ICEB 2012 Proceedings

International Conference on Electronic Business (ICEB)

Fall 10-12-2012

Uncertain Regression Modeling Given the Observational Distributions

Renkuan Guo

Uncertain Regression Modeling Given the Observational Distributions

Renkuan Guo
University of Cape Town, Private Bag, Rondebosch 7701, Cape Town, South Africa

Abstract: In regression theory, the distribution of the error terms occupies a critical position, particularly when switching the data environment from probability theory to uncertainty theory. On the probabilistic platform, the variance-covariance matrix for standard regression model is assumed by an identity matrix with a positive constant multiplier. On the uncertain measure foundation, for given observational distributions, the variance-covariance is an interval-valued matrix. In this paper, we derive the interval-valued variance for given uncertain normal distribution. Further, we derive the interval-valued auto variance matrix for the observational error terms being the members of an uncertain canonical process. This new model may be regarded as an extension to the uncertain canonical process regression models, but its interval-valued variance-covariance matrix is also intrinsic to the uncertain canonical process, which results in an interval-valued weighted regression model.

Keywords: Uncertain measure, uncertain normal distribution, uncertain canonical process, auto covariance matrix, weighted regression model

1. Introduction

Linear regression models [1], [3], [4], [5], [12], [14] are the most familiar statistical models. It is also well-known that in probability theory the Gaussian distributional theory plays a fundamental role and facilitates linear regression modeling.

Linear regression models can work by switching the platform from probability measure into the uncertain measure. The fundamental problem in switching working environment is necessary to define and calculate the variance-covariance matrix of the observations.

Dai et al [2] derived the upper bound of the auto covariance matrix of the uncertain canonical process and Guo et al [6] further developed the upper bound of the auto covariance matrix of the uncertain integral driven process. Based on those foundational efforts in [2] and [6], Guo et al [4] developed the scalar uncertain canonical regression models accordingly.

Nevertheless, we have to aware that Guo's [4] uncertain canonical regression models are very special since the auto covariance matrix with members being scalar numbers by taking the upper limits. Therefore the uncertain canonical process regression is not in its general form, because the uncertain observations could only facilitate interval-valued variance logically.

To a better understanding the nature of working environment switching, let us review Liu's uncertainty theory which was founded in 2007 [7] and refined in [8], [9]. Nowadays uncertainty theory has become a branch of mathematics with many decent applications.

The core concept in uncertainty theory is the uncertain measure, which is a set function defined on a sigma-algebra generated from a non-empty set. Formally, let Ξ be a nonempty set (space), and $\mathfrak{A}(\Xi)$ the σ-algebra on Ξ. Each element, let us say, $A \subset \Xi$, $A \in \mathfrak{A}(\Xi)$ is called an uncertain event. A number denoted as $\bar{\lambda}\{A\}$, $0 \le \bar{\lambda}\{A\} \le 1$, is assigned to event $A \in \mathfrak{A}(\Xi)$, which indicates the uncertain measuring grade with which event $A \in \mathfrak{A}(\Xi)$ occurs. The

Proceedings of the Twelfth International Conference on Electronic Business, Xi'an, China, October 12-16, 2012, 232-241.

normal set function $\mathfrak{M}\{A\}$ satisfies following axioms given by Liu [9].

Axiom 1: (Normality) $\mathfrak{M}\{\Xi\}=1$.

Axiom 2: (Self-Duality) $\mathfrak{M}\{\cdot\}$ is self-dual, i.e., for any $A\in\mathfrak{A}(\Xi)$, $\mathfrak{M}\{A\}+\mathfrak{M}\{A^c\}=1$.

Axiom 3: (σ- Subadditivity) $\mathfrak{M}\left\{\bigcup_{i=1}^{\infty}A_i\right\}\le\sum_{i=1}^{\infty}\mathfrak{M}\{A_i\}$ for any countable event sequence $\{A_i\}$.

Definition I.1: (Liu, [7], [8], [9]) A set function $\mathfrak{M}:\mathfrak{A}(\Xi)\to[0,1]$ satisfies Axioms 1-3 is called an uncertain measure. The triple $(\Xi,\mathfrak{A}(\Xi),\mathfrak{M})$ is called an uncertainty space.

Definition I.2: (Liu, [7], [8], [9]) An uncertainty variable is a measurable function ξ from an uncertainty space $(\Xi,\mathfrak{A}(\Xi),\mathfrak{M})$ to the set of real numbers, i.e., for any Borel set B of real numbers, the set $\{\tau\in\Xi:\xi(\tau)\subset B\in\mathfrak{B}(\mathbb{R})\}\in\mathfrak{A}(\Xi)$, i.e., the pre-image of B is an event within Ξ.

Definition I.3: (Liu, [7], [8], [9]) The uncertain distribution $\Lambda:\mathbb{R}\to[0,1]$ of an uncertain variable ξ on $(\Xi,\mathfrak{A}(\Xi),\mathfrak{M})$ is

$$\Lambda(x)=\mathfrak{M}\{\tau\in\Xi\,|\,\xi(\tau)\le x\}\tag{1}$$

An uncertain variable is completely specified by its uncertain measure, while an uncertain variable is only partially defined by the corresponding uncertain distribution. In other words, using an uncertain distribution to specify an uncertain variable, certain uncertainty will be brought in.

In this paper, we intend to extend the uncertain canonical regression models developed by Guo et al [4] from scalar version into interval-valued version. The interval-valued models do not need to estimate the weight matrix as in fitting the classical weighted regression models using data replications. Rather, the interval weight matrix is intrinsic to the uncertain canonical process without data involvement of any kind.

The structure of the paper are as follows: Section II reviews the classical regression model theory. In Section III, we will derive the interval-valued variance for a given uncertain normal distribution. Section IV contributes the development of the intrinsic interval-valued auto covariance matrix of a standard uncertain canonical process. In Section V, we develop an interval-valued version of uncertain canonical process regression model. The last section offers a few concluding remarks.

2. A Review of Basic Regression Model

In statistical linear regression theory, [3], [12], [14], the basic assumptions are:

Assumption 1: The model takes a form:

$$y_i=\beta_0+\beta_1x_{1,i}+\ldots+\beta_{p,i}x_{p,i}+\varepsilon_i,\ i=1,2,\cdots,n\tag{2}$$

Assumption 2: The error terms satisfy several conditions:

(1) zero mean

$$\mathrm{E}[\varepsilon_i]=0,\ i=1,2,\cdots,n\tag{3}$$

(2) constant variance (homoscedasticity)

$$V[\varepsilon_i]=\sigma^2,\ i=1,2,\cdots,n\tag{4}$$

(3) mutually uncorrelated

$$\mathrm{E}\left[\varepsilon_i\varepsilon_j\right]=0,\ i\ne j,\ i,j=1,2,\cdots,n\tag{5}$$

Assumption 3: $x_1,x_2,\cdots,x_p$ are not random variables. They are fixed values of explanatory variables, with

$$\mathrm{E}[y\,|\,\underline{x}]=\underline{x}^t\underline{\beta}.\tag{6}$$

Assumption 4: $\varepsilon_i\overset{d}{\sim}N(0,\sigma^2),\ i=1,2,\cdots,n$.

In matrix presentation, Eq. (36) can be written as

$$\underline{y}=X\underline{\beta}+\underline{\varepsilon}\tag{7}$$

where the design or regressor matrix X is given by

$$X_{n\times(p+1)}=\begin{bmatrix}1 & x_{11} & \cdots & x_{1p}\\ 1 & x_{21} & \cdots & x_{2p}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & x_{n1} & \cdots & x_{np}\end{bmatrix}=\begin{bmatrix}\underline{1} & \underline{x}_1 & \cdots & \underline{x}_p\end{bmatrix}$$

$$\text{and }\ \underline{\beta}=\begin{bmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_p\end{bmatrix},\quad \underline{\varepsilon}=\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{bmatrix},\quad \underline{y}=\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{bmatrix}. \tag{8}$$

It is obvious that the regression model variance-covariance is $\sigma^2 I_{n\times n}$. The ordinary least square method will give the model estimation and analysis. The Gauss-Markov theorem guarantees the BLUE property [3], [13] of the coefficient estimators. If the regression model variance-covariance is not $\sigma^2 I_{n\times n}$ but $\sigma^2 V_{n\times n}$, then the weighted least square estimation is necessary. But the estimation of model variance-covariance matrix $V_{n\times n}$ requires sampling replications at each specific explanatory variable combination, which requires the model analyst to plan and collect a large sample. Nevertheless, Gaussian process regression, [12], [14], is a Bayesian regression model which take the intrinsic variance-possible remedy, whose variance-covariance matrix uses the prior covariance of the members of a delicate-chosen Gaussian process.

3. Interval Variance of an Uncertain Normal Distribution

Differently from a probability distribution, which fully specifies every measurable event, the uncertainty distribution can only characterize events and equivalently, the events $\{\xi > z\}$ for any $z\in\mathbb{R}$. In uncertainty theory, only uncertain measure, rather than uncertain distribution, can fully specify an arbitrary measurable event.

Definition III.1: (Liu, [7], [8], [9]) The uncertainty variables $\xi_1,\xi_2,\cdots,\xi_n$ on $(\Xi,\mathfrak{A},\lambda)$ are said to be independent if

$$\lambda\left\{\bigcap_{i=1}^{n}\{\xi_i\in B_i\}\right\}=\min_{1\le i\le n}\lambda\{\xi_i\in B_i\}$$

for any Borel sets $B_1,B_2,\cdots,B_n\in\mathfrak{B}(\mathbb{R})$

Theorem III.2: (Liu, [7], [8], [9]) Let $\Psi_{\xi_1},\Psi_{\xi_2},\cdots,\Psi_{\xi_n}$ be uncertainty distributions for the univariate uncertainty variables $\xi_1,\xi_2,\cdots,\xi_n$ on $(\Xi,\mathfrak{A}(\Xi),\lambda)$ respectively. Let $\Psi_{(\xi_1,\xi_2,\cdots,\xi_n)}$ be the joint distribution of uncertainty vector $(\xi_1,\xi_2,\cdots,\xi_n)$. If $\xi_1,\xi_2,\cdots,\xi_n$ are independent, then

$$\Psi_{(\xi_1,\xi_2,\cdots,\xi_n)}(x_1,x_2,\cdots,x_n)=\min_{1\le i\le n}\Psi_{\xi_i}(x_i) \tag{9}$$

for any real numbers $x_1,x_2,\cdots,x_n\in\mathbb{R}$

Definition III.3: (Liu, [7], [8], [9]) An uncertain variable ξ is called normal if the uncertain distribution takes the form

$$\Psi_{\xi}(x)=\frac{1}{1+e^{-\frac{\pi}{\sqrt{3}\sigma}(x-\mu)}},\quad x\in\mathbb{R} \tag{10}$$

Definition III.4: (Liu, [7], [8], [9]) An uncertain normal distribution is standard if the uncertain distribution takes the form

$$\Psi_{\xi_0}(x)=\frac{1}{1+e^{-\frac{\pi}{\sqrt{3}}x}},\quad x\in\mathbb{R} \tag{11}$$

It is obvious that an uncertain normal variable ξ can be expressed by an standard uncertain normal variable ξ_0

$$\xi=\sigma\xi_0+\mu \tag{12}$$

Definition III.5: (Liu, [7], [8], [9]) Let ξ be a uncertainty variable on an uncertainty measure space $(\Xi,\mathfrak{A}(\Xi),\lambda)$. The expectation ξ is defined by

$$\mathrm{E}[\xi]=\int_0^{+\infty}\lambda\{\xi>r\}\,dr-\int_{-\infty}^{0}\lambda\{\xi\le r\}\,dr \tag{13}$$

provided that at least one of the two integrals exists.

Theorem III.6: (Liu, [9]) Let ξ be an uncertain variable with a given uncertain distribution function Ψ_ξ having a finite expectation μ. Then the upper bound of the variance of ξ is

$$\sigma_{\xi,u}^2 = 2\int_0^{+\infty} \left(r\left(1-\Psi_\xi\left(r+\mu\right)\right) + \left(\Psi_\xi\left(\mu-r\right)\right)\right) dr \tag{14}$$

and the lower bound of the variance of ξ is

$$\sigma_{\xi,l}^2 = \int_0^{+\infty} \left(1-\Psi_\xi\left(\mu+\sqrt{x}\right)\right) dx \vee \int_0^{+\infty} \Psi_\xi\left(\mu-\sqrt{x}\right) dx \tag{15}$$

Proof:

$$\begin{aligned}
&\mathrm{E}\left[(\xi-\mu)^2\right] \\
&= \int_0^{+\infty} \lambda\left\{(\xi-\mu)^2 \ge x\right\} dx \\
&= \int_0^{+\infty} \lambda\left\{\left\{\xi \ge \mu+\sqrt{x}\right\} \cup \left\{\xi \ge \mu-\sqrt{x}\right\}\right\} dx \\
&\le \int_0^{+\infty} \left(\lambda\left\{\xi \ge \mu+\sqrt{x}\right\} + \lambda\left\{\xi \ge \mu-\sqrt{x}\right\}\right) dx \\
&= \int_0^{+\infty} \left(\left(1-\Psi_\xi\left(\mu+\sqrt{x}\right)\right) + \left(\Psi_\xi\left(\mu-\sqrt{x}\right)\right)\right) dx \\
&= 2\int_\mu^{+\infty} \left(\left(r-\mu\right)\left(1-\Psi_\xi\left(r\right)\right) + \left(\Psi_\xi\left(2\mu-r\right)\right)\right) dr \\
&= 2\int_0^{+\infty} s\left(1-\Psi_\xi\left(s+\mu\right)\right) ds + 2\int_0^{+\infty} s\Psi_\xi\left(\mu-s\right) ds \\
&= \sigma_{\xi,u}^2
\end{aligned} \tag{16}$$

By changing variable $\sqrt{x} = r$. On the other hand,

$$\begin{aligned}
&\mathrm{E}\left[(\xi-\mu)^2\right] \\
&= \int_0^{+\infty} \lambda\left\{(\xi-\mu)^2 \ge x\right\} dx \\
&= \int_0^{+\infty} \lambda\left\{\left\{\xi \ge \mu+\sqrt{x}\right\} \cup \left\{\xi \ge \mu-\sqrt{x}\right\}\right\} dx \\
&\ge \int_0^{+\infty} \left(\lambda\left\{\xi \ge \mu+\sqrt{x}\right\} dx + \right) dx \vee \int_0^{+\infty} \lambda\left\{\xi \ge \mu-\sqrt{x}\right\} dx \\
&= \int_0^{+\infty} \left(1-\Psi_\xi\left(\mu+\sqrt{x}\right)\right) dx \vee \int_0^{+\infty} \Psi_\xi\left(\mu-\sqrt{x}\right) dx \\
&= 2\int_0^{+\infty} r\left(1-\Psi_\xi\left(r+\mu\right)\right) dr \vee 2\int_0^{+\infty} r\Psi_\xi\left(\mu-r\right) dr \\
&= \sigma_{\xi,l}^2
\end{aligned} \tag{17}$$

Definition III.7: The ratio $\sigma_{\xi,l}^2 / \sigma_{\xi,u}^2$, denoted by κ, is called as the spread ratio of an uncertain variable, i.e.,

$$\kappa = \frac{\sigma_{\xi,l}^2}{\sigma_{\xi,u}^2} = \frac{\int_0^{+\infty} s\left(1-\Psi_\xi\left(s+\mu\right)\right) ds \vee \int_0^{+\infty} s\Psi_\xi\left(\mu-s\right) ds}{\int_0^{+\infty} s\left(1-\Psi_\xi\left(s+\mu\right)\right) ds + \int_0^{+\infty} s\Psi_\xi\left(\mu-s\right) ds} \tag{18}$$

Remark III.8: The variance of an uncertain variable given the uncertain distribution function, denoted by $V[\xi]$, is no longer a scalar positive real-valued number. The spread ratio, κ, represents the uncertainty degree in the variance of an uncertain variable given the distribution function. It is obvious that the spread ratio

$$\kappa \in \left[\frac{1}{2}, 1\right] \tag{19}$$

It is also worthwhile to comment that according to Moore [13], an interval number contains interval uncertainty. For example, interval number $\vec{\alpha} = [1,2]$ means any element belonging to this interval $\alpha \in [1,2]$ may be the representative of $\vec{\alpha}$ with a certain risk level. Thus κ represents the risk in the variance of an uncertain variable given the distribution only.

Theorem III.9: The variance of an uncertain variable given the uncertain distribution function, $V[\xi]$, is an interval-value:

$$\sigma_{\xi,l}^2 = \kappa\sigma_{\xi,u}^2 \le V[\xi] \le \sigma_{\xi,u}^2 \tag{20}$$

Proof: The variance of an uncertain variable

$$V[\xi] = \mathrm{E}\left[(\xi-\mu)^2\right] \tag{21}$$

and according to **Definition III.7**,

$$\kappa = \frac{\sigma_{\xi,l}^2}{\sigma_{\xi,u}^2}$$

Therefore

$$\sigma_{\xi,l}^2 = \kappa\sigma_{\xi,u}^2$$

Combining the results of **Theorem III.6**, we can conclude that:

$$\sigma_{\xi,l}^2 = \kappa\sigma_{\xi,u}^2 \le V[\xi] \le \sigma_{\xi,u}^2.$$

Theorem III.10: Let ξ be an uncertain normal variable defined by **Definition III.3**, Eq.(11), then

$$\kappa = \frac{1}{2}. \tag{22}$$

Proof:

$$\begin{aligned}
&\mathrm{E}\left[(\xi-\mu)^2\right] \\
&\le \int_0^{+\infty}\left(\left(1-\Psi_\xi\left(\mu+\sqrt{x}\right)\right)+\left(\Psi_\xi\left(\mu-\sqrt{x}\right)\right)\right)dx \\
&= \int_0^{+\infty}\left(1-\Psi_\xi\left(\mu+\sqrt{x}\right)\right)dx + \int_0^{+\infty}\Psi_\xi\left(\mu-\sqrt{x}\right)dx \\
&= \int_0^{+\infty}\left(1-\left(1+\exp\left(-\frac{\pi}{\sqrt{3}}\left(\frac{\left(\mu+\sqrt{x}\right)-\mu}{\sigma}\right)\right)\right)^{-1}\right)dx \\
&+\int_0^{+\infty}\left(1+\exp\left(-\frac{\pi}{\sqrt{3}}\left(\frac{\left(\mu-\sqrt{x}\right)-\mu}{\sigma}\right)\right)\right)^{-1}dx
\end{aligned}$$

i.e.,

$$\begin{aligned}
&\mathrm{E}\left[(\xi-\mu)^2\right] \\
&\le \int_0^{+\infty}\left(1-\left(1+\exp\left(-\frac{\pi}{\sqrt{3}}z\right)\right)^{-1}\right)d(\sigma z)^2 \\
&+\int_0^{+\infty}\left(1-\left(1+\exp\left(-\frac{\pi}{\sqrt{3}}(-z)\right)\right)^{-1}\right)d(\sigma z)^2 \\
&= 2\sigma^2\int_0^{+\infty}\left(1-\left(1+\exp\left(-\frac{\pi}{\sqrt{3}}z\right)\right)^{-1}\right)dz \\
&+2\sigma^2\int_0^{+\infty}\left(1-\left(1+\exp\left(\frac{\pi}{\sqrt{3}}z\right)\right)^{-1}\right)dz \\
&= \sigma_{\xi,u}^2
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&\mathrm{E}\left[(\xi-\mu)^2\right] \\
&= \int_0^{+\infty}\lambda\left\{(\xi-\mu)^2 \ge x\right\}dx \\
&= \int_0^{+\infty}\lambda\left\{\left\{\xi \ge \mu+\sqrt{x}\right\}\cup\left\{\xi \ge \mu-\sqrt{x}\right\}\right\}dx \\
&\ge \int_0^{+\infty}\left(\lambda\left\{\xi \ge \mu+\sqrt{x}\right\}dx+\right)dx \vee \int_0^{+\infty}\lambda\left\{\xi \ge \mu-\sqrt{x}\right\}dx \\
&= \int_0^{+\infty}\left(1-\Psi_\xi\left(\mu+\sqrt{x}\right)\right)dx \vee \int_0^{+\infty}\Psi_\xi\left(\mu-\sqrt{x}\right)dx \\
&= 2\sigma^2\left[\int_0^{+\infty}z\left(1-\left(1+\exp\left(-\frac{\pi}{\sqrt{3}}z\right)\right)^{-1}\right)dz \vee \int_0^{+\infty}z\left(1+\exp\left(\frac{\pi}{\sqrt{3}}z\right)\right)^{-1}dz\right] \\
&= \sigma_{\xi,l}^2
\end{aligned} \tag{23}$$

Now, we need to show that

$$\begin{aligned}
&\int_0^{+\infty}z\left(1-\left(1+\exp\left(-\frac{\pi}{\sqrt{3}}z\right)\right)^{-1}\right)dz \\
&= \int_0^{+\infty}z\left(1+\exp\left(\frac{\pi}{\sqrt{3}}z\right)\right)^{-1}dz
\end{aligned} \tag{24}$$

Therefore,

$$\begin{aligned}
&\int_0^{+\infty}z\left(1-\left(1+\exp\left(-\frac{\pi}{\sqrt{3}}z\right)\right)^{-1}\right)dz \\
&= \int_0^{+\infty}z\left(1-\left(1+\frac{1}{\exp\left(\frac{\pi}{\sqrt{3}}z\right)}\right)^{-1}\right)dz \\
&= \int_0^{+\infty}z\left(1-\frac{\exp\left(\frac{\pi}{\sqrt{3}}z\right)}{\exp\left(\frac{\pi}{\sqrt{3}}z\right)+1}\right)dz \\
&= \int_0^{+\infty}z\left(1-\frac{\exp\left(\frac{\pi}{\sqrt{3}}z\right)}{\exp\left(\frac{\pi}{\sqrt{3}}z\right)+1}\right)dz \\
&= \int_0^{+\infty}z\frac{1}{\exp\left(\frac{\pi}{\sqrt{3}}z\right)+1}dz \\
&= \int_0^{+\infty}z\left(1+\exp\left(\frac{\pi}{\sqrt{3}}z\right)\right)^{-1}dz
\end{aligned} \tag{25}$$

Finally,

$$\kappa = \frac{\sigma^2_{\xi,l}}{\sigma^2_{\xi,u}} = \frac{\int_0^{+\infty}\left(1-\left(1+\exp\left(-\frac{\pi}{\sqrt{3}}z\right)\right)^{-1}\right)dz}{2\int_0^{+\infty}\left(1-\left(1+\exp\left(-\frac{\pi}{\sqrt{3}}z\right)\right)^{-1}\right)dz} = \frac{1}{2}$$

Remark III.11: In the family of uncertain variables given the distributional form, the spread ratio $\kappa \in [1/2,1]$ and the length of the interval variance is $(1-\kappa)\sigma^2_{\xi,u}$ in general. When the uncertain distribution is normal, the spread ratio κ reaches the minimum value 1/2. In other words, the interval variance of an uncertain normal variable defined by Eq.(11) enjoys the longest length, i.e., $\sigma^2_{\xi,u}/2 \triangleq \sigma^2/2$.

4. The Auto covariance of an uncertain canonical process

An uncertain process $\{\xi_t,\ t\geq 0\}$ is a family of uncertainty variables indexed by t and taking values in the state space $\mathbb{S}\subseteq\mathbb{R}$.

Definition IV.1: (Liu, [7], [8], [9]) Let $\{C_t, t\geq 0\}$ be an uncertain process. If

(1) $C_0 = 0$ and all the trajectories of realizations are Lipschitz-continuous;
(2) $\{C_t, t\geq 0\}$ has stationary and independent increments;
(3) Every increment $C_{t+s}-C_s$ is a normal uncertainty variable with expected value 0 and variance t^2, i.e., the uncertainty distribution of $C_{t+s}-C_s$ is

$$\Psi_{C_{t+s}-C_s}(z) = \left(1+\exp\left(-\frac{\pi z}{\sqrt{3}t}\right)\right)^{-1} \tag{26}$$

then $\{C_t, t\geq 0\}$ is called an uncertain canonical process.

Definition IV.2: The auto covariance is the covariance between two members C_s and C_t, denoted by $\mathrm{E}[C_sC_t]$ for $\forall s,t,\ s<t$.

Theorem IV.3: Assuming that $\{C_t, t\geq 0\}$ is a standard uncertain canonical process. Then for $\forall s<t$, the auto covariance for a standard uncertain canonical process $\{C_t, t\geq 0\}$ is

$$\sigma_{s,t} = \left[\frac{st}{2}, st\right] \tag{27}$$

Proof: Notice that $C_\tau = \tau\xi_0$, $\tau\in[0,+\infty)$ is an arbitrary index. Thus it is logical to state that $C_t = (s/t)C_s$ because

$$\frac{C_t}{C_s} = \frac{t\xi_0}{s\xi_0} = \frac{t}{s} \tag{28}$$

which implies the linear relation $C_t = (t/s)C_s$. Furthermore,

$$\mathrm{E}[C_t] = \mathrm{E}[C_s] = 0 \tag{29}$$

Thus,

$$\begin{aligned}&\mathrm{E}[C_sC_t]\\ &= \mathrm{E}\left[C_s\left(\frac{t}{s}\right)C_s\right]\\ &= \left(\frac{t}{s}\right)\mathrm{E}\left[(C_s)^2\right]\\ &= \left(\frac{t}{s}\right)\mathrm{E}\left[(s\xi_0)^2\right]\\ &= stV[\xi_0]\end{aligned} \tag{30}$$

Recall for standard uncertain normal distribution, the interval variance is

$$V[\xi_0] = \left[\frac{1}{2},1\right] \tag{31}$$

Finally,

$$\mathrm{E}[C_sC_t] = \left[\frac{st}{2}, st\right] \tag{32}$$

Remark IV.3: The linearity between any two members of a standard uncertain canonical process $\{C_t, t\geq 0\}$ should be understood as the sense at $\lambda\{C_t = (t/s)C_s\} = 1$.

Definition IV.4: The auto covariance matrix $\Sigma = (\sigma_{s_is_j})_{n\times n}$, of members $C_{s_1}, C_{s_2}, \cdots, C_{s_n}$ in a standard uncertain canonical process $\{C_t, t\geq 0\}$ is defined by

$$\Sigma \triangleq \left(\sigma_{s_is_j}\right)_{n\times n} = \begin{bmatrix} V\left[C_{s_1}\right] & \sigma_{s_1s_2} & \cdots & \sigma_{s_1s_n} \\ \sigma_{s_1s_2} & V\left[C_{s_2}\right] & \cdots & \sigma_{s_2s_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{s_1s_n} & \sigma_{s_2s_n} & \cdots & V\left[C_{s_n}\right] \end{bmatrix} \tag{33}$$

Theorem IV.5: The auto covariance matrix $\Sigma = \left(\sigma_{s_is_j}\right)_{n\times n}$, of members $C_{s_1}, C_{s_2}, \cdots, C_{s_n}$ in a standard uncertain canonical process $\{C_t, t \ge 0\}$ is an interval matrix, whose element is

$$\Sigma \triangleq \left(\sigma_{s_is_j}\right)_{n\times n} = \left(\left[\frac{s_is_j}{2}, s_is_j\right]\right)_{n\times n} \tag{34}$$

Proof: According to Definition IV.4, the auto covariance matrix

$$\Sigma \triangleq \left(\sigma_{s_is_j}\right)_{n\times n}$$

whose $(i,j)^{th}$ element (at i^{th} row and j^{th} column) is given by Theorem IV.2: Eq.(27):

$$\sigma_{s,t} = \left[\frac{st}{2}, st\right].$$

Thus we reach the conclusion.

Remark IV.6: The upper bound of the auto covariance derived in this paper $\sup\left(\mathrm{E}\left[C_{s_i}C_{s_j}\right]\right) = s_is_j$ is much simplified comparing to the expression derived by Dai et al [2]. Nevertheless the current result still making intuitive sense if letting $s_j \to s_i$ the limit will be the upper bound of the variance $\sigma^2_{C_{s_i},u} = \sup\left(\mathrm{E}\left[\left(C_{s_i}\right)^2\right]\right) = s_i^2$. The elements of the interval auto covariance matrix $\Sigma = \left(\sigma_{s_is_j}\right)_{n\times n}$, of members $C_{s_1}, C_{s_2}, \cdots, C_{s_n}$ in a standard uncertain canonical process $\{C_t, t \ge 0\}$ are consistent. Now we are ready to investigate the uncertain canonical regression modeling issue.

5. The General Uncertain Canonical Process Regression

The general uncertain regression model to be proposed will take a conditional form

$$\begin{aligned} &y_i = \beta_0 + \beta_2 x_{1i} + \cdots + \beta_m x_{mi} + \sigma C_{s_i} \\ &s_i,\ i = 1, 2, \cdots, n \end{aligned} \tag{35}$$

where y is the response variable, $x_1, x_2, \cdots, x_m$ are m explanatory variables without uncertain influences, and σC_{s_i}, $i = 1, 2, \cdots, n$ are the uncertain error terms from a standard uncertain canonical process $\{C_t, t \ge 0\}$.

The matrix version of (33) is

$$\underline{y} = X\underline{\beta} + \underline{C} \tag{36}$$

where

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},\ X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{m1} \\ 1 & x_{12} & \cdots & x_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{mn} \end{bmatrix},\quad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix},\ \underline{C} = \begin{bmatrix} \sigma C_{s_1} \\ \sigma C_{s_2} \\ \vdots \\ \sigma C_{s_n} \end{bmatrix} \tag{37}$$

Theorem V.1: The basic uncertain regression model in Eq. (36) or Eq. (37) have the following properties:

(1) $\mathrm{E}\left[\underline{C}\right] = \underline{0}$;

(2) $\mathrm{E}\left[\underline{y}\right] = X\underline{\beta}$;

(3) The variance-covariance matrix $V\left[\underline{C}\right] = \mathrm{E}\left[\underline{C}\underline{C}^t\right]$ is the interval-valued auto covariance matrix Σ multiplying a constant positive multiplier σ^2, i.e., $\sigma^2\Sigma$.

Therefore the uncertain canonical regression is an interval-valued weighted regression model.

Proof: Based on the Definition IV.1 of a standard uncertain canonical process $\{C_t, t \ge 0\}$, $C_{s_1}, C_{s_2}, \cdots, C_{s_n}$ are members of the uncertain canonical process, whose indices are times s_i, $i = 1, 2, \cdots, n$, and therefore $C_{s_i} = C_{s_i} - C_0 = s_i\xi_0$ are uncertainty variables with an uncertain normal distribution

$$\Psi_{\xi_0}(z) = \left(1 + \exp\left(-\frac{\pi z}{\sqrt{3}}\right)\right)^{-1} \tag{38}$$

Then, the expectation $\mathrm{E}\left[C_{s_i}\right]=0$ and the variance $V\left[C_{s_i}\right]=\left[s_i^2/2, s_i^2\right]$. Hence $\mathrm{E}\left[\underline{C}\right]=\underline{0}$, $\mathrm{E}\left[\underline{y}\right]=X\underline{\beta}$ are proved.

As to Property (3) in the Definition IV.1, we noted the error vector is composed of $\sigma C_{s_1}, \sigma C_{s_2}, \cdots, \sigma C_{s_n}$ sequentially, according to Theorem IV.5, it is simple to establish that

$$\begin{aligned} &\sigma^2 \Sigma \\ &= \sigma^2 \mathrm{E}\left[\underline{C}\underline{C}^t\right] \\ &= \sigma^2 \left(\sigma_{s_i s_j}\right)_{n\times n} \\ &= \sigma^2 \left(\left[s_i s_j/2, s_i s_j\right]\right)_{n\times n} \end{aligned} \tag{39}$$

The regression model is an interval-valued weighted one since the variance-covariance is not $\sigma^2 I_{n\times n}$ but rather is $\sigma^2\left(\left[s_i s_j/2, s_i s_j\right]\right)_{n\times n}$.

Remark V.2: From the model formulation and Theorem V.1 statement and proof, it should be emphasized that a key feature of this basic uncertain regression model lies in the error term assumption: errors $\sigma C_{s_1}, \sigma C_{s_2}, \cdots, \sigma C_{s_n}$ are members of an uncertain canonical process, whose indices are s_i, $i=1,2,\cdots,n$. In classical regression model, the error terms are not taken sequentially. In other words, sequential order of error terms plays a fundamental role in uncertain regression. In that sense, it is logical to describe an uncertain regression model as a sequential regression model. Furthermore, if we define a scalar variance-covariance matrix $\forall \Pi \in \Sigma = \left(\left[s_i s_j/2, s_i s_j\right]\right)_{n\times n}$ according to Moore's interval uncertainty [13], then the scalar matrix Π would facilitate an coefficient vector estimator with BLUE property of the so-called scalar uncertain canonical regression in terms of the Gauss-Markov Theorem, [3], [12], [15].

Theorem V.3: (Gauss-Markov Theorem) The estimator of the coefficient vector in the scalar uncertain regression defined by Remark V.2 is BLUE (Best - minimum variance, Linear, Unbiased Estimator), i.e., $\hat{\underline{\beta}} = \left(X^t\Pi^{-1}X\right)^{-1} X^t\Pi^{-1}\underline{y}$ is BLUE.

Proof: The estimator of $\underline{\beta}$, $\hat{\underline{\beta}} = \left(X^t\Pi^{-1}X\right)^{-1} X^t\Pi^{-1}\underline{y}$, is linear in response values $y_2, y_2, \cdots, y_n$ as is obvious from the a $(m+1)\times n$ matrix pre-multiplier. The unbiasedness follows from:

$$\begin{aligned} &\mathrm{E}\left[\hat{\underline{\beta}}\right] \\ &= \mathrm{E}\left[\left(X^t\Pi^{-1}X\right)^{-1} X^t\Pi^{-1}\underline{y}\right] \\ &= \left(X^t\Pi^{-1}X\right)^{-1} X^t\Pi^{-1}\mathrm{E}\left[\underline{y}\right] \\ &= \left(X^t\Pi^{-1}X\right)^{-1} X^t\Pi^{-1}X\underline{\beta} = \underline{\beta} \end{aligned} \tag{40}$$

Then, let us assume another unbiased estimator $\tilde{\underline{\beta}} = A\underline{y} = \left(\left(X^t\Pi^{-1}X\right)^{-1} X^t\Pi^{-1} + D\right)\underline{y}$, then

$$\begin{aligned} &\mathrm{E}\left[\tilde{\underline{\beta}}\right] \\ &= \mathrm{E}\left[\left(\left(X^t\Pi^{-1}X\right)^{-1} X^t\Pi^{-1} + D\right)\underline{y}\right] \\ &= \left(\left(X^t\Pi^{-1}X\right)^{-1} X^t\Pi^{-1} + D\right)\mathrm{E}\left[\underline{y}\right] \\ &= \left(\left(X^t\Pi^{-1}X\right)^{-1} X^t\Pi^{-1}X + DX\right)\underline{\beta} \\ &= \left(I + DX\right)\underline{\beta} \end{aligned} \tag{41}$$

Then $DX = 0$ is implied. Now, let us examine the variance of $\tilde{\underline{\beta}}$,

$$\begin{aligned} V\left[\tilde{\underline{\beta}}\right] &= V\left[\left(X^t\Pi^{-1}X\right)^{-1} X^t\Pi^{-1}\underline{y} + D\underline{y}\right] = V\left[\hat{\underline{\beta}} + D\underline{y}\right] \\ &= V\left[\hat{\underline{\beta}}\right] + D\mathrm{E}\left[\underline{C}\underline{C}^t\right]D^t = V\left[\hat{\underline{\beta}}\right] + D\Pi^{-1}D^t \ge V\left[\hat{\underline{\beta}}\right] \end{aligned} \tag{42}$$

Remark V.4: More importantly, we should be fully aware that the scalar variance-covariance matrix Π is an arbitrary one in the autovariance matrix defined by the members in a standard uncertain canonical process $\{C_t, t\ge 0\}$, therefore in the general uncertain canonical regression with interval-valued auto covariance matrix $\left(\left[s_i s_j/2, s_i s_j\right]\right)_{n\times n}$ the BLUE property should be held for the interval-valued estimator of the coefficient vector in certain sense. Furthermore, it is important to emphasize that the auto covariance interval matrix Σ is available according to **Theorem IV.5**. There is no need to estimate it by replicated observations from each combination of the explanatory variables.

6. Conclusion

Although the uncertain measure theory is very abstract, it is necessary to develop some uncertain statistical models for those practitioners in applied scientific or engineering fields. In this paper, we develop regression model given the error vector taken from a standard canonical process with a positive multiplier. Due to interval variance fact, the covariance can only take interval form accordingly, since the distribution functions for members in a standard canonical process are given but the uncertain measures are not. The development takes the advantage in a standard canonical process, whose member can be represents by standard uncertain normal variable, thus the linear relationship between members is established. In general, for uncertain processes we can also develop the interval-valued auto covariance matrix as long as they are uncertain canonical process driven, similar to [8].

Acknowledgements

This research is partially supported financially by the South African National Research Foundation (FR2011040400096).

References

[1] Billard, L. and Diday, E., Regression Analysis for Interval-Valued Data, In: Data Analysis, Classification and Related Methods, Proceedings of the 7th Conference of the International Federation of Classification Societies, Springer-Verlag, Belgium, 2000, pp. 369-374.

[2] Dai, W., Guo, R. Guo, D. and Cui, Y.H. , The Autocorrelation of a Standard Uncertain Canonical Process. In the Proceedings of the Second International Conference on Uncertainty Theory (ICUT2011 http://orsc.edu.cn/icut), August 6-11, 2011, Lhasa, Tibet, China, pp 6-12, Editors: Kaoru Hirota, Dan A. Ralescu, Jin Peng. ISSN 2079-5238. Copyright © 2011 by International Consortium for Uncertainty Theory.

[3] Draper, N. and Smith, H., *Applied Regression Analysis*, 2nd Edition, John Wiley & Sons Inc, New York, USA, 1966.

[4] Guo, R., Guo, D., Dunne, T. and Thiart, C., DEAR Model – The Theoretical Foundation, *Journal of Uncertain Systems*, vol.3, no. 1, 36-51, 2009.

[5] Guo, R. and Guo, D., Random Fuzzy Variable Foundation for Grey Differential Equation Modelling, *Soft Computing*, vol. 13, no. 2, 185-201, January, 2009.

[6] Guo, R., Cui, Y.H. and Guo, D., Uncertainty Linear Regression Models, 2011, *Journal of Uncertainty Systems*.

[7] Guo, R., Dai, W., Guo, D., Cui, Y.H. and Dunne, T. Uncertain Canonical Process Regression Models. In the Proceedings of the Second International Conference on Uncertainty Theory (ICUT2011 http://orsc.edu.cn/icut), August 6-11, 2011, Lhasa, Tibet, China, pp 36-44, Editors: Kaoru Hirota, Dan A. Ralescu, Jin Peng. ISSN 2079-5238. Copyright © 2011 by International Consortium for Uncertainty Theory.

[8] Guo, R., Guo, D., and Cui, Y.H., The Autocorrelation of an Uncertain Integral Driven Process. In the Proceedings of the Second International Conference on Uncertainty Theory (ICUT2011 http://orsc.edu.cn/icut), August 6-11, 2011, Lhasa, Tibet, China, pp 116-122, Editors: Kaoru Hirota, Dan A. Ralescu, Jin Peng. ISSN 2079-5238. Copyright © 2011 by International Consortium for Uncertainty Theory.

[9] Liu, B.D., *Uncertainty Theory: An Introduction to Its Axiomatic Foundations*, Second Edition, Springer-Verlag Heidelberg, Berlin, 2007.

[10] Liu, B.D., *Uncertainty Theory: A Branch of Mathematics of Modelling Human Uncertainty*, Springer-Verlag, Berlin, 2010.

[11] Liu, B.D., *Uncertainty Theory*, 2012. (Drafted Fourth Edition)

[12] Myers, R.H., *Classical and Modern Regression with Applications*, Duxbury Press, 2nd Edition, March 28, 2000.

[13] Moore, R.E., *Interval Analysis*, Prentice-Hall, Englewood Cliff, NJ., 1966.

[14] Rasmussen, C.E. and Williams, C.K.I., *Gaussian Processes for Machine Learning*, MIT Press, 2006.

[15] Searle, S.R., *Matrix Algebra Useful For Statistics*, John Wiley & Sons Inc, New York, USA, 1982.

[16] Williams, C.K.I. and Rasmussen, C.E., Gaussian processes for regression, In: Touretzky, D. S., Mozer, M. C. , and Hasselmo, M. E. , (eds), *Advances in Neural Information Processing Systems*, Volume 8, pp. 598-604, MIT Press, 1996.