The Korean Communications
in Statistics Vol. 5, No. 2, 1998
pp. 287-295

A Random Fuzzy Linear Regression Model[1)]

Changhyuck Oh[2)]

Abstract

A random fuzzy linear regression model is introduced, which includes both randomness and fuzziness. Estimators for the parameters are suggested, which are derived mainly using properties of randomness.

1. Introduction

Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{1}$$

where β's are crisp regression parameters, x_i are given crisp constants, and ε_i are independent and identically distributed normal random variables with mean 0 and variance σ^2. Residuals ε_i are introduced to represent deviations between observed values of Y_i and its expectations. In general it is considered that these deviations are results of omit of many factors from the model. Linear regression models are widely used in many fields. See Kutner et al.(1996). But when interval or fuzzy set data for the output are given, the classical regression model seems not appropriate.

Tanaka et al. (1982) assume that deviations between observed values and their expected values come from the fuzziness of the system parameters and introduce a fuzzy linear regression model. For one system variable, it can be written as

$$Y_i = A_0 + A_1 x_i, \quad i = 1, 2, \ldots, n \tag{2}$$

where A_0 and A_1 are triangular fuzzy sets, x_i's are given crisp numbers. And by properties of triangular fuzzy sets, Y_i is also a triangular fuzzy set. Since Tanaka et al. (1982) has

---

1) This research was supported by Korea Science and Engineering Foundation 1997, 971-0105-031-1
2) Professor, Department of Statistics, Yeungnam University, Kyungsan 712-749, Korea.

introduced the fuzzy linear regression, many works on it have been worked out. See Tanaka (1987), Tanaka and Watada (1988), Bardossy (1990), and Savic and Pedrycz (1991).
However, as Redden and Woodal(1994) point out, certain fuzzy regression models have an infinite number of solutions for parameter estimation, wider fuzzy linear regression intervals when there are more data and that it is hard to interpret fuzzy linear regression intervals.

On the other hand, Näther et al. (1990) consider a linear regression model which contains both fuzziness and randomness. When the simple linear regression is considered for the center points, it becomes

$$\tilde{Y}_i = [\, Y_i, \Delta_i]_L, \quad i = 1, 2, \ldots, n \tag{3}$$

where $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ is regression model, Y_i and Δ_i are independent and ε_i random variables with mean 0 and variance $\sigma^2(\varepsilon)$ and Δ_i are positive random variables with mean δ_i and variance $\sigma^2(\Delta_i) > 0$. Here ε_i and Δ_i are assumed to be independent. ε_i represents randomness of the location of the observation and Δ_i the fuzziness of the observation. And $[a, b]_L$ represents a symmetric LR-type fuzzy set with center a and width b. Näther et al. (1990) obtain a formula for the "best linear unbiased estimator" of $\eta = c'\beta$ where $\beta' = (\beta_0, \beta_1)$ and $c' = (c_1, c_2)$ but realize that it is highly complicated to obtain the solution. Therefore instead of solving the formula, under very restricted conditions that ε_i and Δ_i are independently and identically distributed Gaussian random variables and that $E(\Delta_i) \geq 3\sigma(\Delta_i)$, an estimator $\hat{\eta}$ of η are given by

$$\hat{\eta} = \left[c'(F'F)^{-1}F'\,Y,\ \sum_{i=1}^{n} |\,\lambda_i\,|\ \Delta_i \right]_L \tag{4}$$

where $F = (\mathbf{1}, \mathbf{x})$ is the design matrix with $\mathbf{1}' = (1, 1, \ldots, 1)$ and $\mathbf{x}' = (x_1, x_2, \ldots, x_n)$, and $\mathbf{Y}' = (y_1, y_2, \ldots, y_n)$. For estimators of β_0, β_1, and $\beta_0 + \beta_1 x$, we have $c' = (1, 0)$, $c' = (0, 1)$, and $c' = (1, x)$ respectively. Here, every crisp value $\eta = c'\beta$ is estimated by a fuzzy set. For $c' = (1, x)$, width of a fuzzy set is estimated by a linear combination of the absolute value of linear functions of x, contradicts to the assumption of identically distributedness for Δ_i.

On the other hand Diamond(1992) also suggests a linear regression in which both fuzziness and randomness are contained,

$$Y_i = B_0 + B_1 x_i + E_i, \quad i = 1, 2, \ldots, n \tag{5}$$

where $B_0=[B_0^- \cdot B_0^+]$ and $B_1=[B_1^- \cdot B_1^+]$ are unknown symmetric triangular fuzzy sets and $E_i=[E_i^- \cdot E_i^+]$ is a symmetric triangular random fuzzy set, $E_i^- < E_i^+$ are order statistics from uniform distribution from $(-a, a)$, $a > 0$, and x_i are given crisp numbers. Here $[e^- \cdot e^+]$ represents a symmetric triangular fuzzy set with left and right end points e^- and e^+, respectively. Maximum likelihood estimators of B_0^-, B_0^+, B_1^-, and B_1^+ are then given by

$$
\begin{aligned}
\widehat{B}_1^{\,-} &\in \left[\max_{I_>} w_{ij}^-(a),\ \max_{I_<} w_{ij}^-(a) \right] \\
\widehat{B}_1^{\,+} &\in \left[\max_{I_>} w_{ij}^+(a),\ \max_{I_<} w_{ij}^+(a) \right] \\
\widehat{B}_0^{\,-} &\in \left[\max_{1 \le i \le n} (Y_i^- - B_1^- x_i) - a,\ \max_{1 \le i \le n} (Y_i^- - B_1^- x_i) + a \right], \\
\widehat{B}_0^{\,+} &\in \left[\max_{1 \le i \le n} (Y_i^+ - B_1^+ x_i) - a,\ \max_{1 \le i \le n} (Y_i^+ - B_1^+ x_i) + a \right]
\end{aligned}
\tag{6}
$$

where $I_> = \{(i,j): i,j=1,\dots,n,\ i \neq j, x_i > x_j\}$, $I_< = \{(i,j): i,j=1,\dots,n,\ i \neq j, x_i < x_j\}$, $w_{ij}^-(a)=(y_i^- - y_j^- - 2a)/(x_i - x_j)$, and $w_{ij}^+(a)=(y_i^+ - y_j^+ - 2a)/(x_i - x_j)$. Note that estimator are not unique since the uniform distribution assumed for random variables E_i^- and E_i^+. Moreover the value a is assumed to be known.

In the following section, a linear regression which contains both fuzziness and randomness will be introduced and estimators will be suggested. Examples will be given.

2. A Random Fuzzy Linear Regression Model

In this article we are interested in triangular fuzzy numbers whose membership function is given by

$$
\mu_A(x) = \begin{cases} 1-(m-x)/u & \text{for } m-u < x \le m \\ 1-(x-m)/v & \text{for } m < x \le m+v \\ 0 & \textit{other wise} \end{cases} \tag{7}
$$

where $-\infty < m < \infty$, $u > 0$ and $v > 0$. We denote a triangular fuzzy number as $A=[u, m, v]_T$, where m, u, v are called the center, the left width, and the right width, respectively. When $u=v$, we denote it by $A=[m, u]_{ST}$.

Let (Ω, T, P) be a probability space and $K(R)$ the set of all fuzzy sets A on R with upper semi-continuous normalized membership function and compact support $\{x \in R : \mu_A(x) > 0\}$.

Definition(Näther et al.(1990)) A mapping $\tilde{Y} \mid \Omega \to K(R)$ is a random fuzzy set if every α-cut of $\tilde{Y}$, $\tilde{Y}_\alpha(w) = \{x \in R : \mu_{\tilde{Y}(w)} \geq \alpha\}$ is a compact random set.

For general discussions on fuzzy random sets, see Kwakernaak (1978), Puri and Ralescu (1986), and Zhang et al. (1993). Let U, Y, and V be independent random variables on (Ω, T, P). Assume U and V are positive. Let be the set of all triangular fuzzy set T on R with membership function (7). Then clearly $T(R) \subset K(R)$. Let $g : R^3 \to T(R)$ be a function defined by $g(u, y, v) = [u, y, v]_T$. Then $g(U, Y, V) = [U, Y, V]_T$ is clearly a triangular random fuzzy set. For triangular random fuzzy set $[U, Y, V]_T$, the following theorem is satisfied.

Theorem 1. (Näther et al.(1990)) For triangular random fuzzy set $[U, Y, V]_T$, we have

$$E[U, Y, V]_T = [E(U), E(Y), E(V)]_T \tag{8}$$

Consider a triangular random fuzzy set observations

$$[U_i, Y_i, V_i]_T, \quad i = 1, 2, \ldots, n \tag{9}$$

where

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{10}$$

$$U_i = \alpha_0 + \alpha_1 (x_i - \bar{x})^2 + \tau_i \tag{11}$$

$$V_i = \gamma_0 + \gamma_1 (x_i - \bar{x})^2 + \zeta_i \tag{12}$$

x_i are crisp constant numbers and β_0, β_1, $\alpha_0 \geq 0$, $\alpha_1 \geq 0$, $\gamma_0 \geq 0$, $\gamma_1 \geq 0$ are unknown crisp regression parameters. Random variables ε_i are independent and identically normally distributed with mean 0 and variance σ^2. Positive random variables τ_i are also assumed to be identically distributed and independent. Also ζ_i are assumed to be positive and identically distributed and independent. We also assume all ε_i, , τ_i, and ζ_i are independent.

In (9) the center of the random fuzzy set is assumed to follow a classical simple regression, the left and right widths are assumed to be positive, symmetric about $\bar{x}$, convex quadratic regressions. Convexity for U and V, roughly speaking, are assumed to ensure smaller fuzziness at the center of x_i's.

For the linear regression (10), we apply the maximum likelihood estimators of β_0, and β_1

$$\hat{\beta}= \sum_{i=1}^{n}(Y_i-\bar{Y})(x_i-\bar{x})/\sum_{i=1}^{n}(x_i-\bar{x})^2 \quad \text{and} \quad \hat{\beta}_0=\bar{Y}-\hat{\beta}_1\bar{x} \tag{13}$$

where $\bar{x}=\sum_{i=1}^{n}x_i/n$ and $\bar{Y}=\sum_{i=1}^{n}Y_i/n$. And thus the regression line is estimated by

$$\hat{Y}=\hat{\beta}_0+\hat{\beta}_1 x \tag{14}$$

See Kutner et al. (1996) for estimation in the simple linear regression.

From now on, without loss of generality, we assume that $\bar{x}=0$. Estimators $\hat{\alpha}_0$ and $\hat{\alpha}_1$ of α_0 and α_1 in (11) are obtained by solving the nonlinear programming problem of choosing and α_0 to α_1 satisfy

$$\begin{aligned} &\text{Minimize} \quad \sum_{i=1}^{n}(U_i-\alpha_0-\alpha_1 x_i^2)^2 \\ &\text{subject to} \quad U_i \geq \alpha_0+\alpha_1 x_i^2 \text{ for all } i \text{ and } \alpha_0 \geq 0 \text{ and } \alpha_1 \geq 0 \end{aligned} \tag{15}$$

Thus the regression line is estimated by

$$\hat{\alpha}_0+\hat{\alpha}_1 x_i^2, \quad i=1,2,\dots,n \tag{16}$$

For (12), estimators for γ_0 and γ_1 are obtained by the same way as estimators of α_0 and α_1, *i.e.*, solve the nonlinear programming problem of choosing γ_0 and γ_1 to satisfy

$$\begin{aligned} &\text{Minimize} \quad \sum_{i=1}^{n}(V_i-\gamma_0-\gamma_1 x_i^2)^2 \\ &\text{subject to} \quad V_i \geq \gamma_0+\gamma_1 x_i^2 \text{ for all } i \text{ and } \gamma_0 \geq 0 \text{ and } \gamma_1 \geq 0 \end{aligned} \tag{17}$$

The estimated regression line is then

$$\widehat{\gamma}_0 + \widehat{\gamma}_1 x^2 \tag{18}$$

The expectation $E(\tau_i)$ of τ_i is estimated by a method of moments type estimator, i.e.,

$$\bar{t}_U = \sum_{i=1}^{n} (U_i - \widehat{\alpha}_0 - \widehat{\alpha}_1 x_i^2)/n \tag{19}$$

Note that $\bar{t}_U$ since $U_i - \widehat{\alpha}_0 - \widehat{\alpha}_1 x_i^2 \geq 0$ for $i = 1, 2, \ldots, n$. And the variance $Var(\tau_i)$ of τ_i is estimated by

$$\hat{s}_U^{\,2} = \sum_{i=1}^{n} (U_i - \widehat{\alpha}_0 - \widehat{\alpha}_1 x_i^2 - \bar{t}_U)^2/n \tag{20}$$

Similarly, the expectation $E(\zeta_i)$ and the variance $Var(\zeta_i)$ of ζ_i are estimated by

$$\bar{t}_V = \sum_{i=1}^{n} (V_i - \widehat{\gamma}_0 - \widehat{\gamma}_1 x_i^2)/n \tag{21}$$

$$\hat{s}_V^{\,2} = \sum_{i=1}^{n} (V_i - \widehat{\gamma}_0 - \widehat{\gamma}_1 x_i^2 - \bar{t}_V)^2/n \tag{22}$$

We then estimate $E(U_i)$ by

$$\widehat{U}_i = \hat{\alpha}_0 + \hat{\alpha}_1 (x_i - \bar{x})^2 + \bar{\tau}_U \tag{23}$$

And we estimate $E(V_i)$ by

$$\widehat{V}_i = \hat{\gamma}_0 + \hat{\gamma}_1 (x_i - \bar{x})^2 + \bar{\tau}_V \tag{24}$$

Since $E[U, Y, V]_T = [E(U), E(Y), E(V)]_T$ by Theorem 1, we suggest an estimator for the fuzzy set $E[U, Y, V]_T$ as

$$[\ \widehat{U}_i,\ \widehat{Y}_i,\ \widehat{V}_i]_T \tag{25}$$

Instead of minimizing the variance of random fuzzy sets, we first estimate expectations of three independent random variables Y_t, U_t, and V_t separately using appropriate regression lines and methods of moments and then we put them into the fuzzy set to estimate it. Therefore we remains mainly in the field of statistics to estimate random fuzzy set.

We consider an artificial data to demonstrate our estimators.

Example 1. Suppose data are given $(x_1, \tilde{Y}_1)=(1,[4,2]_{ST})$, $(x_2, \tilde{Y}_2)=(2,[4,2]_{ST})$, $(x_3, \tilde{Y}3)=(3,[5,1]_{ST})$, $(x_4, \tilde{Y}_4)=(4,[7,2]_{ST})$, $(x_5, \tilde{Y}_5)=(5,[8,2]_{ST})$. First we assume model (9) and obtain estimates. The center line $E(Y)=\beta_0+\beta_1 x$ is estimated with data (1,4), (2,4), (3,5), (4,7), (5,8) by

$$\hat{y}=2.3+1.1x$$

from (14). We solve nonlinear programming problem (15) with data (1,2), (2,2), (3,1), (4,2), (5,2) to get the estimate $\hat{\alpha}_0$ and of α_0 and α_1. We obtain that $\hat{\alpha}_0=1$, $\hat{\alpha}_1=0.25$, and $\hat{\alpha}_0+\hat{\alpha}_1(x-3)^2=1+0.25(x-3)^2$. The expectation $E(\tau)$ of τ are then estimated $\bar{t}_U=0.5/5=0.1$ and thus the estimate for $E(U)$ is given by

$$\hat{u}=0.1+0.25(x-3)^2=1.1+0.25(x-3)^2.$$

Therefore the expectation $E(Y+U)=E(Y)+E(U)$ of upper end points $Y+U$ of the random fuzzy set is estimated by

$$\hat{y}+\hat{u}=3.4+1.1x+0.25(x-3)^2.$$

Since $Y=V$, the expectation $E(Y-V)=E(Y)-E(V)$ of lower end points $Y-V$ is estimated by

$$\hat{y}-\hat{v}=1.2+1.1x-0.25(x-3)^2$$

On the other hand if we assume model (3) of Näther et al. (1990), then from (4), estimate of real value $E(Y)=\beta_0+\beta_1 x$ are given by a fuzzy set

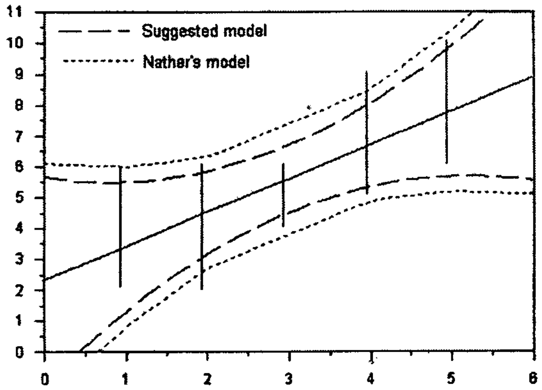
$$[\hat{y}, \hat{\delta}(x)]_{ST}$$

where

$$\hat{\delta}(x)=0.2\times|x-1|+0.4\times|x-2|+0.2+0.4\times|x-4|+0.2\times|x-5|.$$

Even though Näther et al. (1990) assume that variance of Δ does not dependent on x, estimated upper and lower end points does depend on x.

Figure 1 Estimated Regression Lines



References

[1] Bardossy, A. (1990) Note on fuzzy regression, *Fuzzy Sets and Systems*, 37, 65-75.

[2] Diamond, P. (1992) Least Squares and maximum likelihood regression for fuzzy linear models, *Fuzzy Regression Analysis*, Omnitech press, Warsaw, Poland.

[3] Kutner, M. H., Nachtschiem, C. J. and W. Wasserman (1996) *Applied Linear Statistical Models*, 4th ed., Richard & Irwin.

[4] Kwakernaak, H. (1978) Fuzzy random variable: Definition and theorems, *Information Sciences*, 15, 1-29.

[5] Näther, W. and Albrecht, M. (1990) Linear regression with random fuzzy observations, *Statistics*, 21, No. 4, 51-531.

[6] Puri, M. L. and Ralescu, D. A. (1986) Fuzzy random variables, *Journal of Mathematical Analysis and Applications*, 114, 409-422.

[7] Redden, D. and Woodall, W. H. (1994) Properties of certain fuzzy linear regression methods, *Fuzzy Sets and Systems*, 64, 361-375.

[8] Savic, D. A. and Pedrycz, W. (1991) Evaluation of fuzzy linear regression models, *Fuzzy Sets and Systems*, 23, 51-63.

[9] Tanaka, H. (1987) Fuzzy data analysis by possibilistic linear models, *Fuzzy Sets and Systems*, 24, 363-375.

[10] Tanaka, H., Uejima, S. and Asai, K. (1982) Linear regression analysis with fuzzy model, *IEEE Transactions on System Management Cybern.* SMC-12, 903-907.

[11] Tanaka, H and Watada, J. (1988) Possibilistic linear system and their application to linear

regression model, *Fuzzy Sets and Systems*, 27, 1-15.

[12] Zhang, D., Ouyang, H., Lee, E. S., and Yager, R. R. (1993) On fuzzy random sets and their mathematical expectations, *Information Sciences*, 72, 123-142.