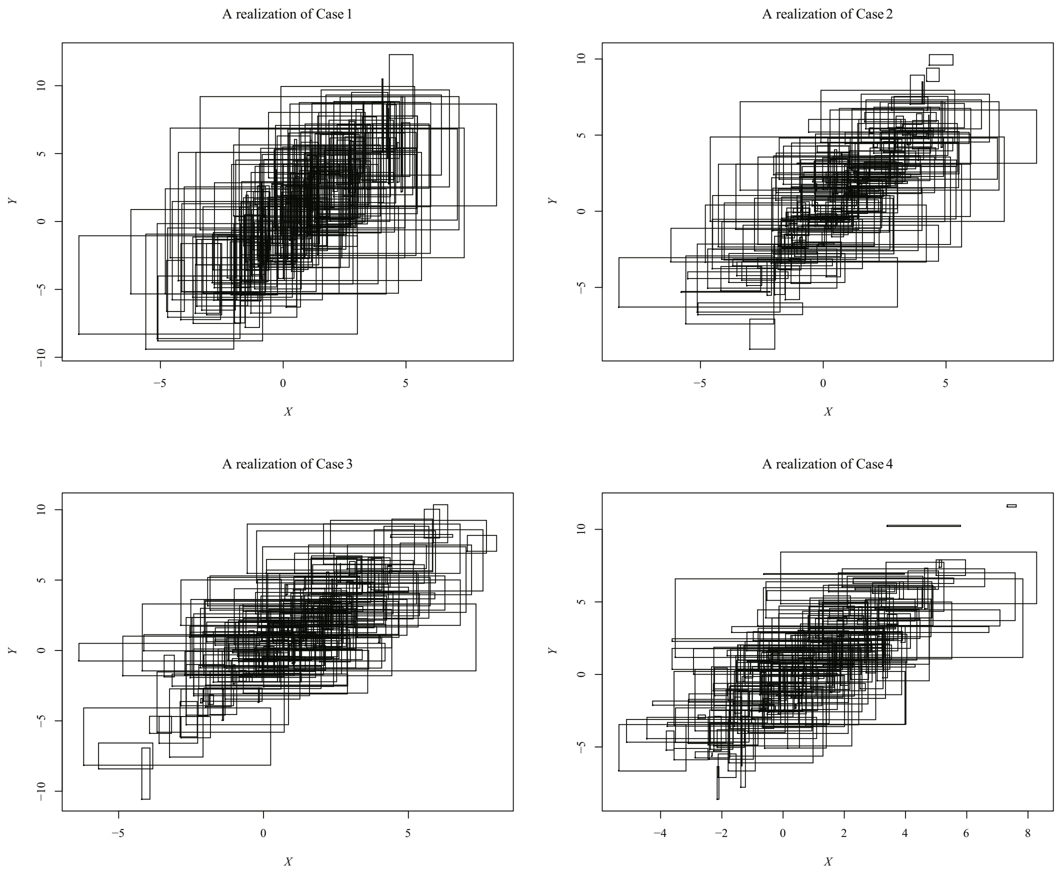

Figure 1 Realizations of Cases 1–4, $n=200$

Table 1 The mean values of λ and ϖ

		Case 1	Case 2	Case 3	Case 4
$n = 200$	λ	0.0055	0.5388	0.5744	0.6078
	ϖ	2.2578	2.7076	4.6471	6.9289
$n = 1000$	λ	0.0054	0.5385	0.5744	0.6082
	ϖ	2.2587	2.8504	4.8891	7.1604

Table 2 Variances and MSEs of coefficients in Case 1 and Case 2

coef	$n = 200$				$n = 1000$			
	Two Step		Three Step		Two Step		Three Step	
	Var	MSE	Var	MSE	Var	MSE	Var	MSE
				Case 1				
β_{10}	0.2113	0.2114	0.0238	0.0238	0.0385	0.0385	0.0043	0.0043
β_{11}	0.0128	0.0128	0.0045	0.0045	0.0024	0.0024	0.0008	0.0008
β_{12}	0.0131	0.0131	0.0044	0.0044	0.0024	0.0024	0.0008	0.0008
C_1	4910.664	4912.055	0.0095	0.0096	560.3382	560.4934	0.0020	0.0020
β_{20}	0.2058	0.2063	0.0231	0.0232	0.0381	0.0381	0.0048	0.0048
β_{21}	0.0123	0.0123	0.0041	0.0041	0.0024	0.0024	0.0008	0.0008
β_{22}	0.0126	0.0126	0.0042	0.0042	0.0023	0.0023	0.0008	0.0008
C_2	4432.905	4437.006	0.0106	0.0106	574.5513	574.6673	0.0019	0.0020
σ_1^2	0.0507	0.0536	0.0510	0.0523	0.0099	0.0104	0.0100	0.0102
σ_2^2	0.0487	0.0532	0.0492	0.0505	0.0102	0.0116	0.0104	0.0106
ρ	0.0033	0.0034	0.0031	0.0031	0.0006	0.0007	0.0006	0.0006
				Case 2				
β_{10}	2.5095	2.5108	0.0990	0.1003	0.3044	0.3045	0.0160	0.0162
β_{11}	0.0494	0.0494	0.0057	0.0058	0.0065	0.0065	0.0010	0.0010
β_{12}	0.0483	0.0483	0.0062	0.0062	0.0064	0.0064	0.0009	0.0009
C_1	4.7392	4.7435	0.0222	0.0225	0.5532	0.5533	0.0041	0.0044
β_{20}	2.4467	2.4470	0.0456	0.0469	0.3311	0.3323	0.0094	0.0096
β_{21}	0.0494	0.0495	0.0040	0.0041	0.0071	0.0071	0.0009	0.0009
β_{22}	0.0490	0.0490	0.0043	0.0043	0.0069	0.0070	0.0008	0.0008
C_2	4.6187	4.6188	0.0088	0.0102	0.5964	0.5983	0.0013	0.0017
σ_1^2	0.0586	0.1387	0.0402	0.1332	0.0088	0.0807	0.0074	0.0805
σ_2^2	0.1641	0.1994	0.0388	0.1339	0.0206	0.0651	0.0076	0.0699
ρ	0.0082	0.0106	0.0075	0.0101	0.0017	0.0051	0.0016	0.0052

For each case with sample sizes $n = 200, 1000$, repeat 1000 times. The obtained empirical variances and mean squared errors of the estimators are listed in Tables 2 and 3, where "Two

Step" represents the two step estimation method, and "Three Step" represents the three step estimation method. In the third step, we use the adjusted estimators (20) and (21).

Table 3 Variances and MSEs of coefficients in Case 3 and Case 4

coef	$n=200$				$n=1000$			
	Two Step		Three Step		Two Step		Three Step	
	Var	MSE	Var	MSE	Var	MSE	Var	MSE
				Case 3				
β_{10}	10.1431	10.1474	0.1294	0.4013	0.9127	0.9144	0.0215	0.2917
β_{11}	0.1123	0.1124	0.0086	0.0127	0.0149	0.0150	0.0013	0.0055
β_{12}	0.1096	0.1096	0.0094	0.0139	0.0150	0.0150	0.0015	0.0056
C_1	28.9029	28.9234	0.0316	0.5330	1.5511	1.5529	0.0058	0.4986
β_{20}	7.7357	7.7430	0.0937	0.3552	0.6223	0.6241	0.0171	0.2995
β_{21}	0.0771	0.0772	0.0079	0.0126	0.0098	0.0098	0.0011	0.0052
β_{22}	0.0759	0.0759	0.0077	0.0118	0.0098	0.0098	0.0011	0.0053
C_2	21.3443	21.3664	0.0134	0.4732	1.1284	1.1312	0.0019	0.4901
σ_1^2	0.1856	1.2077	0.0884	1.1648	0.0226	1.0271	0.0191	1.0587
σ_2^2	0.2857	0.2862	0.0495	0.0591	0.0288	0.0293	0.0088	0.0183
ρ	0.0109	0.0211	0.0093	0.0205	0.0021	0.0156	0.0019	0.0151
				Case 4				
β_{10}	41.2569	41.3446	0.1721	1.1583	2.8529	2.8530	0.0296	1.0548
β_{11}	0.2993	0.2995	0.0124	0.0252	0.0355	0.0355	0.0020	0.0140
β_{12}	0.2793	0.2798	0.0125	0.0244	0.0350	0.0350	0.0022	0.0140
C_1	82.9252	82.9982	0.0514	1.8212	4.8394	4.8397	0.0083	1.7613
β_{20}	20.6290	20.6914	0.2416	1.3097	1.3366	1.3372	0.0312	1.0788
β_{21}	0.1362	0.1367	0.0178	0.0322	0.0160	0.0160	0.0021	0.0150
β_{22}	0.1412	0.1419	0.0170	0.0313	0.0156	0.0157	0.0020	0.0152
C_2	40.1221	40.1884	0.0383	1.7553	2.3089	2.3103	0.0025	1.7346
σ_1^2	1.0952	5.6554	0.1816	5.1745	0.0437	4.9133	0.0349	4.9614
σ_2^2	0.5340	0.5532	0.0607	0.0609	0.0418	0.0445	0.0097	0.0098
ρ	0.0237	0.0728	0.0239	0.0788	0.0031	0.0577	0.0028	0.0589

From the values of ϖ in Table 1 we know that the heteroscedasticity happens in all above cases. Through comparing the results in Table 2 and Table 3, we can find out that three step estimators have smaller variances and MSEs. This shows the proposed three step estimation method is more efficient.

Notice that the two step estimators of C_1 and C_2 are very poor in Case 1. This may due to $g_{1,i} \ll g_{2,i}$, which leads to $\lambda_i \approx 0$. Table 1 also verify the values of λ in Case 1 are very small. Besides, the variances and MSEs of other estimators in Case 1 are smaller than those in

Cases 2–4. This may be because of the lower strength of heteroscedasticity in Case 1.

4.2 An Application: Air Quality Index Data

In this subsection, we study the linear relationship between the concentration of fine particulate matter ($PM_{2.5}$) and nitrogen dioxide (NO_2) in the air through the constrained interval-valued linear regression model. Data are collected from 1586 air monitoring stations in China, 2017.10.1, and downloaded from beijingair.sinaapp.com. Take the minimums and maximums of $PM_{2.5}$ and NO_2 which are collected by each air monitoring station in this day as interval-valued observations. After deleting the stations with missing values we have 1481 interval-valued observations. Let $PM_{2.5}$ be the response variable, NO_2 be the explanatory variable, and plot them in Figure 2. For the convenience of calculation, we divide the observations of $PM_{2.5}$ and NO_2 by 10. Furthermore, assume the distances among these air monitoring stations are far enough so that they can be considered as independent.

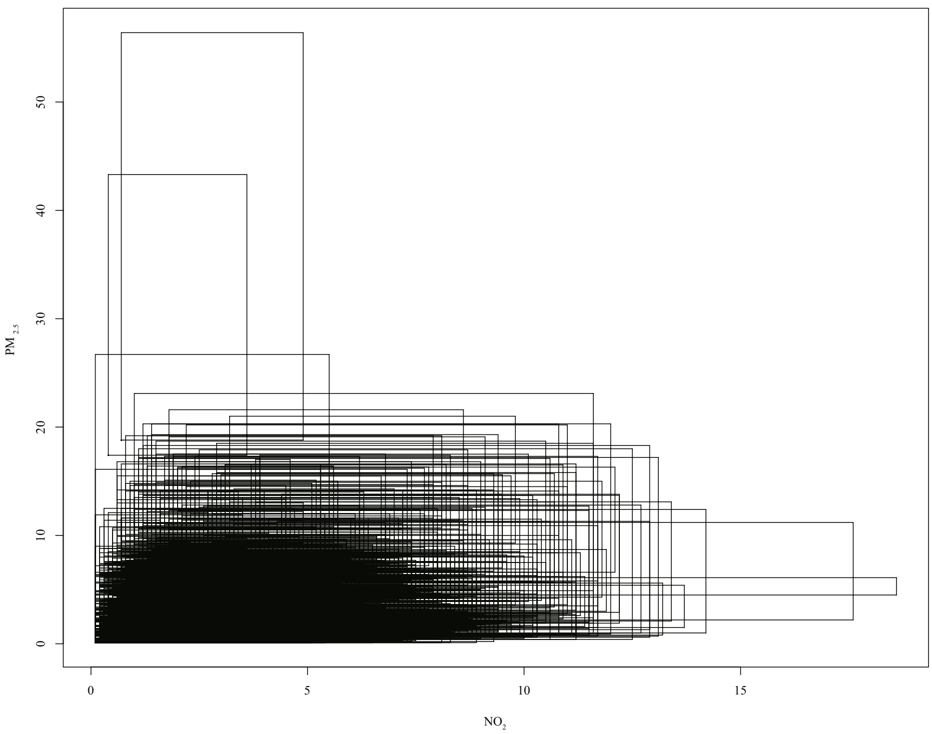

Figure 2 The air quality index data

Divide the data into two groups randomly, one is the training group and the other is the testing group. The training group is applied to estimate the parameters of models, which contains 1331 samples (about 90%). The testing group is used to predict, and the sample size is 150 (about 10%). We use CM, CRM, CCRM methods and constrained interval-valued linear regression model to fit the training data respectively. Then we calculate the predictions with the testing data.

In order to make a comparison, some evaluation criteria are introduced below. Denote $\widehat{Y}_i = [\underline{\widehat{y}}_i, \overline{\widehat{y}}_i]$ as the prediction of $Y_i = [\underline{y}_i, \overline{y}_i]$, $i = 1, 2, \cdots, 150$. The square root of mean square error of lower bound (RMSE_L) and the square root of mean square error of upper bound (RMSE_U) are

$$\text{RMSE}_\text{L} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\underline{\widehat{y}}_i - \underline{y}_i)^2}, \quad \text{RMSE}_\text{U} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\overline{\widehat{y}}_i - \overline{y}_i)^2}.$$

The mean square error for interval-valued data (IMSE) is

$$\text{IMSE} = \frac{1}{n}\sum_{i=1}^{n} d_2^2(\widehat{Y}_i, Y_i) = \frac{1}{n}\sum_{i=1}^{n}\left[(\underline{\widehat{y}}_i - \underline{y}_i)^2 + (\overline{\widehat{y}}_i - \overline{y}_i)^2\right].$$

The mean relative error of interval-valued data (IMRE) is

$$\text{IMRE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i^c - \widehat{y}_i^c|}{y_i^r + \widehat{y}_i^r} = \frac{1}{n}\sum_{i=1}^{n}\frac{|\overline{y}_i - \overline{\widehat{y}}_i + \underline{y}_i - \underline{\widehat{y}}_i|}{\overline{y}_i + \overline{\widehat{y}}_i - \underline{y}_i - \underline{\widehat{y}}_i}.$$

The effective coverage rate (ECR) is

$$\text{ECR} = \frac{1}{n}\sum_{i=1}^{n}\frac{\text{wid}(\widehat{Y}_i \cap Y_i)}{\text{wid}(\widehat{Y}_i)},$$

where $\cap$ is the intersection of two intervals, wid is the length of interval. Notice that the denominators of criteria ECR and IMRE may be equal to 0. Therefore, in this paper, 0.001 is added to the denominators of these two criteria. If the values of RMSE_L, RMSE_U, IMSE and IMRE are close to 0, and the value of ECR is close to 1, the predicted values are more accurate.

The criteria are calculated and listed in Table 4. We find out that three step predictions have the smallest RMSE_L, RMSE_U, IMSE, IMRE and the largest ECR. This shows the efficiency of three step estimation method for constrained interval-valued linear regression model in prediction.

Table 4 Results of different methods

	RMSE_L	RMSE_U	IMSE	IMRE	ECR
CRM	0.9104	2.9410	9.4786	0.3404	0.5958
CCRM	0.9104	2.9410	9.4786	0.3404	0.5958
Two Step	0.8743	2.9534	9.4870	0.3438	0.6053
Three Step	**0.8523**	**2.9405**	**9.3729**	**0.3403**	**0.6180**

5 Conclusion

This paper basically solved the problem of heteroscedasticity in [6] by proposing a three step estimation procedure. Simulations and an application show the proposed three step estimation can reduce the variances and MSEs of parameter estimators effectively, which means the proposed method has certain advancement. Besides, several problems such as the significance tests of the constrained regression equation and coefficients, variable selection method, regression diagnostics etc deserve to study.

References

[1] Wang X, Li S, and Denœux T, Interval-valued linear model, *International Journal of Computational Intelligence Systems*, 2015, **8**(1): 114–127.

[2] Gil M A, Lubiano M A, Montenegro M, et al., Least squares fitting of an affine function and strength of association for interval-valued data, *Metrika*, 2002, **56**(2): 97–111.

[3] Gil M A, Colubi A, and Montenegro M, Testing linear independence in linear models with interval-valued data, *Computational Statistics and Data Analysis*, 2007, **51**(6): 3002–3015.

[4] Fagundes R, Souza R, and Cysneiros F, Interval kernel regression, *Neurocomputing*, 2014, **128**(5): 371–388.

[5] Engle R F and Gallo G M, A multiple indicators model for volatility using intra-daily data, *Journal of Econometrics*, 2006, **131**(1): 3–27.

[6] González-Rivera G and Lin W, Constrained regression for interval-valued data, *Journal of Business and Economic Statistics*, 2013, **31**(4): 473–490.

[7] Wang X, Zhang Z, and Li S, Set-valued and interval-valued stationary time series, *Journal of Multivariate Analysis*, 2016, **145**: 208–223.

[8] Aumann R J, Integrals of set-valued functions, *Journal of Mathematical Analysis and Applications*, 1965, **12**(1): 1–12.

[9] Puri M L and Ralescu D A, A strong law of large numbers for Banach space valued random sets, *The Annals of Probability*, 1983, **11**: 222–224.

[10] Vitale R A, L_p metrics for compact, convex sets, *Journal of Approximation Theory*, 1985, **45**: 280–287.

[11] Aubin J P and Frankowska H, *Set-Valued Analysis*, Birkhauser, Basle, Switzerland, 1990.

[12] Hu S and Papageorgiou N S, *Handbook of Multivalued Analysis*, Kluwer Academic Publishers, Boston, MA, 1997.

[13] Li S and Ogura Y, Convergence of set valued sub- and super-martingales in the Kuratowski-Mosco sense, *The Annals of Probability*, 1998, **26**(3): 1384–1402.

[14] Li S and Ogura Y, Convergence of set valued and fuzzy valued martingales, *Fuzzy Sets and Systems*, 1999, **101**: 453–461.

[15] Li S, Ogura Y, and Kreinovich V, *Limit Theorems and Applications of Set-Valued and Fuzzy Set-Valued Random Variables*, Kluwer Academic Publishers (Now Springer), Dordrecht, 2002.

[16] Yang X and Li S, The D_p-metric space of set-valued random variables and its application to covariances, *International Journal of Innovative Computing, Information and Control*, 2005, **1**: 73–82.

[17] Billard L and Diday E, Symbolic regression analysis, *Classification Data Analysis and Knowledge Organization, Proceedings of the Eighteenth Conference of the International Federation of Classification Societies* (*IFCS*'02), Springer, 2002, 281–288.

[18] Billard L and Diday E, Regression analysis for interval-valued data, *Data Analysis, Classification*

and Related Methods, Proceedings of the Seventh Conference of the International Federation of Classification Societies (*IFCS*'00), Springer, 2000, 369–374.

[19] Lima N E and Carvalho F, Centre and range method for fitting a linear regression model to symbolic interval data, *Computational Statistics and Data Analysis*, 2008, **52**(3): 1500–1515.

[20] Lima N E and Carvalho F, Constrained linear regression models for symbolic interval-valued variables, *Computational Statistics and Data Analysis*, 2010, **54**(2): 333–347.

[21] Blanco-Fernández A, Corral N, and González-Rodríguez G, Estimation of a flexible simple linear model for interval data based on set arithmetic, *Computational Statistics and Data Analysis*, 2011, **55**(9): 2568–2578.

[22] Sun Y, Linear regression with interval-valued data, *Wiley Interdisciplinary Reviews: Computational Statistics*, 2016, **8**(1): 54–60.

[23] Guo J and Hao P, Constrained center and range joint model for interval-valued symbolic data regression, *Computational Statistics and Data Analysis*, 2017, **116**: 106–138.

[24] Wang H, Guan R, and Wu J, CIPCA: Complete-information-based principal component analysis for interval-valued data, *Neurocomputing*, 2012, **86**: 158–169.

[25] Wang H, Guan R, and Wu J, Linear regression of interval-valued data based on complete information in hypercubes, *Journal of Systems Science and Systems Engineering*, 2012, **21**(4): 422–442.

[26] Souza L, Souza R, Amaral G, et al., A parametrized approach for linear regression of interval data, *Knowledge-Based Systems*, 2017, **131**: 149–159.

[27] Hsu H and Wu B, Evaluating forecasting performance for interval data, *Computers and Mathematics with Applications*, 2008, **56**(9): 2155–2163.

[28] Amemiya T, Regression analysis when the dependent variable is truncated normal, *Econometrica*, 1973, **41**(6): 997–1016.

[29] Amemiya T, *Advanced Econometrics*, Harvard University Press, Cambridge, 1985.

[30] Heckman J, The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models, *Annals of Economic and Social Measurement*, 1976, **5**(4): 475–492.

[31] Heckman J, Dummy endogenous variables in a simultaneous equation system, *Econometrica*, 1978, **46**(4): 931–959.

[32] Heckman J, Sample selection bias as a specification error, *Econometrica*, 1979, **47**(1): 153–161.

[33] Li F, Li S, Tang N, et al., Constrained interval-valued linear regression model, *Proceedings of the 20th International Conference on Information Fusion*, 2017, No. 8009676, DOI: 10.23919/ICIF.2017.8009676.

[34] Lawson C and Hanson R, *Solving Least Squares Problems*, Prentice-Hall, Upper Saddle River, 1974.

[35] Nath G, Moments of a linearly truncated bivariate normal distribution, *Australian Journal of Statistics*, 1972, **14**: 97–102.

[36] Orme C, On the uniqueness of the maximum likelihood estimator in truncated regression models, *Econometric Reviews*, 1989, **8**(2): 217–222.

[37] van der Vaart A, *Asymptotic Statistics*, Cambridge University Press, Cambridge, 1998.

[38] Cribari-Neto F and Lima M, Sequence of improved standard errors under heteroskedasticity of unknown form, *Journal of Statistical Planning and Inference*, 2011, **141**(11): 3617–3627.