Journal of Uncertain Systems
Vol.3, No.1, pp.31-51, 2009
Online at: www.jus.org.uk

DEAR Model －The Theoretical Foundation

Renkuan Guo[1,*], Danni Guo[2], Tim Dunne[1], and Christien Thiart[1]
[1]*Department of Statistical Sciences, University of Cape Town*
Private Bag, Rondebosch 7701, Cape Town, South Africa
[2]*Kirstenbosch Research Center, South African National Biodiversity Institute*
Private Bag X7, Claremont 7735, Cape Town, South Africa

Received 10 August 2007; Accepted 20 January 2008

Abstract

In this paper, we merge the differential equation model, regression model and credibility measure based fuzzy mathematics proposed by Liu [13] into a new differential equation associated regression model (abbreviated as DEAR model). The creation of the DEAR model does not only offer a rigorous treatment of the grey differential equation problem proposed by Deng [3] on the random fuzzy variable theoretical foundation, but also increases the variety of model choices greatly. Furthermore, we develop a multivariate DEAR model for the quantitative modeling based on multivariate small sample data. Multivariate DEAR models will be able to establish the quantitative relationship among the main response factor vectors and the covariate vectors, which is a major improvement of information extraction under sparse data availability. Finally, we point out some potential application areas of the new modeling family.

Keywords: DEAR model, Coupling principle, random fuzzy variable, average chance distribution

1 Introduction

When facing a system in the real world, the dynamic law governing the system will be the aim of study. Statisticians, mathematicians or engineers often use the linearization idea to establish a linear model (geometrically a hyper-plane in multi-dimensional space) on the region where the sampling data are collected to approximate the true nonlinear dynamics (high dimensional surface or curve) according to linear model theory. Without any doubt, non-linear statistical models may be also established based on maximum likelihood theory (or other criterion, say, maximum entropy). However, the non-linear statistical modeling is required large sample for accuracy. A rule of thumb for the ratio of the number of sample points to the number of parameters to be estimated is 15 to 1.

In real life, data collection requires manpower, equipment resources, and money and thus is expensive. Furthermore, in today's fast globalization business environments, product life circles are often shortening. For example, in nano industries, every three months a new generation of nano chips is generated. In such circumstances, there is no chance to collect of data. In other words, it is necessary to address the problem analyzing of real world complicated systems without adequate sample information available.

The fundamental challenge becomes a problem whether a system dynamic law (nonlinear in nature) can be revealed in terms of small sample data information. Utilizing small sample data in simple linear regression is well-developed in statistics, while the differential equation is a well-known model for describing nonlinear law in many scientific fields. These facts lead to the question whether a linear modeling of small sample could a generate nonlinear law.

In this paper, we merge differential equation theory, linear regression theory, and random fuzzy variable theory into a new mathematical theory of a small-sample oriented system analysis, which is named as DEAR, an abbreviation of differential equation associated regression.

The structure of this paper is stated as follows. Section 2 introduces the DEAR concept and a classification of the DEAR modeling family. In Section 3, we discuss the role of the Coupling Principle in DEAR model formation and the error structure from coupling of differential equation and regression. Section 4 discusses the parameter estimation

* Corresponding author. Email: rguo@worldonline.co.za (R. Guo).

in a univariate DEAR model. Section 5 is a brief discussion of the formation of a multivariate DEAR model for a single exploratory variable. Section 6 discusses the bivariate DEAR model which details the relevant theoretical developments. A brief conclusion is given in Section 7. In the Appendix, there is a brief review of Liu's [13,14] credibility measure axiomatic foundation and random fuzzy variable theory.

2 Concept of DEAR Model, Classification

In engineering theory, particularly, in modern control theory, it is often convenient to use a hypothesized differential equation to describe the dynamic law of a continuous system, for example, a p^{th}-order univariate differential equation, which proposes an unknown *d*-dimensional parameter vector $\underline{\theta}^T = (\theta_1, \theta_2, \cdots, \theta_d)$.

However, until the unknown parameter vector $\underline{\theta}$ is estimated by data information obtained from sampling from the system itself, the system dynamic law is still unspecified. As we noted in the introduction, DEAR uses a linear regression approach to achieve nonlinear modeling aim, which may seem contradictory. How could this aim be achieved? The question deserves an intuitive explanation.

Without loss of generality, a simple linear differential equation $dx/dt = \alpha + \beta x$ is used in this paper for illustrative purposes. Let $\hat{x}_i^{(1)}$ denote an approximation to the primitive function $x(t)$ at t_i, and let $\Delta x_i/\Delta t_i$ be an approximation to the derivative function dx/dt at t_i, with $\Delta x_i = x(t_i) - x(t_{i-1})$, $\Delta t_i = t_i - t_{i-1}$.

Definition 2.1: If a dynamic system governed by $dx/dt = \alpha + \beta x$ is observed through *n* samples at its derivative level, denoted by $X^{(0)} = \left\{x_1^{(0)}, x_2^{(0)}, \cdots, x_n^{(0)}\right\}$, the coupled equation system

$$\begin{cases} \dfrac{dx}{dt} = \alpha + \beta x \\ x_i^{(0)} = \alpha + \beta \hat{x}_i^{(1)} + \varepsilon_i,\ i = 2, 3, \cdots, n \end{cases} \tag{1}$$

is called Type I DEAR model.

Definition 2.2: If a dynamic system governed by Eq. (1) is sampled at its primitive level with sample size *n*, denoted by $X^{(1)} = \left\{x(t_1), x(t_2), \cdots, x(t_n)\right\}$, the coupled equation system

$$\begin{cases} \dfrac{dx}{dt} = \alpha + \beta x \\ \dfrac{\Delta x_i}{\Delta t_i} = \alpha + \beta x(t_i) + \varepsilon_i,\ i = 2, 3, \cdots, n \end{cases} \tag{2}$$

is called Type II DEAR model.

The second equations in Eq. (1) and Eq. (2) are called the coupled regressions, while the first equations are called the associated differential equations.

Let us examine Type I DEAR model first. The system dynamics are governed by the linear differential equation $dx/dt = \alpha + \beta x$, or equivalently, a nonlinear functional $x(t) = f(t; \alpha, \beta)$. If the sample could be very large, it is possible to perform a non-linear statistical modeling in terms of standard maximum likelihood procedures to estimate a system parameter $\underline{\theta} = (\alpha, \beta)$. However, if only small-sample observations are available, the "best" modeling exercise is to fit a simple regression model $\hat{x}(t) = \hat{\gamma}_0 + \hat{\gamma}_1 t$, called primitive regression, for approximating the system dynamics $x(t) = f(t; \alpha, \beta)$. Figure 1 shows that the blue-dot straight line $\hat{x}(t) = \hat{\gamma}_0 + \hat{\gamma}_1 t$ may poorly approximate a nonlinear curve $x(t) = f(t; \alpha, \beta)$ in the (t, x) space (or (t, x)-coordinate system).

Let us consider the case where sampling observations are collected at first-order derivative level, denoted as $X^{(0)} = \left\{x_1^{(0)}, x_2^{(0)}, \cdots, x_n^{(0)}\right\}$. By a linear transformation, approximations to primitive function level observations are obtained, denoted by $\left\{x(t_1), \hat{x}(t_2), \cdots, \hat{x}(t_n)\right\}$, say, by partial sum. In terms of Type I DEAR model thinking, we first fit the coupled regression, i.e., the second equation in the DEAR system of Eq. (2) in the (x, x') space (or (x, x')-coordinate system), where x' denotes the derivative of x with respect to t, i.e., $x' = dx/dt$.

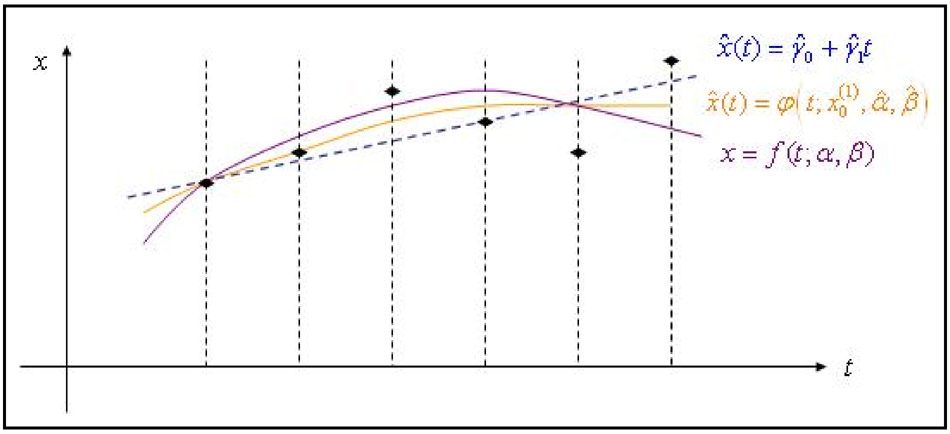

Figure 1: Two approximations to nonlinear curve $x(t) = f(t;\alpha,\beta)$ in (t, x) space

From the fitting of the coupled regression, $x_i^{(0)} = \alpha + \beta \hat{x}_i^{(1)} + \varepsilon_i$, the estimator of parameter $\underline{\theta} = (\alpha, \beta)$, denoted by $\hat{\underline{\theta}} = (\hat{\alpha}, \hat{\beta})$ is obtained. Now, in the (x, x') space, we fit the straight line $\hat{x}' = \hat{\alpha} + \hat{\beta}\hat{x}$ to approximate the straight line $x' = \alpha + \beta x$. It is obvious that goodness-of-fit measures for this model could be very good even with small samples..

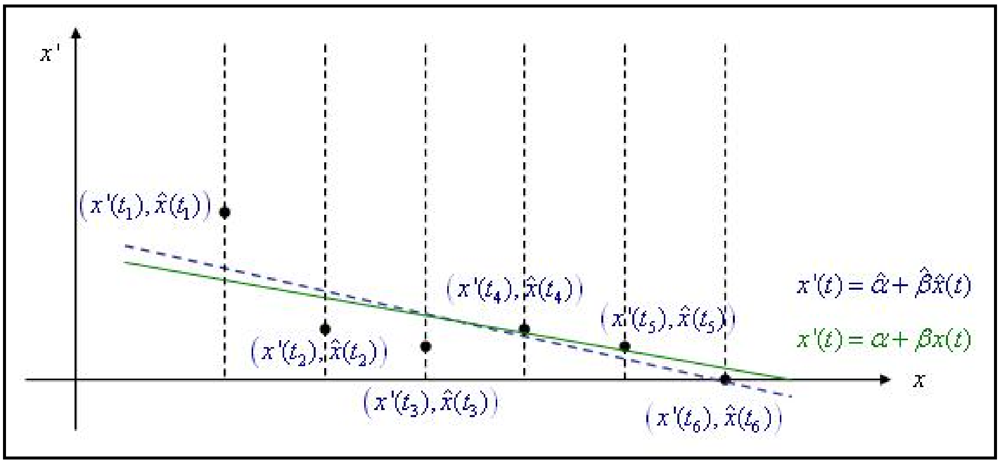

Figure 2: Type I approximation in (x, x') space

Once the parameter $\underline{\theta} = (\alpha, \beta)$ is obtained, by solving the approximated linear differential equation $dx/dt = \hat{\alpha} + \hat{\beta}x$, we will obtain an approximated nonlinear curve $x' = \varphi(t; x_0^{(1)}, \hat{\alpha}, \hat{\beta})$, (yellow-colored curve in Figure 1), which is expected to approximate the primitive curve with relatively high accuracy.

Let us consider the case in which the sampling observations are collected at primitive function level, denoted as $X^{(1)} = \{x(t_1), x(t_2), \cdots, x(t_n)\}$. Then in terms of DEAR Type II model thinking, the derivatives could be approximated, for example, by the divided difference, i.e., $\Delta x_i / \Delta t_i$, or other approaches available. Just as shown in Figure 3, we fit $\hat{x}' = \Delta x / \Delta t = \hat{\alpha} + \hat{\beta}x$ for the approximating line $x' = \alpha + \beta x$. Similarly, the estimated parameter $\hat{\underline{\theta}} = (\hat{\alpha}, \hat{\beta})$ will lead from the nonlinear approximation $x' = \varphi(t; x_0^{(1)}, \hat{\alpha}, \hat{\beta})$ to the primitive function $x(t) = f(t;\alpha,\beta)$ in (t, x) space (shown in Figure 1).

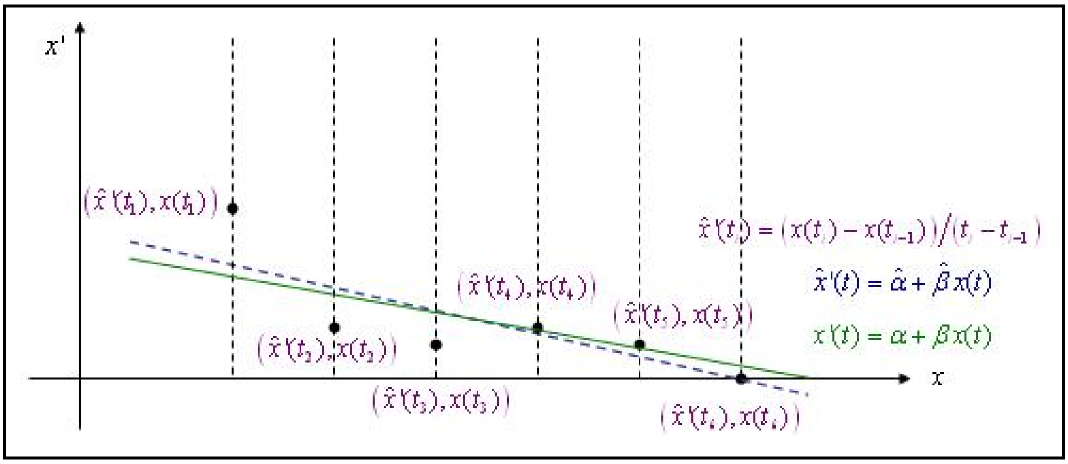

Figure 3: Type II approximation in (x, x') space

Remark 2.3: As hinted by the intuitive discussion of Type I and Type II DEAR models, the associated differential equation should take a linear form in system parameters, i.e., $\theta_1, \theta_2, \cdots, \theta_d$, whenever possible. The maximal dimensionality of $\underline{\theta}$ should be $d_{\max} = n-2$. Furthermore, the hypothesized differential equation model should include minimal numbers of unknown parameters, $\underline{\theta}$.

To further discussion, it is necessary to define the divided differences and partial sum for approximating primitive function respectively.

Definition 2.4: Given a set of *n* sampling observations at p^{th} order derivative level over $(t_1, t_2, \cdots, t_n)$, denoted by $X^{(-p+1)} = \left\{x^{(-p+1)}(t_1), x^{(-p+1)}(t_2), \cdots, x^{(-p+1)}(t_n)\right\}$, then the approximations to the k^{th} order derivatives are denoted by $\hat{X}^{(-k+1)} = \left\{\hat{x}^{(-k+1)}(t_{p-k}), \hat{x}^{(-k+1)}(t_{p-k+1}), \cdots, \hat{x}^{(-k+1)}(t_n)\right\}$, where $k = 1, 2, \cdots, p-2$.

Then the approximations to the first order derivatives dx/dt and primitive function $x(t)$, are denoted by

$$\hat{X}^{(0)} = \left\{\hat{x}^{(0)}(t_{p-1}), \hat{x}^{(0)}(t_p), \cdots, \hat{x}^{(0)}(t_n)\right\}$$

and

$$\hat{X}^{(1)} = \left\{\hat{x}^{(1)}(t_p), \hat{x}^{(1)}(t_{p+1}), \cdots, \hat{x}^{(1)}(t_n)\right\},$$

respectively. It is obvious that the notation in Definition 2.4 is consistent with Definition 2.1. The question of how to obtain $\hat{X}^{(-p+2)}, \hat{X}^{(-p+3)}, \cdots, \hat{X}^{(0)}, \hat{X}^{(1)}$ is a matter of approximations to relevant integrations

$$\int \left(d^{(p)}x/dt^p\right)dt\,,\ \int \left(d^{(p-1)}x/dt^{p-1}\right)dt\,, \cdots\,, \int \left(d^{(2)}x/dt^2\right)dt\,,\ \int (dx/dt)dt\,.$$

For example, given the first-order (approximated) values

$$\hat{X}^{(0)} = \left\{\hat{x}^{(0)}(t_{p-1}), \hat{x}^{(0)}(t_p), \cdots, \hat{x}^{(0)}(t_n)\right\},$$

the approximate values of primitive function may be obtained by

$$\hat{x}^{(1)}(t_k) = \sum_{j=2}^{k} \hat{x}^{(0)}(t_j)(t_j - t_{j-1}) \approx \int_{t_1}^{t_k} \left(\frac{dx}{dt}\right) dt\,.$$

Definition 2.5: Given a set of *n* sampling observations at p^{th} order derivative level over $(t_1, t_2, \cdots, t_n)$, denoted by $X^{(-p+1)} = \left\{x^{(-p+1)}(t_1), x^{(-p+1)}(t_2), \cdots, x^{(-p+1)}(t_n)\right\}$, then a pair of equations is called a (p^{th} order) Type I DEAR model if

$$\begin{cases} \dfrac{d^{(p)}x}{dt^{p}} = \varphi\left(\dfrac{d^{(p-1)}x}{dt^{p-1}}, \dfrac{d^{(p-2)}x}{dt^{p-2}}, \cdots, \dfrac{dx}{dt}, x; \underline{\theta}\right) & (a) \\ x^{(-p+1)}(t_k) = \varphi\left(\hat{x}^{(-p+2)}(t_k), x^{(-p+3)}(t_k), \cdots, \hat{x}^{(-1)}(t_k), x^{(0)}(t_k), \hat{x}^{(1)}(t_k); \underline{\theta}\right) + \varepsilon_k \quad k = p, p+1, \cdots, n. & (b) \end{cases} \tag{3}$$

Eq. (3a) is called the associated differential equation and Eq. (3b) is called the coupled regression model.

It is obvious that Definition 2.5 is a generalization of *1st* order DEAR Type I model. The Type II (p^{th} order) DEAR model can be defined similarly.

Definition 2.6: Given a set of n sampling observations at primitive function level, $x(t)$, denoted by $X^{(1)} = \left(x^{(1)}(t_1), x^{(1)}(t_2), \cdots, x^{(1)}(t_n)\right)$, then a pair of equations is called (p^{th} order) Type II DEAR model if

$$\begin{cases} \dfrac{d^{(p)}x}{dt^{p}} = \varphi\left(\dfrac{d^{(p-1)}x}{dt^{p-1}}, \dfrac{d^{(p-2)}x}{dt^{p-2}}, \cdots, \dfrac{dx}{dt}, x; \underline{\theta}\right) & (a) \\ \hat{x}^{(-p+1)}(t_k) = \varphi\left(\hat{x}^{(-p+2)}(t_k), \hat{x}^{(-p+3)}(t_k), \cdots, \hat{x}^{(-1)}(t_k), \hat{x}^{(0)}(t_k), x^{(1)}(t_k); \underline{\theta}\right) + \varepsilon_k \quad k = p, p+1, \cdots, n. & (b) \end{cases} \tag{4}$$

In Type II model, obtaining the values of $\hat{X}^{(-p+1)}, \hat{X}^{(-p+2)}, \cdots, \hat{X}^{(-1)}, \hat{X}^{(0)}$ is matter of approximating the derivatives $d^{(p)}x/dt^{p}, d^{(p-1)}x/dt^{p-1}, \cdots, d^{(2)}x/dt^{2}, dx/dt$. Typically, the divided difference is a primary approximation approach, for example,

$$\hat{x}^{(0)}(t_i) = \frac{\Delta x_i}{\Delta t_i} = \alpha x(t_i) + \beta x^{m}(t_i).$$

However, in numerical analysis literature, there are many efficient approaches for derivative approximations.

If we examine the DEAR models in Eq. (3) and Eq. (4) in detail, we will find that a DEAR model starts with an associated differential equation, then, the coupled regression model is specified in the discretized form of the associated differential equation, in turn, in terms of coupling regression model. The parameters specifying the associated differential equation are estimated under a least-squares criterion.

Furthermore, the solution to the associated differential equation (or the discretized solution) evaluated with data-assimilated parameter estimates, is used for system analysis or prediction. We should emphasize here that the way a DEAR model uses system sampling information to solve the associated differential equation is different from that in common algorithms for solving a differential equation numerically. In a DEAR model, we will obtain a closed-form functional solution (i.e., the primitive function) to the associated differential equation with optimal data-assimilated parameters. The availability of the closed-form primitive function $x(t)$, will provide great conveniences in the further investigation on the system under study. We acknowledge that the idea of obtaining a closed-form solution to the associated differential equation was suggested by the founder of Grey Mathematics, Deng [3]. The DEAR models defined in Definitions 2.5 and 2.6 have a common feature that both of them start with (a hypothesized) differential equation model and then the coupling regression model. Therefore, they are differential equation motivated regression (abbreviated as DEMR) models, Guo *et al.* [10].

It is possible that in modeling real world data, there is no hypothesized differential equation as such a *priori*. What we may do is to search a best fitted regression model, for example, a set of system data $\{x(t_i), i = 1, 2, \cdots, n\}$ is collected and a fitted regression model takes the form

$$\hat{x}^{(0)}(t_i) = \frac{\Delta x_i}{\Delta t_i} = \alpha x(t_i) + \beta x^{m}(t_i).$$

Then, the associated differential equation is a Bernoulli equation of the form

$$\frac{dx}{dt} + p(t)x^{2} = q(t)x^{m}.$$

Then a model can be established as

$$\begin{cases} \dfrac{\Delta x_i}{\Delta t_i} = \alpha x(t_i) + \beta \\ x^m(t_i) + \varepsilon_i \; \dfrac{dx}{dt} = \alpha x + \beta x^m. \end{cases} \tag{5}$$

It should be fully appreciated that the estimated Bernoulli equation $\frac{dx}{dt} = \hat{\alpha}x + \hat{\beta}x^m$, which results in a solution

$$\varphi(t;\hat{\alpha},\hat{\beta}) = \sqrt[1-m]{\frac{\hat{\beta}}{\hat{\alpha}}\left(e^{\hat{\alpha}(1-m)(t-t_1)} - 1\right) + c(t_1)},\ (m \neq 0,1)$$

for producing the nonlinear approximation to the true system dynamics $x = f(t;\alpha,\beta,\gamma)$.

The example discussed actually begins from a fitted regression and then searching for the best matched (coupled) differential equation from a differential equation family. This kind of nonlinear modeling idea can be called regression motivated differential equation (abbreviated RMDE) modeling. For the distinction between DEMR and RMDE models, we always put the motivated model first. For example, in the RMDE model of Eq. (5), the regression comes first and the differential equation comes second.

In summary, DEAR models are classified into two families: DEMR and RMDE. Each family is classified into two subfamilies: Type I (sampling at derivative level and thus primitive function requires approximation) and Type II (sampling at primitive function level and thus approximation to the derivative is necessary), as in Table 1.

Table 1: DEAR family and its classifications

	DEAR	
DEAR subfamily	DEMR	RMDE
Type I (data at derivative level)	$\begin{cases} \dfrac{dx}{dt} = \alpha + \beta x \\ x_i^{(0)} = \alpha + \beta \hat{x}_i^{(1)} + \varepsilon_i,\ i = 1,2,\cdots,n \end{cases}$	$\begin{cases} x_i^{(0)} = \alpha + \beta \hat{x}_i^{(1)} + \varepsilon_i,\ i = 1,2,\cdots,n \\ \dfrac{dx}{dt} = \alpha + \beta x \end{cases}$
TYPE II (data at primitive level)	$\begin{cases} \dfrac{dx}{dt} = \alpha + \beta x \\ \hat{x}_i^{(0)} = \dfrac{\Delta x^{(1)}(t_k)}{\Delta t_k} = \alpha + \beta x_i^{(1)} + \varepsilon_i,\ i = 1,2,\cdots,n \end{cases}$	$\begin{cases} \hat{x}_i^{(0)} = \dfrac{\Delta x^{(1)}(t_k)}{\Delta t_k} = \alpha + \beta x_i^{(1)} + \varepsilon_i,\ i = 1,2,\cdots,n \\ \dfrac{dx}{dt} = \alpha + \beta x \end{cases}$

We should emphasize that the new DEAR modeling family proposed by Guo *et al*. [7, 8, 9, 10], includes very rich members and therefore the DEAR modeling family will have potentially wide applications. In data modeling exercises if a linear differential equation (or non-linear in some cases) has the close-form solution its coupled regression can be identified for fitting the data collected. Conversely, a very good linear regression model may be "pointed" to a differential equation with close-form solution. Therefore, in either case, the approximated solution of the associated differential equation with fitted parameters obtained from regression modeling should provide accurate predictions. Furthermore, DEAR modeling idea may facilitate an accurate model goodness of fit without large sample requirement.

Linear regression modeling practices revealed that sample data collected possesses both local features and global features. Goodness of fit, as a measure of modeling accuracy, is local. The solution (with estimated parameters) to the associated differential equation is global. DEAR modeling is actually seeking a balance between local and global features carried by sample data.

For illustrating the richness and easiness of DEAR family, we list seven elementary models in Table 2. Note that the order of differential equation: $p = 1, 2$ and φ a linear function in (4a).

Table 2: Seven elementary models in the Type II DEMR subfamily

DEAR model	Order $p=1$	Order $p=2$
A	$\begin{cases} \dfrac{dx}{dt}=\alpha_0+\alpha_1 x \\ \dfrac{\Delta x(t_k)}{\Delta t_k}=\alpha_0+\alpha_1 x(t_k)+\varepsilon_k \end{cases}$	$\begin{cases} \dfrac{d^2x}{dt^2}=\alpha_0+\alpha_1 x+\alpha_2\dfrac{dx}{dt} \\ \dfrac{\Delta x(t_k)-\Delta x(t_{k-1})}{\Delta t_k}=\alpha_0+\alpha_1 x(t_k) \\ \qquad +\alpha_2\dfrac{\Delta x(t_k)}{\Delta t_k}(k)+\varepsilon_k \end{cases}$
B	$\begin{cases} \dfrac{dx}{dt}=\alpha_0 e^{\delta t}+\alpha_1 x \\ \dfrac{\Delta x(t_k)}{\Delta t_k}=\alpha_0 e^{\delta t_k}+\alpha_1 x(t_k)+\varepsilon_k \end{cases}$	$\begin{cases} \dfrac{d^2x}{dt^2}=\alpha_0 e^{\delta t}+\alpha_1 x+\alpha_2\dfrac{dx}{dt} \\ \dfrac{\Delta x(t_k)-\Delta x(t_{k-1})}{\Delta t_k}=\alpha_0 e^{\delta t_k}+\alpha_1 x(t_k) \\ \qquad +\alpha_2\dfrac{\Delta x(t_k)}{\Delta t_k}+\varepsilon_k \end{cases}$
C	$\begin{cases} \dfrac{dx}{dt}=\alpha_0\sin(\omega t+\varpi)+\alpha_1 x \\ \dfrac{\Delta x(t_k)}{\Delta t_k}=\alpha_0\sin(\omega t_k+\varpi)+\alpha_1 x(t_k)+\varepsilon_k \end{cases}$	$\begin{cases} \dfrac{d^2x}{dt^2}=\alpha_0\sin(\omega t+\varpi)+\alpha_1 x+\alpha_2\dfrac{dx}{dt} \\ \dfrac{\Delta x(t_k)-\Delta x(t_{k-1})}{\Delta t_k}=\alpha_0\sin(\omega t_k+\varpi)+\alpha_1 x(t_k) \\ \qquad +\alpha_2\dfrac{\Delta x(t_k)}{\Delta t_k}+\varepsilon_k \end{cases}$
D	$\begin{cases} \dfrac{dx}{dt}=\alpha_0 e^{\delta t}\sin(\omega t+\varpi)+\alpha_1 x \\ \dfrac{\Delta x(t_k)}{\Delta t_k}=\alpha_0 e^{\delta t_k}\sin(\omega t_k+\varpi)+\alpha_1 x(t_k)+\varepsilon_k \end{cases}$	$\begin{cases} \dfrac{d^2x}{dt^2}=\alpha_0 e^{\delta t}\sin(\omega t+\varpi)+\alpha_1 x+\alpha_2\dfrac{dx}{dt} \\ \dfrac{\Delta x(t_k)-\Delta x(t_{k-1})}{\Delta t_k}=\alpha_0 e^{\delta t_k}\sin(\omega t_k+\varpi)+\alpha_1 x(t_k) \\ \qquad +\alpha_2\dfrac{\Delta x(t_k)}{\Delta t_k}+\varepsilon_k \end{cases}$
E*	$\begin{cases} \dfrac{dx}{dt}=\alpha_0 p_q(t)+\alpha_1 x \\ \dfrac{\Delta x(t_k)}{\Delta t_k}=\alpha_0 p_q(t_k)+\alpha_1 x(t_k)+\varepsilon_k \end{cases}$	$\begin{cases} \dfrac{d^2x}{dt^2}=\alpha_0 p_q(t)+\alpha_1 x+\alpha_2\dfrac{dx}{dt} \\ \dfrac{\Delta x(t_k)-\Delta x(t_{k-1})}{\Delta t_k}=\alpha_0 p_q(t_k)+\alpha_1 x(t_k) \\ \qquad +\alpha_2\dfrac{\Delta x(t_k)}{\Delta t_k}+\varepsilon_k \end{cases}$
F*	$\begin{cases} \dfrac{dx}{dt}=\alpha_0 e^{\delta t}p_q(t)+\alpha_1 x \\ \dfrac{\Delta x(t_k)}{\Delta t_k}=\alpha_0 e^{\delta t_k}p_q(t_k)+\alpha_1 x(t_k)+\varepsilon_k \end{cases}$	$\begin{cases} \dfrac{d^2x}{dt^2}=\alpha_0 e^{\delta t}p_q(t)+\alpha_1 x+\alpha_2\dfrac{dx}{dt} \\ \dfrac{\Delta x(t_k)-\Delta x(t_{k-1})}{\Delta t_k}=\alpha_0 e^{\delta t_k}p_q(t_k)+\alpha_1 x(t_k) \\ \qquad +\alpha_2\dfrac{\Delta x(t_k)}{\Delta t_k}+\varepsilon_k \end{cases}$
G*	$\begin{cases} \dfrac{dx}{dt}=\alpha_0 p_q(t)\sin(\omega t+\varpi)+\alpha_1 x \\ \dfrac{\Delta x(t_k)}{\Delta t_k}=\alpha_0 p_q(t_k)\sin(\omega t_k+\varpi)+\alpha_1 x(t_k)+\varepsilon_k \end{cases}$	$\begin{cases} \dfrac{d^2x}{dt^2}=\alpha_0 p_q(t)\sin(\omega t+\varpi)+\alpha_1 x+\alpha_2\dfrac{dx}{dt} \\ \dfrac{\Delta x(t_k)-\Delta x(t_{k-1})}{\Delta t_k}=\alpha_0 p_q(t_k)\sin(\omega t_k+\varpi)+\alpha_1 x(t_k) \\ \qquad +\alpha_2\dfrac{\Delta x(t_k)}{\Delta t_k}+\varepsilon_k \end{cases}$

Note: (*) involves a q^{th}-order polynomial function: $p_q(t)=p_0+p_1t+\cdots+p_qt^q$ $(q>1)$.

.

3 The Coupling Principle and Error Structure of a DEAR Model

Based on the discussions of DEAR model in Section 2, it is fairly clear that organically coupling a differential equation and a regression together does change the nature of data-oriented modeling exercises. Such coupling generates a different optimality and efficiency in terms of a small sample of observations. In the paper of Guo, Guo and Thiart [6], the nature of coupling a differential equation and regression was examined for the first time and accordingly summarized as the Coupling Principle. For an overall intuitive picture of the coupling principle, we list the relevant components and the discretization rule according to the coupling principle in Table 3.

Table 3: Coupling principle (Rule) in univariate first-order DEMR model

Term	Motivated DE	Coupled REG
DISCRETIZATION RULE BETWEEN MOTIVATED DE AND COUPLED REG		
Intrinsic feature	Continuous	Discrete
Independent Variable	t	$t_k,\ k=1,2,\cdots,n$
*1*st-order Derivative	$\frac{dx}{dt}$	$\hat{x}^{(0)}(t_k)=\frac{\Delta x^{(1)}(t_k)}{\Delta t_k}=\frac{x^{(1)}(t_k)-x^{(1)}(t_{k-1})}{t_k-t_{k-1}}$
*p*th-order Derivative	$\frac{d^{(p)}x(t)}{dt^p}$	$\hat{x}^{(-p-1)}(t_k)\triangleq\frac{\Delta\hat{x}^{(-p+1)}(t_k)}{\Delta t_k}=\frac{\hat{x}^{(-p+2)}(t_k)-x^{(-p+2)}(t_{k-1})}{t_k-t_{k-1}}\approx\left.\frac{d^{(p)}x}{dt^p}\right\|_{t=t_k}$
Primitive function	$x(t)$	$x(t_k)$
Model Formation	$\frac{dx(t)}{dt}=\alpha+\beta x(t)$	$x^{(0)}(t_k)=\alpha+\beta\hat{x}(t_k)+\varepsilon_k$ or $\frac{\Delta x(t_k)}{\Delta t_k}=\alpha+\beta x(t_k)+\varepsilon_k$

A fundamental idea here is that an approximated derivative of the dynamic law $x(t)$ is obtained by divided difference, and the primitive is linked to the partial sum as an approximation to the integration (i.e., the primitive function).

In the literature of classical statistical linear models, the error terms in a regression model, denoted as $\varepsilon_i,\ i=1,2,\cdots,n$, are typically assumed to have zero mean and constant variance, i.e., $\mathrm{E}[\varepsilon]=0$ and $\mathrm{Var}[\varepsilon]=\sigma^2$. For the convenience in hypothesis testing, $\varepsilon_i,\ i=1,2,\cdots,n$ are assumed to be sampled from a normal distribution with zero mean and constant variance, i.e., $N(0,\sigma^2)$.

However, we should be fully aware that when carrying the coupling principle back and forward between the associated differential equation and coupled regression, approximation errors are introduced. In other words, the discretization and alternation between the associated differential equation and the coupled regression, in terms of the Coupling Principle, will bring in new error which is different from the random sampling error of the $N(0,\sigma^2)$ variable. Use of a divided difference $\Delta x(t_k)/\Delta t_k=(x(t_k)-x(t_{k-1}))/(t_k-t_{k-1})$ to replace a derivative $(dx/dt)_{t=t_k}$ and use of the average accumulated partial sum $\hat{x}(t_k)$ to replace the primitive function $x(t_k)$ during the discretization between the associated differential equation and the coupled regression give rise to new error.

Our simulation studies have shown that the coupling-introduced error is dependent upon the grid size h, or equivalently upon the total number of approximations *N*. The simulation evidence has shown that the larger the count of nodes on the approximating grid, or equivalently, the smaller the approximating grid size, the smaller the coupling discretization error. However, the coupling discretization error and the approximating grid do not satisfy a deterministic functional relation. What we can see is the functional relation has a particular degree of belongingness. In other words, the coupling discretization process induces a fuzzy error term, denoted as $\tilde{e}$ with a credibility distribution function $\Lambda_{\tilde{e}}(\cdot)$.

For example, a simulation study of the error occurrence frequencies when approximating $\cos(\pi/2)$ by $(\sin(\pi/2)-\sin(\pi/2+\Delta x))/\Delta x$, gave the following graph.

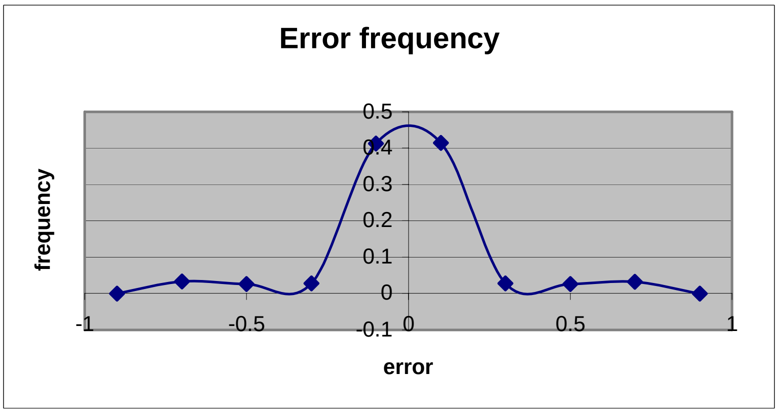

Figure 4: Error occurrence frequency

In general the discretization error term of a DEAR model is fuzzy because of the stratification feature and the vague nature of the error occurrences. In this sense, the coupled regression in Eq. (4b) and Eq. (5b) are actually cases of a special random fuzzy regression model.

As a standard approach, the fuzzy error component e_i may be assumed to be a triangular fuzzy variable with a membership function

$$\mu_e(s) = \begin{cases} \dfrac{s+g}{g} & \text{if } -g \le s < 0 \\ \dfrac{g-s}{g} & \text{if } 0 \le s \le g \\ 0 & \text{otherwise,} \end{cases}$$

which has a fuzzy mean zero. Therefore, the composite error term appearing in the differential equation associated regression is $\varepsilon_i = e_i + \tau_i$, $i = 2,3,\cdots,n$, which is a sequence of random fuzzy variables, because of the nature of the summation of a fuzzy variable, e, and a random variable, τ with $\mathrm{E}[\tau]=0$ and $\mathrm{Var}[\tau]=\sigma^2$, according to Liu [13,14].

Finally, we need also stress that the coupling-generated errors need not be a zero mean fuzzy variable. In general, we may assume a triangular membership function specified by parameters (a,b,c) as shown in Eq. (A8) in the Appendix.

4 An M-Estimator for a Univariate DEAR Model

We emphasize here that the coupled regression is in nature a random fuzzy regression model. The literature is limited but very complicated when the random fuzzy regression model is handled at random fuzzy set level, for example, Bandemer [1], Bardossy [2], Diamond [4], Kacprzyk and Fredrizzi [11], Korner [12], Tanaka and Watada [16], and others. However, based on the credibility measure theoretical foundation developed by Liu and Liu [15], Liu [13,14], a random fuzzy variable is defined as a mapping from the credibility space $(\Theta, 2^{\Theta}, \mathrm{Cr})$ to a set of random variables and thus the random fuzzy regression modeling will be handled at (scalar) functional level. The fundamental change in modeling level will lead to random fuzzy regression, highly similar to that in statistical regression models (or linear models based on probabilistic foundation) rather than these developments at set level.

As we pointed out in Section 3, the coupled regression either in Type I DEAR model or Type II DEAR model is random fuzzy regression in nature and thus the usual statistical regression theory should not be directly utilized.

Without loss of generality, we use the coupled regression in a simplest DEAR model form $Y_i = \alpha + \beta x_i + \varepsilon_i$, and a differentiable function of the random fuzzy error term is assumed to take a form

$$\Psi(\varepsilon_i) = g\left(-w\left(Y_i - (\alpha + \beta x_i)\right)^2\right).$$

Definition 4.1(*M*-function for random fuzzy variable): Let $\{x_1, x_2, \cdots, x_n\}$ be a simple random fuzzy data sampled drawing from a given random fuzzy population, where parameter-vector $\underline{\theta}^T = (\theta_1, \theta_2, \cdots, \theta_d)$, $d \geq 1$ Then the *M*-function is defined by

$$\mathbb{C}(\theta_1, \theta_2, \cdots, \theta_d \mid \{x_1, x_2, \cdots, x_n\} = \sum_{i=1}^{n} \Psi(x_i \mid \theta_1, \theta_2, \cdots, \theta_d)$$

For a given simple random sample $\{x_1, x_2, \cdots, x_n\}$, the optimal data-assimilated parameter-vector $\underline{\theta}^T = (\theta_1, \theta_2, \cdots, \theta_d)$ maximizes the *M*-function.

Mathematically, the parameter search can be performed by solving a nonlinear *M*-function equation system, which is called an average chance *M*-function

$$\begin{cases} \dfrac{\partial \mathbb{C}}{\partial \theta_1} = \displaystyle\sum_{i=1}^{n} \dfrac{\partial \Psi(x_i; \underline{\theta})}{\partial \theta_1} = 0 \\ \quad \vdots \\ \dfrac{\partial \mathbb{C}}{\partial \theta_d} = \displaystyle\sum_{i=1}^{n} \dfrac{\partial \Psi(x_i; \underline{\theta})}{\partial \theta_d} = 0. \end{cases} \tag{6}$$

The solution to the *M*-function equation system is called an *M*-estimator for $(\theta_1, \theta_2, \cdots, \theta_\gamma)^T$.

Theorem 4.2: The *M*-estimator for coupled regression coefficients $\underline{\Gamma} = (\alpha, \beta)^T$, denoted by $\hat{\underline{\Gamma}} = (\hat{\alpha}, \hat{\beta})^T$ is the solution to the *M*-functional equation system

$$\begin{cases} \displaystyle\sum_{i=1}^{n} \delta\left(-w\left(Y_i - (\hat{\alpha} + \hat{\beta} x_i)\right)^2\right)\left(Y_i - (\hat{\alpha} + \hat{\beta} x_i)\right) = 0 \\ \displaystyle\sum_{i=1}^{n} \delta\left(-w\left(Y_i - (\hat{\alpha} + \hat{\beta} x_i)\right)^2\right)\left(Y_i - (\hat{\alpha} + \hat{\beta} x_i)\right) x_i = 0, \end{cases} \tag{7}$$

where $\delta(\cdot)$ is the derivative of $g(\cdot)$.

The proof is very straightforward by taking the derivative of Eq. (6) with respect to α and β respectively. Equivalently, Eq. (7) can be written in a weighted least-squares normal equation

$$X^T W^{-1} X \underline{\Gamma} = X^T W^{-1} \underline{Y},$$

where $d_i = \delta\left(-w\left(Y_i - (a + b x_i)\right)^2\right)$ and

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \; X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \; W^{-1} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}.$$

Furthermore, the coefficient *M*-estimator will have weighted least square presentation form

$$\begin{cases} \hat{\alpha} = \overline{Y}_\delta - b\overline{x}_\delta \\ \hat{\beta} = \dfrac{\sum_{i=1}^{n} \delta\left(-w\left(Y_i - \left(a + bx_i\right)\right)^2\right)\left(x_i - \overline{x}_\delta\right)\left(Y_i - \overline{Y}_\delta\right)}{\sum_{i=1}^{n} \delta\left(-w\left(Y_i - \left(a + bx_i\right)\right)^2\right)\left(x_i - \overline{x}_\delta\right)^2}, \end{cases}$$

where the weighted averages are defined as

$$\begin{cases} \overline{x}_\delta = \sum_{i=1}^{n} \dfrac{\delta\left(-w\left(Y_i - \left(a + bx_i\right)\right)^2\right)}{\sum_{i=1}^{n} \delta\left(-w\left(Y_i - \left(a + bx_i\right)\right)^2\right)} x_i \\ \overline{Y}_\delta = \sum_{i=1}^{n} \dfrac{\delta\left(-w\left(Y_i - \left(a + bx_i\right)\right)^2\right)}{\sum_{i=1}^{n} \delta\left(-w\left(Y_i - \left(a + bx_i\right)\right)^2\right)} Y_i. \end{cases}$$

Theorem 4.3: The coupled regression coefficient *M*-estimator is unbiased, i.e., $\mathrm{E}\left[\hat{\alpha}\right] = \alpha$, $\mathrm{E}\left[\hat{\beta}\right] = \beta$.

The proof is divided into two steps. First setting

$$\kappa_i = \frac{\sum_{i=1}^{n} \delta\left(-w\left(Y_i - \left(a + bx_i\right)\right)^2\right)\left(x_i - \overline{x}_\delta\right)}{\sum_{i=1}^{n} \delta\left(-w\left(Y_i - \left(a + bx_i\right)\right)^2\right)\left(x_i - \overline{x}_\delta\right)^2}, \quad i = 1, 2, \cdots, n.$$

Then, writing $\hat{\beta}$ as $\hat{\beta} = \sum_{i=1}^{n} \kappa_i Y_i$. Finally, it is not difficult to show that $\mathrm{E}\left[\hat{\beta}\right] = \mathrm{E}\left[\sum_{i=1}^{n} \kappa_i Y_i\right] = \beta$.

Theorem 4.4: The estimated variance-covariance matrix of the regression coefficient *M*-estimators is given by

$$\hat{V}_0\left[\hat{\underline{\Gamma}}\right] = \hat{\sigma}^2 \left(X^T W^{-1} X\right)^{-1} X^T V[\underline{Y}] W^{-1} X \left[\left(X^T W^{-1} X\right)^{-1}\right]^T,$$

where

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} h\left(-\left(Y_i - \left(a + bx_i\right)\right)^2\right)\left(Y_i - \left(a + bx_i\right)\right)^2.$$

Let

$$\hat{V}_0\left[\hat{\underline{\Gamma}}\right] = \begin{bmatrix} \hat{\sigma}^2\left(\hat{\alpha}\right) & \hat{\sigma}\left(\hat{\alpha}, \hat{\beta}\right) \\ \hat{\sigma}\left(\hat{\alpha}, \hat{\beta}\right) & \hat{\sigma}^2\left(\hat{\beta}\right) \end{bmatrix}.$$

Corollary 4.5: The estimated variances and the covariance for *M*-estimator $\hat{\alpha}$ and $\hat{\beta}$ respectively, are

$$\hat{\sigma}^2\left(a\right) = \hat{\sigma}^2 \left[\sum_{i=1}^{n} \delta\left(-w\left(Y_i - \left(\hat{\alpha} + \hat{\beta} x_i\right)\right)^2\right)\right]^2$$
$$\times \frac{\sum_{i=1}^{n} \delta^2\left(-w\left(Y_i - \left(\hat{\alpha} + \hat{\beta} x_i\right)\right)^2\right)\left(\dfrac{\sum_{i=1}^{n} \delta\left(-w\left(Y_i - \left(\hat{\alpha} + \hat{\beta} x_i\right)\right)^2\right)x_i^2}{\sum_{i=1}^{n} \delta\left(-w\left(Y_i - \left(\hat{\alpha} + \hat{\beta} x_i\right)\right)^2\right)x_i} - x_i\right)^2}{\sum_{i=1}^{n} \delta\left(-w\left(Y_i - \left(\hat{\alpha} + \hat{\beta} x_i\right)\right)^2\right)\left(x_i - \overline{x}_\delta\right)^2},$$

and

$$\hat{\sigma}^2\left(\hat{\beta}\right)=\hat{\sigma}^2\frac{\sum_{i=1}^{n}\delta^2\left(-w\left(Y_i-\left(\hat{\alpha}+\hat{\beta}x_i\right)\right)^2\right)\left(x_i-\overline{x}_\delta\right)^2}{\sum_{i=1}^{n}\delta\left(-w\left(Y_i-\left(\hat{\alpha}+\hat{\beta}x_i\right)\right)^2\right)\left(x_i-\overline{x}_\delta\right)^2}, \tag{8}$$

and

$$\begin{aligned}\hat{\sigma}\left(\hat{\alpha},\hat{\beta}\right)=&\frac{s^2}{\left[\sum_{i=1}^{n}\delta\left(-w\left(Y_i-\left(\hat{\alpha}+\hat{\beta}x_i\right)\right)^2\right)\sum_{i=1}^{n}\delta\left(-w\left(Y_i-\left(\hat{\alpha}+\hat{\beta}x_i\right)\right)^2\right)\left(x_i-\overline{x}_\delta\right)^2\right]^2}\times\\
&\left[-\left(\sum_{i=1}^{n}\delta\left(-w\left(Y_i-\left(\hat{\alpha}+\hat{\beta}x_i\right)\right)^2\right)x_i^2\right)\times\left(\sum_{i=1}^{n}\delta^2\left(-w\left(Y_i-\left(\hat{\alpha}+\hat{\beta}x_i\right)\right)^2\right)\right)\times\left(\sum_{i=1}^{n}\delta\left(-w\left(Y_i-\left(\hat{\alpha}+\hat{\beta}x_i\right)\right)^2\right)x_i\right)\right.\\
&+\left(\sum_{i=1}^{n}\delta\left(-w\left(Y_i-\left(\hat{\alpha}+\hat{\beta}x_i\right)\right)^2\right)x_i\right)^2\times\left(\sum_{i=1}^{n}\delta^2\left(-w\left(Y_i-\left(\hat{\alpha}+\hat{\beta}x_i\right)\right)^2\right)x_i\right)\\
&+\left(\sum_{i=1}^{n}\delta\left(-w\left(Y_i-\left(\hat{\alpha}+\hat{\beta}x_i\right)\right)^2\right)x_i^2\right)\times\left(\sum_{i=1}^{n}\delta^2\left(-w\left(Y_i-\left(\hat{\alpha}+\hat{\beta}x_i\right)\right)^2\right)x_i\right)\times\left(\sum_{i=1}^{n}\delta\left(-w\left(Y_i-\left(\hat{\alpha}+\hat{\beta}x_i\right)\right)^2\right)\right)\\
&\left.-\left(\sum_{i=1}^{n}\delta\left(-w\left(Y_i-\left(\hat{\alpha}+\hat{\beta}x_i\right)\right)^2\right)x_i\right)\times\left(\sum_{i=1}^{n}\delta^2\left(-w\left(Y_i-\left(\hat{\alpha}+\hat{\beta}x_i\right)\right)^2\right)x_i^2\right)\times\left(\sum_{i=1}^{n}\delta\left(-w\left(Y_i-\left(\hat{\alpha}+\hat{\beta}x_i\right)\right)^2\right)\right)\right].\end{aligned}$$

Theorem 4.4 and Corollary 4.5 pave the way of prediction for a future Y_{n+k} value, denoted by $\hat{Y}_{n+k}$ given x_{n+k}, $k=1,2,\cdots,K$ as well as the variance of $\hat{Y}_{n+k}$.

5 The General Formation of a Mutivariate DEAR Model

A multivariate DEAR (m-dimensional) model may be stated as

$$\begin{cases}\underset{(m\times 1)}{\frac{d\underline{x}^{(1)}(t)}{dt}}=\underset{(m+1)\times m}{B}\,\underset{m\times 1}{\underline{x}^{(1)}}+\underset{m\times 1}{\underline{h}(t)} & (a)\\ \underset{((n-1)\times m)}{X^{(0)}}=\underset{((n-1)\times(m+1))}{Z^{(1)}}\,\underset{((m+1)\times m)}{B}+\underset{((n-1)\times m)}{\mathrm{E}}, & (b)\end{cases} \tag{9}$$

where Eq. (9a) is called a (multivariate) associated differential equation system and Eq. (9b) will be called a coupled multivariate regression model. The first-order vector differential equation in Eq. (9a) is

$$\begin{bmatrix}dx_1^{(1)}/dt\\ dx_2^{(1)}/dt\\ \vdots\\ dx_m^{(1)}/dt\end{bmatrix}=\begin{bmatrix}\beta_{01}&\beta_{21}&\cdots&\beta_{m1}\\ \beta_{02}&\beta_{22}&\cdots&\beta_{m2}\\ \vdots&\vdots&\ddots&\vdots\\ \beta_{0m}&\beta_{2m}&\cdots&\beta_{mm}\end{bmatrix}\begin{bmatrix}1\\ x_1^{(1)}\\ x_2^{(1)}\\ \vdots\\ x_m^{(1)}\end{bmatrix}.$$

In general, a multivariate regression model addresses the relationship between m response variables $Y_1,Y_2,\ldots,Y_m$ and a single set of explanatory variables $z_1, z_2, \ldots, z_m$. Each response variable is assumed to follow its own regression model so that

$$\begin{aligned}Y_1&=\beta_{01}+\beta_{11}z_1+\cdots+\beta_{p1}z_m+\varepsilon_1\\ Y_2&=\beta_{02}+\beta_{12}z_1+\cdots+\beta_{p2}z_m+\varepsilon_2\\ &\vdots\\ Y_m&=\beta_{0m}+\beta_{1m}z_1+\cdots+\beta_{pm}z_m+\varepsilon_m.\end{aligned}$$

The error term $\underline{\varepsilon}^T=\left(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_m\right)$ is assumed to follow $\mathrm{E}\left[\underline{\varepsilon}\right]=\underline{0}_{m\times 1}$ and $\mathrm{Var}\left[\underline{\varepsilon}\right]=\Sigma_{m\times m}$.

However, in the coupled multivariate regression model (8), the response variable Y_i is not an arbitrary one, but $Y_i = \frac{\Delta X(i)}{\Delta t(i)}$, $i = 1, 2, \cdots, m$, which involves the original observations. The explanatory variables $\underline{Z} = (Z_1, Z_2, \cdots, Z_m)^T$ are not arbitrarily chosen but are discretized approximate values of the related primitive functions $\underline{X}^{(1)} = \left(X_1^{(1)}, X_2^{(2)}, \cdots, X_m^{(1)}\right)^T$, i.e.,

$$\begin{cases} x_1(t_k) = \sum_{i=2}^{k} \left(\frac{\widehat{\Delta x_1(t_i)}}{\Delta t_i} \right) \Delta t_i \\ x_2(t_k) = \sum_{i=2}^{k} \left(\frac{\widehat{\Delta x_2(t_i)}}{\Delta t_i} \right) \Delta t_i \\ \vdots \\ x_m(t_k) = \sum_{i=2}^{k} \left(\frac{\widehat{\Delta x_m(t_i)}}{\Delta t_i} \right) \Delta t_i , \end{cases}$$

where $\Delta x_k(1) \equiv x_k(1)$, $k = 1, 2, \cdots, n$. Accordingly the approximate values of primitive functions are calculated as

$$\begin{cases} \hat{x}_1(t_k) = \left(x_1^{(1)}(t_k) + x_1^{(1)}(t_{k-1})\right) \big/ 2 \\ \hat{x}_2(t_k) = \left(x_2^{(1)}(t_k) + x_2^{(1)}(t_{k-1})\right) \big/ 2 \\ \vdots \\ \hat{x}_m(k) = \left(x_m^{(1)}(t_k) + x_m^{(1)}(t_{k-1})\right) \big/ 2, k = 2, 3, \cdots, n. \end{cases}$$

It is necessary to point out that in seeking a solution to Eq. (9a), it is inevitable that we invoke the finite term approximation of matrix e^{Bt}:

$$e^{Bt} = \kappa_0(t) I + \kappa_1(t) Bt + \kappa_2(t)(Bt)^2 + +\kappa_{n-1}(t)(Bt)^{n-1} .$$

To determine the functions $\{\kappa_0(t), \kappa_1(t), \cdots, \kappa_{n-1}(t)\}$, let us define a function

$$r(\lambda) = \kappa_0(t) + \kappa_1(t)\lambda + \kappa_2(t)\lambda^2 + +\kappa_{n-1}(t)\lambda^{n-1} . \quad (10)$$

Now let λ be an eigenvalue of matrix Bt, i.e., $|Bt - \lambda I| = 0$. Then $e^{\lambda_i} = r(\lambda_i)$.

Furthermore, if λ_i is an eigenvalue of multiplicity j, then the following equations hold

$$e^{\lambda_i} = \left[\frac{d^{(k)} r(\lambda)}{d\lambda^k} \right]_{\lambda=\lambda_i}, \quad k = 1, 2, \cdots, j .$$

When the equation system is established for each eigenvalue of matrix Bt, which comprises *n* linear equations, all including e^{λ_i} on the left side, then we can solve the linear equation system for $\{\kappa_0(t), \kappa_1(t), \cdots, \kappa_{n-1}(t)\}$. Section 6 details a special case.

6 A Bivariate DEAR Model Example

It is difficult to discuss the high-dimensional DEAR model in detail. However, a bivariate DEAR model is a quite manageable example for revealing the fundamental features.

6.1 Formation of General Bivariate DEAR Model

A bivariate DEAR model takes the form

$$\begin{cases} \dfrac{d\underline{x}^{(1)}(t)}{dt} = B\underline{x}^{(1)} + \underline{h}(t) & (a) \\ \underset{((n-1)\times 2)}{F} = \underset{((n-1)\times(3))}{Z}\underset{(3\times 2)}{\Lambda} + \underset{((n-1)\times 2)}{\mathrm{E}}, & (b) \end{cases} \tag{11}$$

where

$$\underline{x}^{(1)}(t) = \begin{bmatrix} x^{(1)}(t) \\ y^{(1)}(t) \end{bmatrix}. \tag{12}$$

The first vector-differential equation Eq. (11a) is called the bivariate associated differential equation system, while the second equation Eq. (11b) is called the coupled bivariate regression model.

6.2 Bivariate Differential Equation System

To obtain insight into the bivariate DEAR model, we start with a bivariate associated differential equation system in the same form in Eq. (11a)

$$\frac{d\underline{x}^{(1)}(t)}{dt} = B\underline{x}^{(1)} + \underline{h}(t). \tag{13}$$

A typical example is

$$\begin{cases} \dfrac{dx}{dt} = \alpha_1 + \beta_{11}x + \beta_{12}y \\ \dfrac{dy}{dt} = \alpha_2 + \beta_{21}x + \beta_{22}y. \end{cases} \tag{14}$$

Let

$$B = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix},\ \underline{a} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}.$$

Then the bivariate differential equation system in Eq. (14) can be re-written in a matrix form

$$\frac{d\underline{x}}{dt} = B\underline{x} + \underline{a}. \tag{15}$$

A common geometric interpretation of bivariate differential equation Eq. (12) is that under a two-dimensional coordinate system, a curve can be defined as $\{(x(t), y(t)), t \in \mathbb{T}\}$. It is obvious that in Eq. (10) the vector function $\underline{h}(t) = \underline{a}$ is a constant vector.

The differential equation system

$$\frac{d\underline{x}}{dt} = B\underline{x} \tag{16}$$

is called the homogeneous version of Eq. (13), and we investigate its solution matrix.

Lemma 6.1: Let *B* be a 2×2 matrix as

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}.$$

Define function $r(\lambda) = \gamma_0 + \gamma_1\lambda$, where λ_i is an eigenvalue of matrix Bt, then $e^{\lambda_i} = r(\lambda_i),\ i = 1, 2$. Furthermore, if $\lambda_1 = \lambda_2 = \lambda$, then

$$e^{\lambda_1} = \frac{d}{d\lambda} r(\lambda)\bigg|_{\lambda=\lambda_1}.$$

Lemma 6.2: If the eigenvalues of the matrix Bt are $\lambda_1 \neq \lambda_2$, then the equation system

$$\begin{cases} \gamma_0 + \gamma_1\lambda_1 = e^{\lambda_1} \\ \gamma_0 + \gamma_1\lambda_2 = e^{\lambda_2} \end{cases}$$

defines functions $\gamma_0(t)$ and $\gamma_1(t)$ as

$$\begin{cases} \gamma_0 = \dfrac{1}{\lambda_2 - \lambda_1}\left(\lambda_2 e^{\lambda_1} - \lambda_1 e^{\lambda_2}\right) \\ \gamma_1 = \dfrac{1}{\lambda_2 - \lambda_1}\left(e^{\lambda_2} - e^{\lambda_1}\right), \end{cases} \tag{17}$$

where λ_1 and λ_2 are eigenvalues of matrix *Bt* with

$$\begin{cases} \lambda_1(t) = \left((b_{11} - b_{22}) + \Delta\right)t/2 \\ \lambda_2(t) = \left((b_{11} - b_{22}) - \Delta\right)t/2 \\ \Delta = \sqrt{(b_{11} - b_{22})^2 + 4b_{12}b_{21}} \end{cases}.$$

If the eigenvalues of matrix Bt are $\lambda_1(t) = \lambda_2(t) = \lambda(t)$, then the equation system

$$\begin{cases} \gamma_0(t) + \gamma_1(t)\lambda(t) = e^{\lambda(t)} \\ \gamma_1(t) = e^{\lambda(t)} \end{cases}$$

will define functions γ_0 and γ_1 as

$$\gamma_0 = (1-\lambda)e^{\lambda}, \gamma_1 = e^{\lambda} , \tag{18}$$

respectively. Use of Lemmas 6.1 and 6.2, allow us to obtain a closed form expression for matrix e^{Bt}. We state the result as a theorem.

Theorem 6.3: The homogeneous bivariate linear differential equation system in Eq. (16) has an *elementary solution matrix*

$$X(t) = e^{Bt} = \gamma_0 I + \gamma_1 Bt ,$$

where functions $\gamma_0(t)$ and $\gamma_1(t)$ are defined in Eq. (17), or Eq. (18) respectively.

We proceed to the solution for the nonhomogeneous bivariate differential equation system.

Theorem 6.4: For the bivariate differential equation system in Eq. (13), the general solution to the non-homogeneous system takes the form

$$\underline{x}(t) = e^{Bt}\underline{c} + \int_{t_0}^{t} e^{B(t-s)}\underline{h}(s)\,ds ,$$

where $\underline{c} = [c_1 \quad c_2]^T$ is an arbitrary constant vector.

Example 6.5: Let $\underline{h}(t) = \underline{a} = [\alpha_1 \quad \alpha_2]^T$. Then Eq. (13) becomes Eq. (15). The general solution becomes

$$\underline{x}(t) = e^{Bt}\underline{c} + B^{-1}\left(e^{Bt} - e^{Bt_0}\right)\underline{a} ,$$

where $\underline{c} = [c_1 \quad c_2]^T$ is an arbitrary constant vector. However**,** for 2×2 matrix *Bt***,**

$$e^{Bt} = \gamma_0(t) I + \gamma_1(t) Bt .$$

If $\lambda_1 \neq \lambda_2$, then

$$\begin{aligned} \underline{x}(t) &= e^{Bt}\underline{c} + \int_{t_0}^{t} e^{B(t-s)}\underline{a}\,ds = e^{Bt}\underline{c} + \left(\int_{t_0}^{t} e^{B(t-s)}ds\right)\underline{a} \\ &= e^{Bt}\underline{c} + \left(\int_{t_0}^{t} \left(\gamma_0(t-s)I + \gamma_1(t-s)(t-s)B\right)ds\right)\underline{a} \\ &= e^{Bt}\underline{c} + \left(\int_{t_0}^{t} \gamma_0(t-s)\,ds\right)\underline{a} + \left(\int_{t_0}^{t} \left(\gamma_1(t-s)(t-s)\right)ds\right)B\underline{a}. \end{aligned}$$

Note that

$$\int_{t_0}^{t} \gamma_0(t-s)\,ds = \frac{\theta_2}{\pi_1}\left(e^{\pi_1(t-t_0)}-1\right)+\frac{\theta_1}{\pi_2}\left(e^{\pi_2(t-t_0)}-1\right)$$

and

$$\int_{t_0}^{t} \gamma_1(t-s)(t-s)\,ds = -\frac{1}{\Delta\pi_2}\left(e^{\pi_2(t-t_0)}-1\right)+\frac{1}{\Delta\pi_1}\left(e^{\pi_1(t-t_0)}-1\right),$$

where

$$\pi_1=\left((b_{11}-b_{22})+\Delta\right)/2,\ \pi_2=\left((b_{11}-b_{22})-\Delta\right)/2,\ \theta_1=\frac{\pi_1}{-\Delta},\ \theta_2=\frac{\pi_2}{-\Delta}.$$

6.3 The Coupled Bivariate Regression Model

In the bivariate DEAR model of Eq. (11), the second equation system in Eq. (11b) is just the coupled bivariate regression model by assuming equal-interval observation times),

$$\begin{cases} x^{(0)}(t_k)=\alpha_1+\beta_{11}z_1^{(1)}(t_k)+\beta_{12}z_2^{(1)}(t_k)+\varepsilon_1(k) \\ y^{(0)}(t_k)=\alpha_2+\beta_{21}z_1^{(1)}(t_k)+\beta_{22}z_2^{(1)}(t_k)+\varepsilon_2(k) \end{cases} \quad for \quad k=2,3,\cdots,n.$$

In contrast to the univariate case, the observations are data pairs ($X^{(0)},Y^{(0)}$)={($x^{(0)}(i),y^{(0)}(i)$)}, $i=1,2,\ldots,n$}. Then the approximate primitive function pairs will be generated in a manner similar to the univariate DEAR model case,

$$\begin{cases} x^{(1)}(t_k)=\sum_{i=1}^{k} x^{(0)}(t_i)\Delta t_i \\ y^{(1)}(k)=\sum_{i=1}^{k} y^{(0)}(t_i)\Delta t_i. \end{cases}$$

Thus, the mean pair sequence will be generated as

$$\begin{cases} z_x^{(1)}(t_k)=\left(x^{(1)}(t_k)+x^{(1)}(t_{k-1})\right)/2 \\ z_y^{(1)}(t_k)=\left(y^{(1)}(t_k)+y^{(1)}(t_{k-1})\right)/2 \end{cases} \quad for \quad k=2,3,\cdots,n.$$

With the data as handled by linear transformation of the original data observation vector, we are ready to explain the discretization from the associated differential equation systems into the coupling multivariate regression model. The bivariate (coupling) regression model here suggests the relationship between bivariate responses, $\underline{x}^{(0)},\underline{y}^{(0)}$ and a single set of bivariate predictors, $\underline{z}_1^{(1)},\underline{z}_2^{(1)}$. Each response is assumed to follow its own regression model.

In classical multivariate regression theory, the error term $\underline{\varepsilon}=(\varepsilon_1,\varepsilon_2)^T$ has $\mathrm{E}[\underline{\varepsilon}]=\underline{0}$ and $\mathrm{Var}[\underline{\varepsilon}]=\Sigma$, which indicates the error terms associated with different responses might be correlated.

Let $\left[z_1^{(1)}(k),z_2^{(1)}(k)\right]$ be the values of the bivariate predictors for the k^{th} observed pair in a sequence, and

$$\underline{f}(k)=\left(x^{(0)}(k),y^{(0)}(k)\right)^T,\ \underline{\varepsilon}(k)=\left(\varepsilon_1(k),\varepsilon_2(k)\right)^T$$

the values of responses and errors in k^{th} observation. Then in matrix notation, the "design matrix" takes the form

$$Z_{(n-1)\times 3}=\begin{bmatrix} 1 & z_1^{(1)}(t_2) & z_2^{(1)}(t_2) \\ 1 & z_1^{(1)}(t_3) & z_2^{(1)}(t_3) \\ \vdots & \ddots & \vdots \\ 1 & z_1^{(1)}(t_n) & z_2^{(1)}(t_n) \end{bmatrix}.$$

The response matrix is

$$F_{(n-1)\times 2} = \begin{bmatrix} x^{(0)}(t_2) & y^{(0)}(t_2) \\ x^{(0)}(t_3) & y^{(0)}(t_3) \\ \vdots & \vdots \\ x^{(0)}(t_n) & y^{(0)}(t_n) \end{bmatrix} = \begin{bmatrix} \underline{x}^{(0)} & \underline{y}^{(0)} \end{bmatrix}.$$

The parameter matrix is

$$\Lambda_{3\times 2} = \begin{bmatrix} \alpha_1 & \alpha_2 \\ \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \end{bmatrix} = \begin{bmatrix} \underline{\beta}_1 & \underline{\beta}_2 \end{bmatrix},$$

and finally the error matrix is

$$\mathrm{E}_{(n-1)\times 2} = \begin{bmatrix} \varepsilon_{21} & \varepsilon_{22} \\ \varepsilon_{31} & \varepsilon_{32} \\ \vdots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} \end{bmatrix} = \begin{bmatrix} \underline{\varepsilon}_1 & \underline{\varepsilon}_2 \end{bmatrix} = \begin{bmatrix} \underline{\varepsilon}^T(2) \\ \underline{\varepsilon}^T(3) \\ \vdots \\ \underline{\varepsilon}^T(n) \end{bmatrix}.$$

Then the coupled bivariate linear regression model in matrix form is

$$\underset{((n-1)\times 2)}{F} = \underset{((n-1)\times 3)}{Z}\underset{(3\times 2)}{\Lambda} + \underset{((n-1)\times 2)}{\mathrm{E}}.$$

It is obvious that the both response vector $\underline{x}^{(0)} = Z\underline{\beta}_1 + \underline{\varepsilon}_1$ and response vector $\underline{y}^{(0)} = Z\underline{\beta}_2 + \underline{\varepsilon}_2$ follow their own bivariate linear regressions respectively.

Based the observed data sequence pair $(X^{(0)}, Y^{(0)})$, the weighted least-squares estimates $\hat{\underline{\beta}}_1$ are obtained

$$\hat{\underline{\beta}}_1 = \left(Z^T W^{-1} Z\right)^{-1} Z^T W^{-1} \underline{x}^{(0)}.$$

Similarly,

$$\hat{\underline{\beta}}_2 = \left(Z^T W^{-1} Z\right)^{-1} Z^T W^{-1} \underline{y}^{(0)}.$$

We combine the two least-squares estimates as

$$\hat{\Lambda} = \begin{bmatrix} \hat{\underline{\beta}}_1 & \hat{\underline{\beta}}_2 \end{bmatrix} = \left(Z^T W^{-1} Z\right)^{-1} Z^T W^{-1} \begin{bmatrix} \underline{x}^{(0)} & \underline{y}^{(0)} \end{bmatrix} = \left(Z^T W^{-1} Z\right)^{-1} Z^T W^{-1} F.$$

For any choice of parameter $\mathrm{B} = [\underline{b}_1 \quad \underline{b}_2]$, the matrix of weighted least-squared error is $G(F - Z\mathrm{B})$, where $W^{-1} = G^T G$. Therefore, the error matrix for sum of squares and the cross-products is

$$\left(G(F - Z\mathrm{B})\right)^T \left(G(F - Z\mathrm{B})\right) = \begin{bmatrix} \left(\underline{x}^{(0)} - Z\underline{b}_1\right)^T W^{-1}\left(\underline{x}^{(0)} - Z\underline{b}_1\right) & \left(\underline{x}^{(0)} - Z\underline{b}_1\right)^T W^{-1}\left(\underline{y}^{(0)} - Z\underline{b}_2\right) \\ \left(\underline{y}^{(0)} - Z\underline{b}_2\right)^T W^{-1}\left(\underline{x}^{(0)} - Z\underline{b}_1\right) & \left(\underline{y}^{(0)} - Z\underline{b}_2\right)^T W^{-1}\left(\underline{y}^{(0)} - Z\underline{b}_2\right) \end{bmatrix}.$$

The estimate of parameter matrix Λ actually minimizes the trace of the weighted matrix

$$(F - Z\mathrm{B})^T W^{-1} (F - Z\mathrm{B}),$$

i.e.,

$$\hat{\Lambda} = \min_B \mathrm{tr}\left\{(F - Z\mathrm{B})^T W^{-1}(F - Z\mathrm{B})\right\}$$

and it can be shown that the generalized variance $\left|(F - Z\mathrm{B})^T W^{-1}(F - Z\mathrm{B})\right|$ is minimized by the weighted least-squares estimates $\hat{\Lambda}$. Then the predicted values are

$$\hat{F} = Z\hat{\Lambda} = Z\left(Z^T W^{-1} Z\right)^{-1} Z^T W^{-1} F.$$

As to the weighted matrix W, it is necessary to mention here that W depends upon the object function to be used.

7 Conclusion

In this paper, we propose a new modeling family, called the DEAR. We discuss the mathematical foundation for a DEAR model, which combines the (ordinary) differential equation theory, (statistical) linear model theory and random fuzzy variable theory based on credibility measure and probability measure foundations, into a new small-sample oriented prediction theory. The coupled regression component in a DEAR model is in nature a special random fuzzy regression model. The multivariate DEAR model is also introduced and as illustration, a bivariate DEAR model is explored in detail to offer further insight into multivariate DEAR modeling.

Acknowledgements

This research is partially supported by South African National Research Foundation Grant FA2006042800024.

References

[1] Bandemer, H., Evaluating explicit functional relationship from fuzzy observations, *Fuzzy Sets and Systems*, vol.16, pp.41-52, 1985.
[2] Bardossy, A., Note on fuzzy regression, *Fuzzy Sets and Systems*, vol.37, pp.65-75, 1990.
[3] Deng, J.L., *Grey Systems (Social and Economical)*, The Publishing House of Defense Industry, Beijing, China (*in Chinese*), 1985.
[4] Diamond, P., Interval-valued random functions and the kriging of intervals, *Mathematical Geology*, vol.20, no.3, pp.145-165, 1988.
[5] Diamond, P., *Least Squares and Maximum Likelihood Regression for Fuzzy Linear Models*, Fuzzy Regression Analysis, Physica-Verlag, Heidelberg, 1992.
[6] Guo, D., R. Guo, and C. Thiart, , The Coupling of regression modeling and differential equation model in GM(1,1) modeling and extended GM(1,1) models, *Journal of Grey System*, vol.9, no.2, pp.143-154, 2006.
[7] Guo, D., R. Guo, G.F. Midgley, and I. Ebrahim, DEMR modeling Brunsvigia Litoralis plant distribution in 2004, western region district council, South Africa, *Proceedings of the Sixth International Conference on Information and Management Sciences*, Lhasa, Tibet, China, pp.770-782, 2007.
[8] Guo, R., and D. Guo, Univariate DEMR model-the formation and foundation, *Proceedings of IEEE International Conference on Grey Systems and Intelligent Services*, Nanjing, China, pp 13-18, 2007.
[9] Guo, R., and D. Guo, Multivariate DEMR model-the formation and foundation, *Proceedings of IEEE International Conference on Grey Systems and Intelligent Services*, Nanjing, China, pp 19-25, 2007.
[10] Guo, R., D. Guo, G.F. Midgley, and A.G. Rebelo, Bivariate DEMR modeling of range shifts in Aulax Umbellata Shrubs from 1992 to 2002 in South Africa, *Proceedings of the Sixth International Conference on Information and Management Sciences*, Lhasa, Tibet, China, pp.783-791, 2007.
[11] Kacprzyk, J., and M. Fredrizzi, *Fuzzy Regression Analysis*, Physica-Verlag, Heidelberg, 1992.
[12] Korner, W., Linear statistical inference for random fuzzy data, *Statistics*, vol.29, pp.221-240, 1997.
[13] Liu, B., *Uncertainty Theory: An Introduction to Its Axiomatic Foundations*, Springer-Verlag Heidelberg, Berlin, 2004.
[14] Liu, B., *Uncertainty Theory*, 2nd Edition. Springer-Verlag, Heidelberg, Berlin, 2007.
[15] Liu, Y.K., and B. Liu, Random fuzzy programming with chance measures defined by fuzzy integrals, *Mathematical and Computer Modeling*, vol.36, nos.4-5, pp.509-524, 2002.
[16] Tanaka, H., and J. Watada, Possibilistic linear systems and their applications to the linear regression model, *Fuzzy Sets and Systems*, vol.27, pp.275-289, 1988.
[17] Zadeh, L.A., Fuzzy sets, *Information and Control*, vol.8, pp.338-353, 1965.
[18] Zadeh, L.A., Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems*, vol.1, pp.3-28, 1978.

Appendix: A Review on Axiomatic Fuzzy Credibility Measure Theory

A.1 Axiomatic Fuzzy Credibility Measure Theory

Let Θ be a nonempty set, and 2^{Θ} the power set on Θ. Each element, let us say, $A \subset \Theta$, $A \in 2^{\Theta}$ is called an event. A number denoted as $\mathrm{Cr}\{A\}$, $0 \le \mathrm{Cr}\{A\} \le 1$, is assigned to event $A \in 2^{\Theta}$, which indicates the credibility grade with which event $A \in 2^{\Theta}$ occurs. $\mathrm{Cr}\{A\}$ satisfies the following axioms (Liu, 2004, 2007):

Axiom 1: $\mathrm{Cr}\{\Theta\} = 1$.

Axiom 2: $\mathrm{Cr}\{\cdot\}$ is non-decreasing, i.e., $\mathrm{Cr}\{A\} \le \mathrm{Cr}\{B\}$ whenever $A \subset B$.

Axiom 3: $\mathrm{Cr}\{\cdot\}$ is self-dual, i.e., $\mathrm{Cr}\{A\} + \mathrm{Cr}\{A^c\} = 1$ for any $A \in 2^{\Theta}$.

Axiom 4: $\mathrm{Cr}\{\cup_i A_i\} = \sup_i \left[\mathrm{Cr}\{A_i\}\right]$ for any $\{A_i\}$ with $\sup_i \left[\mathrm{Cr}\{A_i\}\right] < 0.5$.

Axiom 5: Let set functions $\mathrm{Cr}_k\{\cdot\}: 2^{\Theta_k} \to [0,1]$ satisfy Axioms 1-4, and $\Theta = \Theta_1 \times \Theta_2 \times \cdots \times \Theta_p$, then

$$\mathrm{Cr}\{\theta_1, \theta_2, \cdots, \theta_p\} = \mathrm{Cr}_1\{\theta_1\} \wedge \mathrm{Cr}_2\{\theta_2\} \wedge \cdots \wedge \mathrm{Cr}_p\{\theta_p\}$$

for each $\{\theta_1, \theta_2, \cdots, \theta_p\} \in 2^\Theta$.

Definition A.1 ([13,14]): Any set function $\mathrm{Cr}: 2^\Theta \to [0,1]$ satisfies Axioms **1-4** is called a $(\vee, \wedge)$ -credibility measure (or classical credibility measure). The triple $(\Theta, 2^\Theta, \mathrm{Cr})$ is called the $(\vee, \wedge)$ -credibility measure space.

Definition A.2 ([13,14]): A fuzzy variable ξ is a mapping from credibility space $(\Theta, 2^\Theta, \mathrm{Cr})$ to the set of real numbers.

Similar to random variable, a fuzzy variable is fully specified by its distribution function.

Definition A.3 ([13,14]): The credibility distribution $\Phi: \mathbb{R} \to [0,1]$ of a fuzzy variable ξ on $(\Theta, 2^\Theta, \mathrm{Cr})$ is $\Lambda(x) = \mathrm{Cr}\{\theta \in \Theta | \xi(\theta) \le x\}$.

The credibility distribution $\Lambda(x)$ is the accumulated credibility grade that the fuzzy variable ξ takes a value less than or equal to a real-number $x \in \mathbb{R}$.

Definition A.4 [13,14]: Let Φ be the credibility distribution of the fuzzy variable ξ . Then function $\lambda: \mathbb{R} \to [0, +\infty)$ of a fuzzy variable ξ is called a credibility density function such that

$$\Lambda(x) = \int_{-\infty}^{x} \lambda(y) \mathrm{d}y, \quad \forall x \in \mathbb{R} .$$

Zadeh [17] based his fuzzy set concept upon the membership function, which is intuitive and sounds very practical. However, the membership function is not the correct starting point for establishing a set-theoretical foundation of the fuzzy mathematics. Some later developments on possibility measure theory, which was assumed to be the counterpart of probability measure theory, however, failed to behave as expected by Zadeh [18]. In contrast, the axioms for the credibility measure proposed by Liu [13,14] have introduced a set-theoretical foundation. As a traditional treatment, whenever a fuzzy variable is involved, its membership function is given. One should be fully aware that this treatment may not be the natural way to deal with a fuzzy variable. On credibility measure theoretical grounds, a fuzzy variable should be characterized by its credibility distribution first. The corresponding membership is merely an *induced* function and a conventional and convenient mathematical language for describing the fuzzy phenomenon. The credibility measure of an event permits many developments related to the membership function.

Definition A.5 ([13,14]): The (induced) membership function of a fuzzy variable ξ on $(\Theta, 2^\Theta, \mathrm{Cr})$ is

$$\mu(x) = (2\mathrm{Cr}\{\xi = x\}) \wedge 1, \quad x \in \mathbb{R} .$$

Conversely, for a given membership function the credibility measure is determined by the credibility inversion theorem.

Theorem A.6 ([13,14]): Let ξ be a fuzzy variable with membership function μ . Then for $\forall B \subset \mathbb{R}$,

$$\mathrm{Cr}\{\xi \in B\} = \frac{1}{2}\left(\sup_{x \in B} \mu(x) + 1 - \sup_{x \in B^c} \mu(x) \right), \quad B \subset \mathbb{R} .$$

As an example, if the set *B* is degenerate at a point *x*, then

$$\mathrm{Cr}\{\xi = x\} = \frac{1}{2}\left(\mu(x) + 1 - \sup_{y \neq x} \mu(y) \right), \quad \forall x \in \mathbb{R} .$$

Theorem A.7 ([13,14]): Let ξ be a fuzzy variable on $(\Theta, 2^\Theta, \mathrm{Cr})$ with membership function μ. Then its credibility distribution is

$$\Lambda(x) = \frac{1}{2}\left(\sup_{y \le x} \mu(y) + 1 - \sup_{y > x} \mu(y) \right), \ \forall x \in \mathbb{R} .$$

An important class of fuzzy variables is defined by a triangular credibility distribution with three parameters (a, b, c)

$$\Lambda(x) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{2(b-a)} & \text{if } a \le x < b \\ \dfrac{x+c-2b}{2(c-b)} & \text{if } b \le x < c \\ 1 & \text{if } x \ge c. \end{cases}$$

A.2 Random Fuzzy Variable

Liu [13,14] stated that a random fuzzy variable is a mapping from the credibility space $(\Theta, 2^{\Theta}, \mathrm{Cr})$ to a set of random variables. We present a constructive definition.

Definition A.8: A random fuzzy variable, denoted as $\xi = \{X_{\beta(\theta)}, \theta \in \Theta\}$, is a collection of random variables X_β defined on the common probability space $(\Omega, \mathfrak{A}, \Pr)$ and indexed by a fuzzy variable $\beta(\theta)$ defined on the credibility space $(\Theta, 2^{\Theta}, \mathrm{Cr})$.

A random fuzzy variable is a bivariate mapping from $(\Omega \times \Theta, \mathfrak{A} \times 2^{\Theta})$ to the space $(\mathbb{R}, \mathfrak{B})$.

In Liu [13,14], random fuzzy variable theory, we may say that the average chance measure plays an equivalent role similar to that of a probability measure, denoted as $\Pr$, in probability theory.

Definition A.9 ([15]): Let ξ be a random fuzzy variable. Then the average chance measure denoted by $\mathrm{ch}\{\cdot\}$, of a random fuzzy event $\{\xi \le x\}$, is

$$\mathrm{ch}\{\xi \le x\} = \int_0^1 \mathrm{Cr}\{\theta \in \Theta | \Pr\{\xi(\theta) \le x\} \ge \alpha\} \mathrm{d}\alpha .$$

Then function $\Psi(\cdot)$ is called as average chance distribution if and only if $\Psi(x) = \mathrm{ch}\{\xi \le x\}$.

A function $\phi: \mathbb{R} \to \mathbb{R}^+$ such that $\Psi(x) = \int_{-\infty}^{x} \phi(u) du$ is called the average chance density function of random fuzzy variable ξ.

A.3 Normal Random Fuzzy Variable with Triangular Fuzzy Parameter

Let ϕ and Φ be the density and cdf of the standard normal random variable, respectively. Then, for a normal random fuzzy variable with a triangular fuzzy mean defined by parameters (a,b,c), the average chance distribution for the mean is

$$\Psi(x) = \frac{x-a}{2(b-a)}\left(\Phi\left(\frac{x-a}{\sigma}\right) - \Phi\left(\frac{x-b}{\sigma}\right)\right) + \frac{x+c-2b}{2(c-b)}\left(\Phi\left(\frac{x-b}{\sigma}\right) - \Phi\left(\frac{x-c}{\sigma}\right)\right)$$
$$+ \Phi\left(\frac{x-c}{\sigma}\right) - \frac{\sigma}{2(b-a)} \int_{\frac{x-b}{\sigma}}^{\frac{x-a}{\sigma}} u\phi(u)\mathrm{d}u - \frac{\sigma}{2(c-b)} \int_{\frac{x-c}{\sigma}}^{\frac{x-b}{\sigma}} u\phi(u)\mathrm{d}u .$$

The average chance density is

$$\psi(x) = \frac{1}{2(b-a)}\left(\Phi\left(\frac{x-a}{\sigma}\right) - \Phi\left(\frac{x-b}{\sigma}\right)\right) + \frac{x-a}{2(b-a)\sigma}\left(\phi\left(\frac{x-a}{\sigma}\right) - \phi\left(\frac{x-b}{\sigma}\right)\right)$$
$$+ \frac{1}{2(c-b)}\left(\Phi\left(\frac{x-b}{\sigma}\right) - \Phi\left(\frac{x-c}{\sigma}\right)\right) + \frac{x+c-2b}{2(c-b)\sigma}\left(\phi\left(\frac{x-b}{\sigma}\right) - \phi\left(\frac{x-c}{\sigma}\right)\right) + \frac{1}{\sigma}\phi\left(\frac{x-c}{\sigma}\right)$$
$$- \frac{1}{2(b-a)}\left(\frac{x-a}{\sigma}\phi(\frac{x-a}{\sigma}) - \frac{x-b}{\sigma}\phi(\frac{x-b}{\sigma})\right) - \frac{1}{2(c-b)}\left(\frac{x-b}{\sigma}\phi(\frac{x-b}{\sigma}) - \frac{x-c}{\sigma}\phi(\frac{x-c}{\sigma})\right)$$

from the differentiation formula of an integral

$$\frac{d}{dt}\left(\int_{a(t)}^{b(t)} f(x,t)\,dx\right) = \int_{a(t)}^{b(t)} \frac{\partial f(x,t)}{\partial t}\,dx + f\big(b(t),t\big)\frac{db(t)}{dt} - f\big(a(t),t\big)\frac{da(t)}{dt}.$$

Because sampling distributions are critical for the construction of hypothesis tests, we will address sampling of average chance distributions.

The sample mean, denoted as $x = \overline{\xi} = \sum_{i=1}^{n} x_i \Big/ n$, is a normal random fuzzy variable and the average chance distribution of the sampling mean can be obtained by substituting $\sigma/\sqrt{n}$ in place of σ. The reason is obvious because

$$\Psi(x) = \frac{x-a}{2(b-a)}\left(\Phi\left(\frac{x-a}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{x-b}{\sigma/\sqrt{n}}\right)\right) + \frac{x+c-2b}{2(c-b)}\left(\Phi\left(\frac{x-b}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{x-c}{\sigma/\sqrt{n}}\right)\right)$$
$$+ \Phi\left(\frac{x-c}{\sigma/\sqrt{n}}\right) - \frac{\sigma/\sqrt{n}}{2(b-a)}\int_{\frac{x-b}{\sigma/\sqrt{n}}}^{\frac{x-a}{\sigma/\sqrt{n}}} u\,\phi(u)\,\mathrm{d}u - \frac{\sigma/\sqrt{n}}{2(c-b)}\int_{\frac{x-c}{\sigma/\sqrt{n}}}^{\frac{x-b}{\sigma/\sqrt{n}}} u\,\phi(u)\,\mathrm{d}u.$$