Efficiency Measurement for Non-homogeneous Units in a Fuzzy Environment

W.-C. Yeh, C.-K. Hu, F.-B. Liu, and C.-F. Hu

***Abstract*—Data envelopment analysis (DEA) assumes homogeneity among the decision-making units (DMUs) for measuring the relative efficiency of DMUs. In actual practice the efficiency scores generated by DEA may reflect the underlying differences in environments rather than true operational efficiency when DMUs do not operate under similar conditions. This work considers measuring the relative efficiency of non-homogeneous DMUs in a fuzzy environment. A two-stage analysis incorporating imprecise environmental factors into DEA model for measuring the relative efficiency of non-homogeneous units is proposed. A fuzzy regression analysis is employed for adjusting the data of DMUs to account for variations in the fuzzy causes of non-homogeneity. DEA with fuzzy residual errors is then performed to construct a scalar measure of efficiency for all DMUs. An example of the performance assessment of municipalities' solid waste recycling and disposal channels is included. This is the first attempt to estimate the relative efficiency for non-homogeneous DMUs in a fuzzy environment.**

***Index Terms*—Data envelopment analysis, fuzzy optimization, non-homogeneous, fuzzy regression.**

I. Introduction

Measuring the efficiency of decision making units (DMUs) has been a subject of tremendous interest to management practitioners and theorists. Data envelopment analysis (DEA) provides a methodology for measuring the relative efficiency of a group of homogeneous DMUs in the sense that each uses the same inputs and outputs measures and operates in common environments. However, homogeneity conditions might not exist in many real world cases. DMUs could operate under different environmental conditions or lack some inputs or outputs. As an example, consider the case where the DMUs are plants in the same industry that may not all produce the same products [1]. Another is the case of evaluating the relative performance levels of police forces. Crime rates in a given precinct may be influenced by some external environmental factors, e.g., the level of residents' education and percentage of unemployed residents of nearby communities. These socioeconomic characteristics of nearby communities are beyond the control of police departments and might affect police efficiency [2]. When the DMUs are not operated under similar conditions and homogeneous, the efficiency scores generated by DEA may not reflect true managerial and operational efficiency. Therefore, extending DEA to relax the assumption of homogeneity has been important both in theory and in practice.

Researchers have studied this topic from different perspectives depending on the reason for heterogeneity. Potentially, the nonhomogeneous DMU issue could be handled by breaking the set of DMUs into multiple groups, with all members of any group producing the same outputs, and then doing a separate DEA analysis for each group. However, one needs large numbers of DMUs to do this [3]. Another strategy is to adjust for non-homogeneity. A common adjusting technique in the literature is to use the regression equation. Reference [4] calculated an adjusted efficiency score using a regression model with the DEA efficiency scores as the dependent variable to identify managerial inefficiency not caused by extrinsic factors. Reference [5] performed a regression analysis on the DEA using the residuals of the regression analysis as the adjusted scores. Reference [6] proposed a two-stage method to account for non-homogeneity in DMU characteristics. A stepwise multiple regression was implemented on the set of efficiency scores using a set of DMU characteristics that are expected to account for differences in efficiency not attributable to management. DEA based on the adjusted DMU outputs was then performed to produce efficiency scores for non-homogeneous DMUs. Reference [3] proposed two additional methods of compensating for non-homogeneity. Reference [2] applied a three-stage analysis to explicitly incorporate environment factor into the DEA model for measuring the performance of police forces in Taiwan. Reference [7] developed a DEA type of methodology to evaluate non-homogeneous DMUs with different input configurations.

Due to the fuzziness involved in the real-world decision making problems, some data available for efficiency analysis of DMUs might be imprecise or vagueness, e.g., the causes of non-homogeneity might be in the form of qualitative or linguistic. A variety of methodologies and applications in fuzzy DEA has been studied for efficiency measurement in a fuzzy environment. Reference [8] developed a procedure to measure the efficiencies of DMUs with fuzzy observations. The fuzzy measurement concept and the extension principle are adopted to transform the fuzzy DEA model into a traditional DEA model. Reference [9] considered the fuzzy CCR model with asymmetric triangular fuzzy numbers, and [10] proposed the fuzzy BCC model using probabilities to

Manuscript received March 28, 2020; revised July 16, 2020. This work was supported by National Science Council, Taiwan under Grant MOST 108-2221-E-415-006.

W.-C. Yeh is with the Department of IEEM, National Tsing Hua University, Hsinchu, Taiwan (e-mail: yeh@ieee.org).

C. K. Hu is with the Department of International Business, Kao Yuan University, Kaohsiung, Taiwan (e-mail: t80147@cc.kyu.edu.tw).

F.-B. Liu is with the Department of Mechanical and Automation Engineering, I-Shou University, Kaohsiung, Taiwan (e-mail: fliu@isu.edu.tw).

C-. F. Hu is with the Department of Applied Mathematics, National Chiayi University, Chiayi, Taiwan (Corresponding author; e-mail: cfhu@mail.ncyu.edu.tw).

doi: 10.18178/ijtef.2020.11.4.668

conduct the analysis. They employed α -cut to transform the fuzzy DEA model into a linear structure model. For a taxonomy and review of the fuzzy DEA methods, readers can refer to the work by [11]. To the best of our knowledge, no literature to date has studied the efficiency measures for non-homogeneous units in a fuzzy environment.

In this work a two-stage analysis which incorporates the imprecise environmental factors into the fuzzy DEA model for measuring the relative efficiencies of non- homogeneous DMUs with fuzzy data is developed. In the first stage inputs and outputs of non-homogeneous units DMUs are adjusted by employing the fuzzy regression analysis to account for variations in the imprecise causes of non-homogeneity. DEA with fuzzy residual errors is developed to measure the relative efficiency for all DMUs in the second stage of the proposed method. An example of the performance assessment of municipalities' solid waste recycling and disposal channels is provided. The rest of the paper is organized as follows. In Section II, the fuzzy regression analysis is introduced to compensate the non-homogeneity of DMUs. DEA with fuzzy residual errors is then developed to measure the relative efficiency for all DMUs in Section III. An example of evaluating the relative efficiency of municipal reverse logistics channels is included in Section IV. The paper is concluded in Section V.

II. A Methodology for Compensating the Non-homogeneity of DMUs

To compensate the non-homogeneity of DMUs, the simplest adjustment is to estimate inputs and outputs using regression analysis. This work considers adjusting the data of the non-homogeneous DMUs by fuzzy regression analysis with the fuzzy causes of non-homogeneity as the independent variables. Suppose there are *m* inputs, *s* outputs and n DMUs being evaluated. Denote $\tilde{x}_{ij}$ as the i-th fuzzy input and $\tilde{y}_{rj}$ as the *r*-th fuzzy output of the *j*-th DMU, i = 1,2,…,m , r =1,2...,s , j=1,2,...,n. To detect how external factors affect the variation in the inputs and outputs of the non-homogeneous DMUs, we consider the fuzzy regression equations:

$$\hat{y}_{rj} = \alpha_{rj} + \beta_{rj,1}\tilde{t}_{rj,1} + \beta_{rj,2}\tilde{t}_{rj,2} + \cdots + \beta_{rj,c}\tilde{t}_{rj,c} + \tilde{e}'_{rj,1} + \tilde{e}'_{rj,2}\ ,$$

and

$$\hat{x}_{ij} = \alpha_{ij} + \beta_{ij,1}\tilde{t}_{ij,1} + \beta_{ij,2}\tilde{t}_{ij,2} + \cdots + \beta_{ij,c}\tilde{t}_{ij,c} + \tilde{e}_{ij,1} + \tilde{e}_{ij,2}\ ,$$

where $\hat{y}_{rj}$ and $\hat{x}_{ij}$ are the fuzzy estimated values of the actual fuzzy outputs $\tilde{y}_{rj}$ and inputs $\tilde{x}_{ij}$, respectively , $\tilde{t}_1$ ... , $\tilde{t}_c$ are the imprecise external factors, $\tilde{e}_{ij,1}$ and $\tilde{e}'_{rj,1}$ represent the sum of randomness and measurement error, and $\tilde{e}_{ij,2}$ and $\tilde{e}'_{rj,2}$ represent differences due to policy, practice, and operating conditions.

A number of novel approaches to fuzzy regression were proposed after the inception of fuzzy set theory [12]. Reference [13] developed the fuzzy least squares regression which considers minimizing the overall squared error between the observed values and the estimated values. Reference [14] proposed the possibilistic regression approach where the fuzzy linear regression with fuzzy coefficients and crisp input variables was introduced. Both of the approaches to fuzzy regression analysis have been investigated extensively [15]-[18]; however, they are sensitive to outliers [19]. Reference [20] suggested the least absolute deviation estimators to construct the fuzzy regression model and showed that the proposed approach is more robust than the least squares deviation method when the data contains fuzzy outliers. With the development of computing technology, the least absolute deviation method was later studied and applied by some researchers [19], [21]-[23].

Taking advantage of the well-developed techniques in fuzzy regression analysis, the method of least absolute deviation is employed to construct the fuzzy regression models for adjusting the inputs and outputs of DMUs in this work. We recall some basic concepts and important properties associated with fuzzy set theory and the least absolute deviation regression analysis.

Definition 1. Suppose that the membership function of the fuzzy number $\tilde{a} = (a_1, a_2, a_3)$ is defined as

$$\mu_{\tilde{a}}(x) = \begin{cases} \dfrac{x - a_1}{a_2}\ , & a_1 \le x \le a_1 + a_2 \\ \dfrac{a_1 + a_2 + a_3 - x}{a_3}\ , & a_1 + a_2 < x \le a_1 + a_2 + a_3 \\ 0\ , & \text{otherwise.} \end{cases}$$

Then we call $\tilde{a} = (a_1, a_2, a_3)$ a triangular fuzzy number with $a_2, a_3 \ge 0$, where a_1, a_2, a_3 are the left point, the width of the left side, and the width of the right side of the triangular fuzzy number $\tilde{a}$, respectively.

Triangular fuzzy numbers are of more importance among the various types of fuzzy numbers. In this work we consider the imprecise data of DMUs as triangular fuzzy numbers. According to Zadeh's extension principle [12], for two triangular fuzzy numbers $\tilde{a} = (a_1, a_2, a_3)$ and $\tilde{b} = (b_1, b_2, b_3)$, we have the following results:

$$\tilde{a} + \tilde{b} = (a_1 + b_1, a_2 + b_2, a_3 + b_3)$$
$$\tilde{a} - \tilde{b} = (a_1 - b_1, a_2 + b_2, a_3 + b_3)$$
$$\tilde{a} \times \tilde{b} = (a_1 \times b_1, |b_1 \times a_2| + |a_1 \times b_2|, |b_1 \times a_3| + |a_1 \times b_3|), \text{ where } a_1 \ne 0\ ,\ b_1 \ne 0.$$

Definition 2. [24] For two triangular fuzzy numbers $\tilde{a} = (a_1, a_2, a_3)$, and $\tilde{b} = (b_1, b_2, b_3)$. The absolute distance between $\tilde{a}$ and $\tilde{b}$ is defined as

$$d(\tilde{a}, \tilde{b}) = |a_1 - b_1| + |a_2 - b_2| + |a_3 - b_3|. \qquad (1)$$

It is shown that d is a distance metric on triangular fuzzy numbers (Li *et al.* 2016a). Suppose that $\tilde{y}$ and $\tilde{t}_1, \cdots, \tilde{t}_c$ are triangular fuzzy numbers and $\alpha, \beta_1, \cdots, \beta_c$ are crisp coefficients. The fuzzy linear regression model with crisp coefficients and fuzzy independent variables can be described as follows:

$$\tilde{y} = \alpha + \beta_1\tilde{t}_1 + \cdots + \beta_c\tilde{t}_c. \qquad (2)$$

For a group of w fuzzy sample data, say $(\tilde{t}_{i1}, \tilde{t}_{i2}, \ldots, \tilde{t}_{ic}, \bar{y}_i)$, $i = 1, 2, \cdots, w$, let the triangular fuzzy number $\tilde{t}_{ik} = (t_{ik,1}, t_{ik,2}, t_{ik,3}), i = 1, 2, \cdots, w, k = 1, 2, \cdots, c$, be the independent variable and $\bar{y}_i = (y_{i,1}, y_{i,2}, y_{i,3})$ be the dependent variable in the fuzzy linear regression model (2). By the extension principle, we have

$$\tilde{y}_i = \alpha + \beta_1 \tilde{t}_{i1} + \beta_2 \tilde{t}_{i2} + \cdots + \beta_c \tilde{t}_{ic} = (\alpha + \sum_{k=1}^{c} \beta_k t_{ik,1}, \sum_{k=1}^{c} |\beta_k t_{ik,2}|, \sum_{k=1}^{c} |\beta_k t_{ik,3}|).$$

To determine the estimated regression coefficients, $\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_c$, the least absolute deviation criterion is employed by minimizing the overall absolute error distance. It can be described as the following optimization problem:

$$\min_{\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_c} \sum_{i=1}^{w} d(\bar{y}_i, \hat{\alpha} + \hat{\beta}_1 \tilde{t}_{i1} + \hat{\beta}_2 \tilde{t}_{i2} + \cdots + \hat{\beta}_c \tilde{t}_{ic}). \quad (3)$$

According to (1), the optimization problem (3) can be rewritten as

$$\min_{\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_c} \sum_{i=1}^{w} |y_{i,1} - \hat{\alpha} - \sum_{k=1}^{c} \hat{\beta}_k t_{ik,1}| + |y_{i,2} - \sum_{k=1}^{c} |\hat{\beta}_k t_{ik,2}|| + |y_{i,3} - \sum_{k=1}^{c} |\hat{\beta}_k t_{ik,3}||, \quad (4)$$

which is equivalent to

$$\begin{aligned} &\min \sum_{i=1}^{w} (\mu_{i1} + \mu_{i2} + v_{i1} + v_{i2} + \varphi_{i1} + \varphi_{i2}) \\ &\text{s.t.} \quad y_{i.1} - \hat{\alpha} - \sum_{k=1}^{c} \hat{\beta}_k t_{ik,1} = \mu_{i1} - \mu_{i2} \\ &\qquad y_{i.2} - \sum_{k=1}^{c} |\hat{\beta}_k t_{ik,2}| = v_{i1} - v_{i2} \\ &\qquad y_{i.3} - \sum_{k=1}^{c} |\hat{\beta}_k t_{ik,3}| = \varphi_{i1} - \varphi_{i2} \\ &\qquad \mu_{i1}, \mu_{i2}, v_{i1}, v_{i2}, \varphi_{i1}, \varphi_{i2} \geq 0. \end{aligned} \quad (5)$$

The estimated regression coefficients, $\hat{\alpha}$, $\hat{\beta}_1, \ldots, \hat{\beta}_c$, are then determined by solving the problem (5). Consequently, the fuzzy estimated value of $\tilde{y}$ can be obtained as follows:

$$\hat{y} = \hat{\alpha} + \hat{\beta}_1 \tilde{t}_1 + \hat{\beta}_2 \tilde{t}_2 + \cdots + \hat{\beta}_c \tilde{t}_c. \quad (6)$$

III. DEA with Fuzzy Residual Errors

After adjusting the imprecise inputs $\tilde{x}_{ij}$ and outputs $\tilde{y}_{rj}$ of DMUs to account for variations in the fuzzy causes of non-homogeneity, we then perform DEA using the differences between actual and estimated fuzzy inputs and outputs. That is, instead of using the actual inputs $\tilde{x}_{ij}$ and outputs $\tilde{y}_{rj}$, we use the residual errors defined as

$$\tilde{x}'_{ij} \triangleq \tilde{x}_{ij} - \hat{x}_{ij}, i = 1,2, \ldots, m, j = 1,2, \ldots, n, \quad (7)$$

and

$$\tilde{y}'_{rj} \triangleq \tilde{y}_{rj} - \hat{y}_{rj}, r = 1,2, \ldots, s, j = 1,2, \ldots, n. \quad (8)$$

The DEA model with inputs $\tilde{x}'_{ij}$ and outputs $\tilde{y}'_{rj}$ can be described as

$$\begin{aligned} &\max_{u,v} \sum_{r=1}^{s} u_r \tilde{y}'_{rp} / \sum_{i=1}^{m} v_i \tilde{x}'_{ip} \\ &\text{s.t.} \quad \sum_{r=1}^{s} u_r \tilde{y}'_{rj} - \sum_{i=1}^{m} v_i \tilde{x}'_{ij} \leq 0, j = 1,2, \ldots, n \\ &\qquad u_r, v_i \geq 0, \forall r = 1,2, \ldots, s, i = 1,2, \ldots, m, \end{aligned} \quad (9)$$

where $u = (u_1, u_2, \ldots, u_s)^T$ and $v = (v_1, v_2, \ldots, v_m)^T$ are the weight vectors to be applied to the inputs and outputs of DMUs, respectively.

Let $\tilde{x}'_{ij} = (x^l_{ij}, x^m_{ij} - x^l_{ij}, x^u_{ij} - x^m_{ij})$ and $\tilde{y}'_{rj} = (y^l_{rj}, y^m_{rj} - y^l_{rj}, y^u_{rj} - y^m_{rj})$ be triangular fuzzy numbers with corresponding membership functions defined as

$$\mu_{\tilde{x}'_{ij}(x)} = \begin{cases} \frac{x - x^l_{ij}}{x^m_{ij} - x^l_{ij}}, & x^l_{ij} \leq x \leq x^m_{ij} \\ \frac{x^u_{ij} - x}{x^u_{ij} - x^m_{ij}}, & x^m_{ij} \leq x \leq x^u_{ij} \\ 0, & \text{otherwise}, \end{cases} \quad i = 1,2, \cdots, m, j = 1,2, \cdots, n,$$

and

$$\mu_{\tilde{y}'_{ij}(x)} = \begin{cases} \frac{y - y^l_{rj}}{y^m_{rj} - y^l_{rj}}, & y^l_{rj} \leq y \leq y^m_{rj} \\ \frac{y^u_{rj} - y}{y^u_{rj} - y^m_{rj}}, & y^m_{rj} \leq y \leq y^u_{rj} \\ 0, & \text{otherwise}, \end{cases} \quad r = 1,2, \cdots, s, j = 1,2, \cdots, n.$$

For each $\alpha \in [0,1]$, the α-level sets of $\tilde{x}'_{ij}$ and $\tilde{y}'_{rj}$ are defined as

$$(\tilde{x}'_{ij})_\alpha = [\alpha x^m_{ij} + (1-\alpha) x^l_{ij}, \alpha x^m_{ij} + (1-\alpha) x^u_{ij}], \quad i = 1,2, \cdots, m, j = 1,2, \cdots, n,$$

and

$$(\tilde{y}'_{rj})_\alpha = [\alpha y^m_{rj} + (1-\alpha) y^l_{rj}, \alpha y^m_{rj} + (1-\alpha) y^u_{rj}], \quad r = 1,2, \cdots, s, j = 1,2, \cdots, n,$$

respectively. Let x'_{ij} and y'_{rj} be variables in the α-level sets of $\tilde{x}'_{ij}$ and $\tilde{y}'_{rj}$, respectively. By introducing α-cuts of objective function and constraints, the problem (9) can be restated as the crisp optimization problem [9]:

$$\begin{aligned} &\max_{u,v} \sum_{r=1}^{s} u_r \tilde{y}'_{rp} / \sum_{i=1}^{m} v_i \tilde{x}'_{ip} \\ &\text{s.t.} \quad \sum_{r=1}^{s} u_r \tilde{y}'_{rj} - \sum_{i=1}^{m} v_i \tilde{x}'_{ij} \leq 0, j = 1,2, \ldots, n, \\ &\qquad \alpha x^m_{ij} + (1-\alpha) x^l_{ij} \leq x'_{ij} \leq \alpha x^m_{ij} + (1-\alpha) x^u_{ij}, \\ &\qquad i = 1,2, \cdots, m, j = 1,2, \cdots, n, \\ &\qquad \alpha y^m_{rj} + (1-\alpha) y^l_{rj} \leq y'_{rj} \leq \alpha y^m_{rj} + (1-\alpha) y^u_{rj}, \\ &\qquad r = 1,2, \cdots, s, j = 1,2, \cdots, n, \\ &\qquad u_r, v_i \geq 0, \forall r = 1,2, \cdots, s, i = 1,2, \cdots, m, \end{aligned} \quad (10)$$

or equivalently

$$\begin{aligned} &\max_{u,v} \sum_{r=1}^{s} u_r y'_{rp} \\ &\text{s.t.} \quad \sum_{i=1}^{m} v_i x'_{ip} = 1, \\ &\qquad \sum_{r=1}^{s} u_r y'_{rj} - \sum_{i=1}^{m} v_i x'_{ij} \leq 0, j = 1,2, \ldots, n, \\ &\qquad \alpha x^m_{ij} + (1-\alpha) x^l_{ij} \leq x'_{ij} \leq \alpha x^m_{ij} + (1-\alpha) x^u_{ij}, \\ &\qquad i = 1,2, \cdots, m, j = 1,2, \cdots, n, \\ &\qquad \alpha y^m_{rj} + (1-\alpha) y^l_{rj} \leq y'_{rj} \leq \alpha y^m_{rj} + (1-\alpha) y^u_{rj}, \\ &\qquad r = 1,2, \cdots, s, j = 1,2, \cdots, n, \\ &\qquad u_r, v_i \geq 0, \forall r = 1,2, \cdots, s, i = 1,2, \cdots, m. \end{aligned} \quad (11)$$

Let $\bar{x}_{ij} = v_i x'_{ij}, i = 1,2, \cdots, m, j = 1,2, \cdots, n$, and $\bar{y}_{rj} = u_r y'_{rj}, r = 1,2, \cdots, s, j = 1,2, \cdots, n$. By these substitutions, the problem (11) can be converted to the following linear programming problem:

$$\begin{aligned} &\max_{u,v} \sum_{r=1}^{s} \bar{y}_{rp} \\ &\text{s.t.} \quad \sum_{i=1}^{m} \bar{x}_{ip} = 1, \\ &\qquad \sum_{r=1}^{s} \bar{y}_{rj} - \sum_{i=1}^{m} \bar{x}_{ij} \leq 0, j = 1,2, \ldots, n, \\ &\qquad v_i(\alpha x^m_{ij} + (1-\alpha) x^l_{ij}) \leq \bar{x}_{ij} \leq v_i(\alpha x^m_{ij} + (1-\alpha) x^u_{ij}), \ i = 1,2, \cdots, m, j = 1,2, \cdots, n, \\ &\qquad u_r(\alpha y^m_{rj} + (1-\alpha) y^l_{rj}) \leq \bar{y}_{rj} \leq u_r(\alpha y^m_{rj} + (1-\alpha) y^u_{rj}), \ r = 1,2, \cdots, s, j = 1,2, \cdots, n, \\ &\qquad u_r, v_i \geq 0, \forall r = 1,2, \cdots, s, i = 1,2, \cdots, m, \end{aligned} \quad (12)$$

For each $\alpha \in (0,1]$, a solution of the fuzzy DEA problem (9) can be obtained by solving the linear programming problem (12).

According to the basic assumption of DEA, the measurement error and randomness can be eliminated. Therefore, we have $\tilde{e}_{rj,1} = \tilde{e}'_{ij,1} = 0$. Consequently, the residual errors must be due to policy and practice. This leads to the following result.

Theorem 1. Assume that $\tilde{e}_{ij,1} = \tilde{e}'_{rj,1} = 0$. If the fuzzy regressions used to estimate the values of the fuzzy inputs and outputs are all perfect fits, that is, $\tilde{e}_{ij,2} = \tilde{e}'_{rj,2} = 0$, then the DEA model (9) with fuzzy residual errors defined in (7) and (8) places all DMUs on the efficiency frontier.

Proof. Let $\hat{x}_{ij}$ and $\hat{y}_{rj}$ be the estimated values of fuzzy inputs and outputs, respectively, and $\tilde{x}_{ij}$ and $\tilde{y}_{rj}$ be the actual values of the fuzzy inputs and outputs, respectively. Consider the fuzzy residual errors $\tilde{x}'_{ij}$ and $\tilde{y}'_{rj}$ defined in (7) and (8), respectively, i.e., $\tilde{x}'_{ij} = \tilde{x}_{ij} - \hat{x}_{ij}$ and $\tilde{y}'_{rj} = \tilde{y}_{rj} - \hat{y}_{rj}$, $j = 1,\cdots,n$, $i = 1,\cdots,m$, and $r = 1,\cdots,s$. Since all regressions estimating the inputs and outputs are perfect fits, then all $\tilde{x}'_{ij} = \tilde{y}'_{rj} = 0$, and all $\tilde{x}'_{ip} = \tilde{y}'_{rp} = 0$. Then the objective function of the fuzzy DEA model (9)

$$\sum_{r=1}^{s} u_r \tilde{y}'_{rp} / \sum_{i=1}^{m} v_i \tilde{x}'_{ip}$$

is indeterminate. Also,

$$\sum_{r=1}^{s} u_r \tilde{y}'_{rj} - \sum_{i=1}^{m} v_i \tilde{x}'_{ij} \leq 0, j = 1,2,\ldots,n,$$

is feasible. Thus, an efficiency score of 1 for each DMU is feasible, and all DMUs can be placed on the efficiency frontier.

IV. An Example

In this section, we applied the proposed method to assess the relative efficiency of the reverse logistics channels of twenty-three municipalities. The modified data studied in Haas [25] is utilized. All the municipalities acknowledge that refuse and recyclables must be collected and disposed of, and each has provided a means to achieve that. Since each municipality operates under a unique set of conditions, the municipalities fail to meet the homogeneity assumption of DEA. Their areas and populations are different. Additionally, the demographic characteristics of their residents vary and this may affect the propensity of the citizenry to participate in their recycling programs [3]. To compensate the non-homogeneity of the twenty-three municipalities, data are adjusted by the fuzzy regression analysis before performing DEA. The DEA model used in the analysis includes two inputs, size of the solid waste stream ($\tilde{x}_1$) and net channel costs ($\tilde{x}_2$), and two outputs, residents' satisfaction to receive curbside solid waste removal service ($\tilde{y}_1$) and the quantity recycled in the municipality ($\tilde{y}_2$). Any or all of these inputs and outputs could potentially be affected by the differences in operating conditions. Table I provides inputs and outputs data for the study.

Based on the study on [25], two external environmental factors which may cause differences in operating conditions among the municipalities were chosen. These factors include the residents' tendencies to participate in the recycling program ($\tilde{t}_1$) and population ($\tilde{t}_2$). Table II provides data for the external environmental factors.

TABLE I: Input and Output Data

DMU	Input		Output	
j	$\tilde{x}_{1j}$	$\tilde{x}_{2j}$	$\tilde{y}_{1j}$	$\tilde{y}_{2j}$
1 Abington	(40823100100)	3558268	(80,10,5)	17466
2 Allentown	(5185250100)	4415240	(85,5,10)	19188
3 Bensalem	(2663510050)	2971000	(20,5,10)	3009
4 Bristol Borough	(488050100)	590734	(50,5,5)	678
5 Bristol Twp.	(2679510050)	2168218	(50,5,10)	4933
6 Cheltenham	(2237350100)	1556099	(90,5,5)	11068
7 Doylestown	(6805100100)	1269131	(80,5,10)	1591
8 Horsham	(1105850100)	1941753	(70,5,10)	2115
9 Lower Makefield	(2161250100)	2603100	(50,10,10)	8067
10 LowerMoreland	(900110050)	768328	(60,5,10)	3543
11 Lower Southampton	(9315100100)	1027374	(40,10,10)	2905
12 Middletown	(2019750100)	2298994	(30,5,10)	4091
13 Montgomery	(61865050)	2003249	(30,5,10)	1309
14 Newtown	(6418100100)	1391120	(40,5,5)	1283
15 Northampton	(1660710050)	1984016	(60,10,10)	5533
16 Plymouth	(87415050)	1098588	(60,5,10)	2343
17 Springfield	(11034100100)	1120890	(50,10,10)	2966
18 Upper Dublin	(1315410050)	1743444	(55,10,10)	3389
19 Upper Moreland	(13190100100)	1664000	(75,5,10)	6413
20 Upper Southampton	(7540100100)	927000	(50,5,10)	2404
21 Warminster	(15399100100)	1760025	(60,5,5)	4048
22 West Norriton	(778510050)	1492020	(30,10,5)	952
23 Whitpain	(802350100)	1269000	(40,5,5)	1544

TABLE II: Environmental Factors

DMU	Environmental factors
j	$(\tilde{t}_{1j}, \tilde{t}_{2j})$
1 Abington	((30,5,5),56444)
2 Allentown	((50,5,5),105090)
3 Bensalem	((60,5,5),57078)
4 Bristol Borough	((40,10,5),10428)
5 Bristol Twp.	((80,5,5),57703)
6 Cheltenham	((80,5,5),34863)
7 Doylestown	((60,5,10),17274)
8 Horsham	((60,5,5),21887)
9 Lower Makefield	((80,10,10),30301)
10 Lower Moreland	((40,5,5),11852)
11 LowerSouthampton	((40,10,10),19896)
12 Middletown	((60,5,5),44106)
13 Montgomery	((70,5,5),18325)
14 Newtown	((60,5,5),15857)
15 Northampton	((70,10,10),37979)
16 Plymouth	((80,10,5),15771)
17 Springfield	((50,10,10),19612)
18 Upper Dublin	((60,5,5),24682)
19 Upper Moreland	((70,10,5),25254)
20 UpperSouthampton	((30,5,5),16175)
21 Warminster	((50,5,5),33419)
22 West Norriton	((60,10,10),15305)
23 Whitpain	((70,10,10),16516)

In the proposed method the least absolute deviation criterion is employed to determine the estimated regression coefficients, $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2$, for each fuzzy data. For instance, to construct the fuzzy regression model for the first output variable $\tilde{y}_1$, the following problem is considered.

$$\min_{\hat{\alpha},\hat{\beta}_1,\hat{\beta}_2} |80-\hat{\alpha}-30\hat{\beta}_1-56444\hat{\beta}_2| + |10-|5\hat{\beta}_1|| + 5 - |5\hat{\beta}_1|| + |85-\hat{\alpha}-50\hat{\beta}_1-105090\hat{\beta}_2| + |5-|5\hat{\beta}_1|| + |10- |5\hat{\beta}_1|| + |20-\hat{\alpha}-60\hat{\beta}_1-57078\hat{\beta}_2| + |5-|5\hat{\beta}_1|| + |10- |5\hat{\beta}_1|| + \cdots + |30-\hat{\alpha}-60\hat{\beta}_1-15305\hat{\beta}_2| + |10-|10\hat{\beta}_1|| + 5 - |10\hat{\beta}_1|| + |40-\hat{\alpha}-70\hat{\beta}_1-16516\hat{\beta}_2| + |5-|10\hat{\beta}_1|| + |5- |10\hat{\beta}_1|| \quad (13)$$

The estimated regression coefficients, $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2$, in (13) for each fuzzy data are shown in Table III.

TABLE III: COEFFICENTS IN THE FUZZY REGRESSION MODEL

	α	β_1	β_2
Outputs			
Residents' satisfaction to receive curbside solid waste removal service($\tilde{y}_1$)	1.75	0.9	0.0001
The quantity recycled in the municipality($\tilde{y}_2$)	1.1998	0.9733	0.1439
Inputs			
Size of the solid waste stream($\tilde{x}_1$)	7.7720	10	0.4717
Net channel costs($\tilde{x}_2$)	12.8042	69.8208	52.0290

After adjusting the fuzzy data of DMUs to account for variations in the causes of non-homogeneity, DEA model with fuzzy residual errors is performed to construct a scalar measure of efficiency for all DMUs. A solution of the DEA with fuzzy residual errors can be obtained by solving the following linear programming problem:

$$\begin{aligned}
&\max \bar{y}_{1p} + \bar{y}_{2p} \\
&\text{s.t. } \bar{x}_{1p} + \bar{x}_{2p} = 1 \\
&\bar{y}_{1j} + \bar{y}_{2j} - \bar{x}_{1j} + \bar{x}_{2j} \le 0, j = 1,2,\dots,23, \\
&v_1(\alpha x_{1j}^m + (1-\alpha)x_{1j}^l \le \bar{x}_{1j} \le v_1(\alpha x_{1j}^m + (1-\alpha)x_{1j}^u, \\
&j = 1,2,\dots,23, \\
&v_2(\alpha x_{2j}^m + (1-\alpha)x_{2j}^l \le \bar{x}_{2j} \le v_2(\alpha x_{2j}^m + (1-\alpha)x_{2j}^u, \\
&j = 1,2,\dots,23, \\
&u_1(\alpha y_{1j}^m + (1-\alpha)y_{1j}^l \le \bar{y}_{1j} \le u_1(\alpha y_{1j}^m + (1-\alpha)y_{1j}^u, \\
&j = 1,2,\dots,23, \\
&u_2(\alpha y_{2j}^m + (1-\alpha)y_{2j}^l \le \bar{y}_{2j} \le u_2(\alpha y_{2j}^m + (1-\alpha)y_{2j}^u, \\
&j = 1,2,\dots,23, \quad (14)
\end{aligned}$$

Table IV shows the efficiency scores and the associated rankings calculated from the DEA model (14). The columns headed "Unadjusted model" represent the results produced by using the unadjusted inputs and outputs. The columns headed "Residual error model" represent the results obtained by the adjustment techniques proposed in this work. In our experiment, the efficiency scores produced by the unadjusted model are spread across the range from 1.0000 to 0.7814 with a mean of 0.8937 and a standard deviation of 0.0802. The efficiency scores produced by the residual error model are spread across a larger range from 1.0000 to 0.1369 with a mean of 0.7400 and a standard deviation of 0.2561. There are six units on the efficiency frontier in the results of the unadjusted model. When external environmental factors were accounted for, Bristol Borough (No. 4), Cheltenham (No. 6), and Newtown (No. 14) leave the efficiency frontier, but Cheltenham (No. 6) retains a high efficiency score. Bristol Borough (No. 4) and Newtown (No. 14) which are on the efficiency frontier in the results of the unadjusted models reduce to relatively low efficiency scores after compensating for non-homogeneity. Moreover, the performance of Lower Moreland (No. 10) and Whitpain (No. 23) which are among the "mid-range" performers in terms of the unadjusted efficiency score, become the top performers when using the residual efficiency. On the other hand, the efficiency ratings of Moreland (No. 10), Plymouth (No. 16) and Whitpain (No. 23) using the residual efficiency, are improved noticeably as compared to its rating in terms of the unadjusted efficiency scores. While Bristol Borough (No. 4) and Newtown (No. 14) exhibit a noticeable decline in their efficiency ratings after compensating for non-homogeneity. Our results imply that the external environmental factors do have some effects on the efficiency scores and rankings of each municipality. The results will be closer to the real level of efficiency after compensating for non-homogeneity.

TABLE IV: EFFICIENCY SCORES AND THE ASSOCIATED RANKINGS

DMU	Unadjusted model		Residual error model	
j	Score	Rank	Score	Rank
1 Abington	0.8894	11	0.4163	20
2 Allentown	1.0000	1	1.0000	1
3 Bensalem	0.7923	21	0.1369	23
4 Bristol Borough	1.0000	1	0.5238	18
5 Bristol Twp.	0.9409	8	0.9878	8
6 Cheltenham	1.0000	1	0.8834	9
7 Doylestown	0.8144	18	0.6741	15
8 Horsham	0.7814	23	0.4824	19
9 Lower Makefield	0.8422	16	0.3942	21
10 LowerMoreland	0.8782	13	1.0000	1
11 LowerSouthampton	0.9035	10	0.9028	10
12 Middletown	0.8487	15	0.6124	17
13 Montgomery	1.0000	1	1.0000	1
14 Newtown	1.0000	1	0.6981	14
15 Northampton	0.9222	9	0.9938	6
16 Plymouth	0.8018	20	0.7238	13
17 Springfield	0.8863	12	0.8354	11
18 Upper Dublin	0.8196	17	0.7663	12
19 Upper Moreland	1.0000	1	1.0000	1
20 Upper Southampton	0.9588	7	0.9883	7
21 Warminster	0.8764	14	0.6142	16
22 West Norriton	0.7905	22	0.3922	22
23 Whitpain	0.8082	19	1.0000	1

V. CONCLUSION

This work considers measuring the relative efficiency of non-homogeneous DMUs with imprecise inputs, outputs, and external environmental factors. A two-stage analysis which incorporates the imprecise environmental factors into the fuzzy DEA model for measuring the relative efficiencies of non-homogeneous DMUs is proposed. To adjust the inputs and outputs of non-homogeneous DMUs, the least absolute deviation regression technique is employed to account for variations in the fuzzy causes of non-homogeneity. DEA with fuzzy residual errors is then performed to construct a scalar measure of efficiency for all DMUs. It is shown that if the fuzzy regression is a perfect fit, then the DEA with fuzzy residual errors places all DMUs on the efficiency frontier. An example on the assessment of the performance of municipalities' solid waste recycling and disposal channels is included. Our results show that these imprecise environmental factors do influence the efficiency scores and rankings of each municipality. This confirms our earlier statement that the efficiency scores generated by DEA may not reflect true managerial and operational efficiency when DMUs are not operated under similar conditions.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

W.-C. Yeh, C.-K. Hu, F.-B. Liu and C.-F. Hu conducted the research. W.-C. Yeh and C.-F. Hu developed the two-stage methodology analysis. C.-K. Hu and F.-B. Liu implemented the numerical application. All authors had approved the final version.

References

[1] W. D. Cook, J. Harrison, R. Imanirad, P. Rouse, and J. Zhu, "Data envelopment analysis with nonhomogeneous DMUs," *Operations Research*, vol. 61, no. 3, pp. 666-676, 2013.

[2] T.-H. Wu, M.-S. Chen, and J.-Y. Yeh, "Measuring the performance of police forces in Taiwan using data envelopment analysis," *Evaluation and Program Planning*, vol. 33, pp. 246-254, 2010.

[3] D. A. Haas and F. H. Murphy, "Compensating for non-homogeneity in decision making units in data envelopment analysis," *European Journal of Operational Research*, vol. 144, no. 3, pp. 530-544, 2003

[4] S. C. Ray, "Resource-use efficiency in public schools: A study of Connecticut data," *Management Science*, vol. 37, pp. 1620-1628, 1991.

[5] J.L. Fizel and T.S. Nunnikhoven, "Technical efficiency of for-profit and non-profit nursing homes," *Managerial and Decision Economics*, vol. 13, pp. 429-439, 1992.

[6] T. R. Sexton, S. Sleeper, S., and Jr. Taggart, "Improving pupil transportation in North Carolina," *Interfaces*, vol. 24, no. 1, pp. 87-103, 1994.

[7] W. Li, L. Liang, W. D. Cook, and J. Zhu, "DEA models for non-homogeneous DMUs with different input configurations," *European Journal of Operational Research*, vol. 254, pp. 946-956, 2016.

[8] C. Kao and S.-T. Liu, "Data envelopment analysis with missing data: an application to university libraries in Taiwan," *Journal of the Operational Research Society*, vol. 51, pp. 897-905, 2000.

[9] S. M. Saati, A. Memariani, and G. R. Jahanshahloo, "Efficiency analysis and ranking of DMUs with fuzzy data," *Fuzzy Optimization and Decision Making*, vol. 3, pp. 255-267, 2002.

[10] S. Lertworasirikul, S.-C. Fang, H. L. W. Nuttle, and J. A. Joines, "Fuzzy BCC model for data envelopment analysis," *Fuzzy Optimization and Decision Making*, vol. 2, pp. 337-358, 2003.

[11] A. Hatami-Marbini, A. Emrouznejad, and M. Tavana, "A taxonomy and review of the fuzzy data envelopment analysis literature: Two decades in the making," *European Journal of Operational Research*, vol. 214, pp. 457-472, 2011.

[12] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. l8, pp. 338-353, 1965.

[13] A. Cel m̃ins, "Least squares model fitting to fuzzy vector data," *Fuzzy Sets and Systems*, vol. 22, pp. 245-269, 1987.

[14] H. Tanaka and P. Guo, *Possibilistic Data Analysis for Operations Research*, NY: Springer-Verlag, 1999.

[15] T. Allahviranloo, A. A. Hosseinzadeh, M. Ghanbari, E. Haghi, and R. Nuraei, "On the new solutions for a fully fuzzy linear system," *Soft Computing*, vol. 18, pp. 95-107, 2014.

[16] C. Roldán, A. Roldán, and J. Martínez-Moreno, "A fuzzy regression model based on distances and random variables with crisp input and fuzzy output data: a case study in biomass production," *Soft Computing*, vol. 16, pp. 785-795, 2012.

[17] H. K. Kim, J. H. Yoon, and Y. Li, "Asymptotic properties of least squares estimation with fuzzy observations," *Information Science*, vol. 178, pp. 439-451, 2008.

[18] L. H. Chen and C. C. Hsueh, "Fuzzy regression models using the least-squares method based on the concept of distance," *IEEE Trans. on Fuzzy System*, vol. 17, pp. 1259-1272, 2009.

[19] J. Li, W. Zeng, J. Xie, and Q. Yin, "A new fuzzy regression model based on least absolute deviation," *Engineering Applications of Artificial Intelligence*, vol. 52, pp. 54-64, 2016.

[20] S. H. Choi and J. J. Buckley, "Fuzzy regression using least absolute deviation estimators," *Soft Computing*, vol. 12, pp. 257-263, 2008.

[21] K. J. Kim, D. H. Kim, and S. H. Choi, "Least absolute deviation estimator in fuzzy regression," *Journal of Computational and Applied Mathematics*, vol. 18, pp. 649-656, 2005.

[22] H. Torabi, and J. Behboodian, "Fuzzy least-absolutes estimates in linear models," *Communications in Statistics - Theory and Methods*, vol. 36, pp. 1935-1944, 2007.

[23] S. M. Taheri and M. Kelkinnama, "Fuzzy linear regression based on least absolute deviations," *Iranian Journal of Fuzzy Systems*, vol. 9, pp. 121-140, 2012.

[24] W. Stahel and S. Weisberg, *Directions in Robust Statistics and Diagnostics*. NY: Springer-Verlag, 2012.

[25] D. H. Haas, "Evaluating the efficiency of municipal reverse logistics channels: An application of data envelopment analysis," Ph.D. dissertation, Temple Univ., 1998.

W.-C. Yeh received the M.S. and Ph.D. degrees from the Department of Industrial Engineering, University of Texas at Arlington, Arlington, TX, USA, in 1990 and 1992, respectively. He is currently a distinguished professor in the Department of Industrial Engineering and Engineering Management, National Tsing Hua University in Taiwan. His research interests include network reliability theory, graph theory, the deadlock problem, linear programming, and scheduling. Prof. Yeh is a Member of INFORMS.

C.-K. Hu received the Ph.D. degree from the Department of Management, I-Shou University, Kaohsiung, Taiwan, in 2004. He is currently an associated professor in the Department of Tourism Management, Kao Yuan University in Taiwan. His research interests include fuzzy decision making on business and management.

F.-B. Liu received the Ph.D. degree from the Department of Mechanical Engineering, North Carolina University, Raleigh, NC, USA, in 1995. He is currently a PROFESSOR in the Department of Mechanical and Automation Engineering, I-Shou University in Taiwan. His research interests include reverse engineering, fuzzy optimization applied on heat transfer, and fluid dynamics.

C.-F. Hu received the M.S. and Ph.D. Degree from the Operations Research Program, North Carolina State University, Raleigh, NC, USA, in 1995 and 1997, respectively. She is currently a professor in the Department of Applied Mathematics, National Chiayi University in Taiwan. Her research interests include fuzzy optimization and decision making.