

行政院國家科學委員會專題研究計畫 成果報告

模糊樣本資料之新統計檢定程序與決策(第3年) 研究成果報告(完整版)

計畫類別：個別型
計畫編號：NSC 96-2416-H-004-014-MY3
執行期間：98年08月01日至99年07月31日
執行單位：國立政治大學應用數學學系

計畫主持人：吳柏林

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫涉及專利或其他智慧財產權，1年後可公開查詢

中 華 民 國 99 年 09 月 30 日

行政院國家科學委員會補助專題研究計畫 ☒ 成果報告
☐ 期中進度報告

模糊樣本資料之新統計檢定程序與決策

計畫類別：☒ 個別型計畫 ☐ 整合型計畫

計畫編號：NSC962416H004014-MY3

執行期間：96 年 8 月 1 日至 99 年 7 月 31 日

執行機構及系所：政大應數系

計畫主持人：吳柏林

共同主持人：

計畫參與人員：

成果報告類型(依經費核定清單規定繳交)：☐ 精簡報告 ☒ 完整報告

本計畫除繳交成果報告外，另須繳交以下出國心得報告：

☐ 赴國外出差或研習心得報告

☐ 赴大陸地區出差或研習心得報告

☒ 出席國際學術會議心得報告

☐ 國際合作研究計畫國外研究報告

處理方式：除列管計畫及下列情形者外，得立即公開查詢

☐ 涉及專利或其他智慧財產權，☐ 一年 ☐ 二年後可公開查詢

中 華 民 國 99 年 8 月 1 日

1. Introduction

Point forecasting gives important information during decision-making processes, especially on economic developments, population policies, management planning or financial controls. Nevertheless, its drawbacks still include: (1) it is not efficient in the marketing application due to the inaccurate forecasting; (2) the model constructed only by the closing price may not illustrate the whole process of daily or monthly trend. The causes of this regretful situation are that the business marketing is full of uncertainty and human being's manipulation. In this paper, we propose interval forecasting approaches, such as the interval moving average, the weighted interval moving average, and ARIMA interval forecasting. We generate interval time series by simulation and apply the proposed forecasting approaches to carry out the interval forecasting. The forecast results are compared by the mean squared interval error and the mean relative interval error. Finally, we take the monthly trading prices of China Steel stock as a case study. In the comparison of forecasting performance, it is found that ARIMA interval forecasting provides more efficiency and flexibility than the traditional ones do.

In the forecasting and analysis of time series, the tendency of the data interval is paid more and more attention, such as daily temperature changes, the fluctuation of the exchange rate, the level price of petroleum etc. According to the characteristic of observed data, traditional time series analysis selects the best fitted model from *a priori* models, such as ARIMA model, ARCH model or the threshold model etc. Due to the uncertainty of the predicted points, intervals are used as the estimated prediction values. Montgomery and Johnson (1976), Abraham and Ledolter (1983), Chatfield (1989) proposed prediction interval by time series of points to carry out the prediction.

Based on different factors and data types, there are a diversity of prediction methods. Chatfield (1993) made a comparison of several different methods. And Diebold and Mariano (1995) proposed the discussion in the respect that different prediction methods have their own pros and cons depending on the time series has a steady tendency or a severe fluctuation. Christoffersen (1998) provided the calculation method of interval forecasting for the risk measurement. Despite there are various methods of interval forecasting due to the variety of the backgrounds and the purposes of research, the data collections are mostly in the basic form of single real numerical variable. Even though there are many interval prediction methods, Chatfield (1996) pointed out that the prediction intervals resulted by those methods are too narrow owing to a lot of factors, one of which is the problem whether the model is appropriate.

This paper proposes the interval forecasting that uses interval time series to carry out the prediction and analysis of intervals, whereas actually the literature of dynamic data analysis applied to interval data is few. We employ the methods of AR(1) and ARCH(1) to make multiple simulations on each single time point in order to find the interval of each time point to design four sets of stable or unstable interval time series. Thus, we can proceed the interval forecasting in accordance with the proposed forecasting methods, which are the interval moving average of order k , the weighted interval moving average of order k , ARIMA((p_c, d_c, q_c) , (p_r, d_r, q_r)) of (c, r) and ARIMA((p_m, d_m, q_m) , (p_l, d_l, q_l) , (p_u, d_u, q_u)) of (m, l, u) .

However, when ARIMA is applied to the analysis of interval time series, the most important step is to consider how to associate the central point and the interval length to make an integrated analysis. Therefore, we denote an interval by the central point and the interval radius instead of the traditional interval expression, complying with the rules of interval operations established by Young (1931) and the properties of the interval algorithm by Dwyer (1951). This paper applies the interval calculation technology to explore the model construction and forecasting of interval time series. First, we must give the appropriate definitions of random interval and its expectation. Then ARIMA is employed to construct the models of the central points and the interval lengths separately. At last, the interval forecasting can be carried out by means of the constructed models.

It is necessary to determine the validity of the forecasting method by means of the estimated error between the predicted data and the simulated data. Chatfield (1996) declared that the error made by an inappropriate interval prediction method is more severe than the error made by a simple point prediction. Therefore, in order to assess the efficiency of interval forecasting, we define several criteria, referring to the traditional efficiency measurements, to evaluate the efficiency of forecasting. We make an integrated analysis of the efficiency by associating the central point and length of the interval, which are the mean squared error

of interval and the mean relative interval error respectively. Finally, regardless of the simulated interval time series or the practical data, using ARIMA to proceed the interval forecasting is more appropriate than using traditional prediction technologies, such as moving average and weighted moving average.

In this paper, we propose the model construction and its theoretical architecture of interval time series analysis and forecasting. We use the monthly highest and lowest prices of stocks as our case study. The mean squared error of interval is taken into account to be compared with the mean of relative interval error to validate the forecasting effect, which will be helpful for the study and judgment on the trend of the future financial market. The interval prediction by interval time series can provide more objective decision space to policymakers.

2. Analysis of Interval Time Series

2.1. The interval data

The traditional social and economic studies have brought in various analyses of interactive relationship and model related to human. In the traditional model construction, we often confront the uncertainty data problem. For instance, should we count the number of yearly enrolled students at the beginning of the year? at midyear? or at the end of the year? The obtained number is often different at different time. For another example, what is the exchange rate of U.S. dollar to Japanese Yen last week? is with the opening quotation? the closing price? or the average of the highest price and the bottom price? The results are also quite different. Wu and Chen (1999) have given an extensive review of literature on this topic. In the social science study or economic research, obviously the answer to these questions is not just true or false. There are lots of uncertain and incomplete information or events so that we can not apply the conventional real number system to process it. In many practical cases under the influences of many factors, the observed data usually appear not only a single numerical value type but also a “range”.

2.2. The operation of interval data

While we consider the data to be of interval type, we must encounter the various problems of interval operations as well as the realistic meanings. Dwyer (1951) called intervals as "range numbers" and defined the relevant operations. The subsequent studies relative to the interval operations continually quote such definitions. Nevertheless, it is still unable to give the standard rules of interval operations on the computer hardware. Hayes (2003) pointed out that the rules of interval operations seem simple, but there often appears a trap of miscalculation in the practical calculations, especially in the case that the element 0 is contained in an interval. The traditional interval operations are given as follows:

$$[a, b] \circ [c, d] = [\min(a \circ c, a \circ d, b \circ c, b \circ d), \max(a \circ c, a \circ d, b \circ c, b \circ d)],$$

where \circ denotes any of $+$, $-$, \times and \div . That means that the interval operation takes the extreme values of $a \circ c$, $a \circ d$, $b \circ c$ and $b \circ d$ as the resulted interval. This definition is feasible for addition, subtraction and multiplication. But in the case of division, such interval operation will lead to a blind spot in computation. For example, $[3, 5] \div [-1, 1]$ will obtain the answer $[-5, 5]$. But actually $0 \in [-1, 1]$ and $[3, 5] \div 0$ is not well defined, so that such the answer $[-5, 5]$ is meaningless.

Furthermore, when we process a set of dynamic data represent by the interval form, we often encounter certain realistic dilemma. For instance, does the value increase or decrease from $[2, 8]$ to $[3, 5]$? We may consider the location variation between intervals. But if we take the interval scale into considerations, this case becomes more complicated. Hence, in this paper we will propose the bivariate parameters, which are (i) the interval radius to express the interval scale and (ii) the interval center to express the interval location, to demonstrate the variations of intervals. For the above example, the interval location is decreased from 5 to 4, whereas the interval radius is also decreased from 3 to 1. Such a new interval expression, which integrates the interval center with the interval radius, is apt to show the

location shift and the length variation of interval data. By the proposed interval expression, we are able to make an appropriate interval forecasting for the interval time series. And under the assistance of computer programs, even more complicated calculation can be easily solved.

2.3. Definitions and Properties of Interval Time Series

The interval time series is the analytical method to apply intervals to the analysis of time series, incorporating with the interval operations, so as to solve the uncertainty of the data. As a result, before constructing and forecasting the model of interval time series, we must first give several definitions relevant to interval time series. The definition of a random interval defined by Nguyen and Wu (2006) is given in Definition 2.1, and the other relevant definitions are given as follows.

Definition 2.1 (A random interval)

Let $X = [a, b]$. If a and b are random variables, then the interval $[a, b]$ is called a random interval.

Definition 2.2 (A random interval of alternate notation, $X = (c, r)$)

Suppose $X = [a, b]$ be a random interval over the real numbers \mathcal{R} , $c = (a + b)/2$ be the center of the interval X , and $r = (b - a)/2$ be the radius of the interval X , then the interval X can be expressed as $X = (c, r)$.

Definition 2.3 (The mean of a random interval $X = (c, r)$)

Let $X = (c, r)$ be a random interval over the real numbers \mathcal{R} . The mean of the random interval X is defined as $E[X] = E[c, r] = (E[c], E[r])$, where $E[c]$ and $E[r]$ are the expectations of the random variables c and r respectively.

In some special conditions, we could know there is a value which manipulates the position and the length of the interval X . For example, if the daily highest price of a stock is b , the daily lowest price is a , and the weighted average price is m , then m is not necessarily the median value c . Hence the distances between m and b is different from that between m and a . There is a similar situation in the weather forecast. If the daily average temperature is m , obviously m is not likely the median value of the highest and lowest temperatures. Therefore, m is called the weighted center, defined as follows.

Definition 2.4 (The weighted center of random interval, m)

Let $X = [a, b]$ be a random interval over the real numbers \mathcal{R} . If there is such a value, denoted as m , which influence position and the length of the random interval X , then the value m is called the weighted center of random interval X .

Definition 2.5 (A random interval, $X = (m, l, u)$)

Suppose $X = [a, b]$ is a random interval over the real numbers \mathcal{R} . Let m be the weighted center of the interval X , $l = m - a$ and $u = b - m$ be the distances between the weighted center and the boundaries of interval X respectively. Then the interval X can be expressed as $X = (m, l, u)$.

Definition 2.6 (The mean of random interval $X = (m, l, u)$)

Let $X = (m, l, u)$ be a random interval over the real numbers \mathcal{R} . The mean of the random interval X is defined as $E[X] = E[m, l, u] = (E[m], E[l], E[u])$, where $E[m]$, $E[l]$ and $E[u]$ are the expectations of random

variables m , l and u respectively.

Definition 2.7 (The interval addition)

Let $X_1 = [a_1, b_1] = (c_1, r_1)$ and $X_2 = [a_2, b_2] = (c_2, r_2)$ be random intervals. The interval addition is defined as follows:

$$\text{Interval addition: } X_1 \oplus X_2 = (c_1, r_1) \oplus (c_2, r_2) = (c_1 + c_2, r_1 + r_2),$$

Definition 2.8 (The interval scalar multiplication)

If $X = [a, b] = (c, r)$ is a random interval and k is a scalar, then the scalar multiplication of X by k is the interval given by $kX = k(c, r) = (kc, |k|r)$.

From the definition 2.7 and 2.8, the the interval subtraction can be obtained as follows:

$$X_1 \ominus X_2 = X_1 \oplus (-X_2) = (c_1, r_1) \oplus (-c_2, r_2) = (c_1 - c_2, r_1 + r_2).$$

Remark1

1. The interval additive identity is the zero interval: $O = [0, 0] = (0, 0)$.
2. The interval additive inverse: $-X = [-b, -a] = (-c, r)$.
3. The interval absolute value: $|X| = |(c, r)| = \begin{cases} \left(\frac{|c|+r}{2}, \frac{|c|+r}{2} \right) & \text{if } 0 \in (c, r) \\ (|c|, r) & \text{if } 0 \notin (c, r). \end{cases}$

Definition 2.9 (The interval length)

Let $X = [a, b] = (c, r)$ be a random interval, the interval length of X is $2r = b - a$, and denoted as $\|X\| = 2r$.

Definition 2.10 (The distance between two intervals)

Let $X_1 = [a_1, b_1] = (c_1, r_1)$ and $X_2 = [a_2, b_2] = (c_2, r_2)$ be two random intervals, then the distance between X_1 and X_2 is defined as $D(X_1, X_2) = \|X_1 \ominus X_2\| = 2(r_1 + r_2)$.

Property 2.1

$$|X_1 \ominus X_2| = 0 \text{ if only if } r_1 = 0 \text{ and } r_2 = 0.$$

Definition 2.11 (Positive intervals and negative intervals)

Let $X = [a, b] = (c, r)$ be a random interval. If the range of X is all positive, i.e. $x \geq 0$ for all $x \in X$, then X is a positive interval. If the range of X is all negative, i.e. $x \leq 0$ for all $x \in X$, then X is a negative interval.

Definition 2.12 (The interval multiplication)

If $X_1 = [a_1, b_1] = (c_1, r_1)$ and $X_2 = [a_2, b_2] = (c_2, r_2)$ are two random intervals, the interval multiplication of X_1 and X_2 is denoted as $X_1 \otimes X_2$.

Because the interval multiplication involves the problems of changing sign and whether zero is in intervals or not, there should consider several situations, described separately as follows.

- (Case 1) If 0 does not belong to any of the intervals X_1 and X_2 , and the two intervals X_1 and X_2 both are *positive intervals* or *negative intervals*, then $X_1 \otimes X_2 = (c_1 c_2 + r_1 r_2, |c_1 r_2 + c_2 r_1|)$.
- (Case 2) If the intervals X_1 and X_2 do not contain 0, and one of them is a positive interval and the other is a negative interval, then $X_1 \otimes X_2 = (c_1 c_2 - r_1 r_2, |c_1 r_2 - c_2 r_1|)$.
- (Case 3) If an interval includes 0, there included both positive and negative numbers in the range of the interval. Thus, if either interval contains 0, the calculation of interval multiplication becomes much more complicated. To simplify the problem, let zero be separated from the interval. We can assume the interval X_1 contains 0 without loss of generality, let $X_1 = [c_1 - r_1, 0] \cup [0, c_1 + r_1] = (c'_1, r'_1) \oplus (c''_1, r''_1) = A_1 \oplus A_2$, where $c'_1 = \frac{c_1 - r_1}{2}$, $r'_1 = \frac{c_1 - r_1}{2}$, $c''_1 = \frac{c_1 + r_1}{2}$, $r''_1 = \frac{c_1 + r_1}{2}$. Notice $A_1 = (c'_1, r'_1)$ is a negative interval and $A_2 = (c''_1, r''_1)$ is a positive interval. Therefore, the interval multiplication can be expressed as $X_1 \otimes X_2 = (A_1 \oplus A_2) \otimes X_2 = (A_1 \otimes X_2) \oplus (A_2 \otimes X_2)$. Then $A_1 \otimes X_2$ and $A_2 \otimes X_2$ satisfy the conditions in (Case 1) and (Case 2). As a result, the interval multiplication $X_1 \otimes X_2$ can be obtained by the interval addition of $A_1 \otimes X_2$ and $A_2 \otimes X_2$.
- (Case 4) If both intervals X_1 and X_2 contain 0, then $X_1 = A_1 \oplus A_2$ and $X_2 = B_1 \oplus B_2$, where A_1, B_1 are negative intervals, and A_2, B_2 are positive intervals. As the same description in (Case 3), the product interval $X_1 \otimes X_2$ can be calculated. Because the calculation is much more complex and is not the main issue in this paper, the detail is not discussed here.

Definition 2.14 (Interval time series)

An interval time series is a sequence of random intervals, $X_t = [a_t, b_t] = (c_t, r_t)$, $t = 1, 2, 3, \dots$, denoted as $\{X_t\} = \{X_t = [a_t, b_t] = (c_t, r_t) \mid t = 1, 2, 3, \dots\}$.

Definition 2.15 (Stationary interval time series of order k)

Let $\{X_t\}$ be an interval time series. If $X_{t+k} - X_t = \Delta_k = (\delta_{c_k}, \varepsilon_{r_k})$ for any $t = 1, 2, 3, \dots$, where $E[\delta_{c_k}] = \delta$ and $E[\varepsilon_{r_k}] = \varepsilon$, then $\{X_t\}$ is called a *stationary interval time series*.

Property 2.2 Let $\{X_t\} = \{X_t = [a_t, b_t] = (c_t, r_t) \mid t = 1, 2, 3, \dots\}$ be a stationary interval time series, then both $\{c_t\}$ and $\{r_t\}$ are stationary time series.

Example 2.1 Let $A = [1, 3] = (2, 1)$, $B = [2, 6] = (4, 2)$, then

$$\begin{aligned} A \oplus B &= (2, 1) \oplus (4, 2) = (2 + 4, 1 + 2) = (6, 3), \\ A \ominus B &= (2, 1) - (4, 2) = (2 - 4, 1 + 2) = (-2, 3), \\ A \otimes B &= (2, 1) \otimes (4, 2) = (2 \cdot 4 + 1 \cdot 2, |2 \cdot 2 + 4 \cdot 1|) = (10, 8), \\ 3A &= 3(2, 1) = (3 \cdot 2, 3 \cdot 1) = (6, 3). \end{aligned}$$

2.4. The Forecasting Models of Interval Time Series

A traditional time series is defined as $\{X_t = x_t, t = 1, 2, 3, \dots\}$, and the corresponding prediction $\hat{X}_t = E[X_t | X_{t-1}, X_{t-2}, \dots, X_1]$ is a point prediction. The traditional prediction model for traditional time series is not capable of being applied directly to the forecasting of interval time series $\{X_t = [a_t, b_t] = (c_t, r_t), t = 1, 2, 3, \dots\}$. The followings are some forecasting models of interval time series.

(1) Interval moving average of order k (IMA)

Let $\hat{c}_t = \frac{c_{t-1} + \dots + c_{t-k}}{k}$, $\hat{r}_t = \frac{r_{t-1} + \dots + r_{t-k}}{k}$, $t = k+1, k+2, k+3, \dots$, then the forecasting of interval time series is $E[X_t | X_{t-1}, X_{t-2}, \dots, X_{t-k}] = [\hat{c}_t - \hat{r}_t, \hat{c}_t + \hat{r}_t] = (\hat{c}_t, \hat{r}_t)$.

(2) Weighted interval moving average of order k (WIMA)

Let $\hat{c}_t = \sum_{i=t-k}^{t-1} p_i c_i$, $\hat{r}_t = \sum_{i=t-k}^{t-1} p_i r_i$, for $t = k+1, k+2, k+3, \dots$, where $p_i = \frac{f_i}{\sum_{j=t-k}^{t-1} f_j}$, $f_i = \frac{\|X_i\|}{\|R\|}$, for $i = t-k, t-k+1, \dots, t-1$, and $R = \left[\min_{t-k \leq j \leq t-1} \{a_j\}, \max_{t-k \leq j \leq t-1} \{b_j\} \right]$, $\|\cdot\|$ denotes the length of an interval. Then the forecasting of interval time series is $E[X_t | X_{t-1}, X_{t-2}, \dots, X_{t-k}] = [\hat{c}_t - \hat{r}_t, \hat{c}_t + \hat{r}_t] = (\hat{c}_t, \hat{r}_t)$.

(3) $ARIMA((p_c, d_c, q_c), (p_r, d_r, q_r))$ of (c, r) interval forecasting ($ARIMA$ IF)

If $\{X_t = (c_t, r_t)\}$ is a stationary interval time series, then $\{c_t\}$ and $\{r_t\}$ are stationary time series, and

$$\begin{aligned} c_t &= \theta + \phi_1 c_{t-1} + \dots + \phi_{p_c} c_{t-p_c} + \delta_t - \theta_1 \delta_{t-1} - \dots - \theta_{q_c} \delta_{t-q_c}, \\ r_t &= \alpha + \beta_1 r_{t-1} + \dots + \beta_{p_r} r_{t-p_r} + \varepsilon_t - \eta_1 \varepsilon_{t-1} - \dots - \eta_{q_r} \varepsilon_{t-q_r} \end{aligned}$$

where $\delta_t \sim WN(0, \sigma_c^2)$ and $\varepsilon_t \sim WN(0, \sigma_r^2)$. Therefore, $E[c_t | c_{t-1}, c_{t-2}, \dots, c_1] = \theta + \phi_1 c_{t-1} + \dots + \phi_{p_c} c_{t-p_c}$, and $E[r_t | r_{t-1}, r_{t-2}, \dots, r_1] = \alpha + \beta_1 r_{t-1} + \dots + \beta_{p_r} r_{t-p_r}$.

Let $\hat{c}_t = E[c_t | c_{t-1}, c_{t-2}, \dots, c_1]$ and $\hat{r}_t = E[r_t | r_{t-1}, r_{t-2}, \dots, r_1]$, then the forecasting of interval time series is $E[X_t | X_{t-1}, X_{t-2}, \dots, X_1] = [\hat{c}_t - \hat{r}_t, \hat{c}_t + \hat{r}_t] = (\hat{c}_t, \hat{r}_t)$.

(4) $ARIMA((p_m, d_m, q_m), (p_l, d_l, q_l), (p_u, d_u, q_u))$ of (m, l, u) interval forecasting

Let $\{X_t = (m_t, l_t, u_t)\}$ be a stationary interval time series, then $\{m_t\}$, $\{l_t\}$ and $\{u_t\}$ are all stationary time series, and

$$\begin{aligned} m_t &= \theta + \phi_1 m_{t-1} + \dots + \phi_{p_m} m_{t-p_m} + \delta_t - \theta_1 \delta_{t-1} - \dots - \theta_{q_m} \delta_{t-q_m} \\ l_t &= \alpha + \beta_1 l_{t-1} + \dots + \beta_{p_l} l_{t-p_l} + \varepsilon_{l,t} - \eta_1 \varepsilon_{l,t-1} - \dots - \eta_{q_l} \varepsilon_{l,t-q_l} \\ u_t &= \tau + \omega_1 u_{t-1} + \dots + \omega_{p_u} u_{t-p_u} + \varepsilon_{u,t} - \varphi_1 \varepsilon_{u,t-1} - \dots - \varphi_{q_u} \varepsilon_{u,t-q_u} \end{aligned}$$

where $\delta_t \sim WN(0, \sigma_m^2)$, $\varepsilon_{l,t} \sim WN(0, \sigma_l^2)$, and $\varepsilon_{u,t} \sim WN(0, \sigma_u^2)$. Then

$$\begin{aligned} E[m_t | m_{t-1}, m_{t-2}, \dots, m_1] &= \theta + \phi_1 m_{t-1} + \dots + \phi_{p_m} m_{t-p_m}, \\ E[l_t | l_{t-1}, l_{t-2}, \dots, l_1] &= \alpha + \beta_1 l_{t-1} + \dots + \beta_{p_l} l_{t-p_l}, \\ E[u_t | u_{t-1}, u_{t-2}, \dots, u_1] &= \tau + \omega_1 u_{t-1} + \dots + \omega_{p_u} u_{t-p_u}. \end{aligned}$$

And let $\hat{m}_t = E[m_t | m_{t-1}, m_{t-2}, \dots, m_1]$, $\hat{l}_t = E[l_t | l_{t-1}, l_{t-2}, \dots, l_1]$, $\hat{u}_t = E[u_t | u_{t-1}, u_{t-2}, \dots, u_1]$, then the forecasting of interval time series is

$$E[X_t | X_{t-1}, X_{t-2}, \dots, X_1] = [\hat{m}_t - \hat{l}_t, \hat{m}_t + \hat{u}_t] = (\hat{m}_t, \hat{l}_t, \hat{u}_t).$$

Example 2.2 Let $\{X_t\} = \{[1, 2], [2, 4], [3, 4], [4, 6], [3, 7]\}$. That is $X_1 = [1, 2] = (1.5, 0.5)$, $X_2 = [2, 4] = (3, 1)$, $X_3 = [3, 4] = (3.5, 0.5)$, $X_4 = [4, 6] = (5, 1)$, and $X_5 = [3, 7] = (5, 2)$. Then by the interval moving average of order 5, we can obtain

$$E[X_6|X_5, X_4, X_3, X_2, X_1] = \left(\frac{1.5+3+3.5+5+5}{5}, \frac{0.5+1+0.5+1+2}{5} \right) \\ = (3.6, 1) = [3.6-1, 3.6+1] = [2.6, 4.6]$$

If the range of intervals is considered as $\Omega=[1, 7]$, then the full length is $R=6$, and $f_1=\frac{1}{6}$, $f_2=\frac{2}{6}$, $f_3=\frac{1}{6}$,

$f_4=\frac{2}{6}$, $f_5=\frac{4}{6}$, we have $p_1=\frac{\frac{1}{6}}{\frac{1}{6}+\frac{2}{6}+\frac{1}{6}+\frac{2}{6}+\frac{4}{6}}=0.1$, $p_2=0.2$, $p_3=0.1$, $p_4=0.2$, $p_5=0.4$. Then the

weighted interval moving average of order 5 is

$$E[X_6|X_5, X_4, X_3, X_2, X_1] \\ = (0.1 \times 1.5 + 0.2 \times 3 + 0.1 \times 3.5 + 0.2 \times 5 + 0.4 \times 5; 0.1 \times 0.5 + 0.2 \times 1 + 0.1 \times 0.5 + 0.2 \times 1 + 0.4 \times 2) \\ = (4.1, 1.3) = [2.8, 5.4]$$

3. The Efficiency Analysis of Interval Time Series Forecasting

The quality of the forecasting result is the most concern of the analysts. In a traditional forecast of time series, it is to compare the distances between the actual values and the predicted values to assess the quality of forecasting. With regard to the interval forecasting, not only the forecasting of interval length, we are also concerned with the location disparity between the predicted interval and the actual interval. Therefore, traditional methods to evaluate the forecasting efficiency of time series are unable to analyze the forecasting efficiency of interval time series. The following will define the criteria for analyzing the efficiency of interval forecasting.

3.1. The Efficiency of Interval Forecasting

Since an interval time series can be represented by either $\{X_t=(c_t, r_t)\}$ or $\{X_t=(m_t, l_t, u_t)\}$, the definitions will be given with respect to these two cases.

Definition 3.1 (Mean squared error of interval position, MSEP)

(1) Let $\{X_t=(c_t, r_t)\}$ be an interval time series, $\hat{X}_t=(\hat{c}_t, \hat{r}_t)$ be the predicted interval. If $\delta_{c_t}=c_t-\hat{c}_t$ is the position error between \hat{X}_t and X_t , then the mean squared error of interval position (MSEP) is given by

$$MSEP = \frac{\sum_{t=1}^s \delta_{c_{n+t}}^2}{s} = \frac{\sum_{t=1}^s (c_{n+t} - \hat{c}_{n+t})^2}{s},$$

where n denotes the current time, s is the number of the preceding intervals, and \hat{c}_t is the estimation of c_t .

(2) Let $\{X_t=(m_t, l_t, u_t)\}$ be an interval time series and $\hat{X}_t=(\hat{m}_t, \hat{l}_t, \hat{u}_t)$ be the predicted interval. If $\delta_{m_t}=m_t-\hat{m}_t$ is the position error between \hat{X}_t and X_t , then the mean squared error of interval position (MSEP) is defined by

$$MSEP = \frac{\sum_{t=1}^s \delta_{m_{n+t}}^2}{s} = \frac{\sum_{t=1}^s (m_{n+t} - \hat{m}_{n+t})^2}{s},$$

where n denotes the current time, s is the number of the preceding intervals, and \hat{m}_t is the estimation of m_t .

Definition 3.2 (Mean squared error of interval length, MSEL)

(1) If the interval time series is $\{X_t = (c_t, r_t)\}$, the predicted interval is $\hat{X}_t = (\hat{c}_t, \hat{r}_t)$, and $\varepsilon_{r_t} = r_t - \hat{r}_t$ is the error between the length of the predicted interval \hat{X}_t and that of the actual interval X_t , then the mean squared error of interval length (MSEL) is given by

$$MSEL = \frac{\sum_{t=1}^s \varepsilon_{r_{n+t}}^2}{s} = \frac{\sum_{t=1}^s (r_{n+t} - \hat{r}_{n+t})^2}{s},$$

where n represents the current time, s is the number of the previous intervals, and \hat{r}_t is the estimation of r_t .

(2) If the interval time series is $\{X_t = (m_t, l_t, u_t)\}$, the predicted interval is $\hat{X}_t = (\hat{m}_t, \hat{l}_t, \hat{u}_t)$, $\varepsilon_{l_t} = l_t - \hat{l}_t$ and $\varepsilon_{u_t} = u_t - \hat{u}_t$ are the length errors between \hat{X}_t and X_t , then the mean squared error of interval length (MSEL) is defined as

$$MSEL = \frac{\sum_{t=1}^s (\varepsilon_{l_{n+t}}^2 + \varepsilon_{u_{n+t}}^2)}{2s} = \frac{\sum_{t=1}^s (l_{n+t} - \hat{l}_{n+t})^2 + \sum_{t=1}^s (u_{n+t} - \hat{u}_{n+t})^2}{2s},$$

where n represents the current time, s is the number of the preceding intervals, and \hat{l}_t and \hat{u}_t are the estimations of l_t and u_t respectively.

Example 3.1 Let the interval time series be $X_1 = [4, 6] = (5, 1)$, $X_2 = [5, 8] = (6.5, 1.5)$, the predicted intervals are $\hat{X}_1 = [2.8, 5.4] = (4.1, 1.3)$ and $\hat{X}_2 = [3.8, 7.8] = (5.8, 2)$. Then the mean squared error of interval position is given by

$$MSEP = \frac{(5 - 4.1)^2 + (6.5 - 5.8)^2}{2} = 0.65,$$

and the mean squared error of interval length is given by

$$MSEL = \frac{(1 - 1.3)^2 + (1.5 - 2)^2}{2} = 0.17.$$

3.2. The Integrated Efficiency of Interval Forecasting

Definition 3.3 (Mean squared error of interval, MSEI)

Let $\{X_t = (c_t, r_t)\}$ be the interval time series. The error between the predicted interval $\hat{X}_t = (\hat{c}_t, \hat{r}_t)$ and the actual interval X_t consists of two parts; the position error δ_c and the length error ε_r . Then, the mean squared error of interval (MSEI) is given by

$$MSEI = \frac{\sum_{t=1}^l (c_{n+t} - \hat{c}_{n+t})^2}{s} + \frac{\sum_{t=1}^l (r_{n+t} - \hat{r}_{n+t})^2}{s} = MSEP + MSEL,$$

where n represents the current time, s is the number of the preceding intervals, and \hat{c}_t is the estimation of c_t , \hat{r}_t is the estimation of r_t .

If the interval time series is $\{X_t = (m_t, l_t, u_t)\}$, then the mean squared error of interval (MSEI) is given by

$$MSEI = \frac{\sum_{t=1}^s (m_{n+t} - \hat{m}_{n+t})^2}{s} + \frac{\sum_{t=1}^s (l_{n+t} - \hat{l}_{n+t})^2 + \sum_{t=1}^s (u_{n+t} - \hat{u}_{n+t})^2}{2s} = MSEP + MSEL,$$

where n represents the current time, s is the number of the preceding intervals, and \hat{m}_t , \hat{l}_t and \hat{u}_t are the estimations of m_t , l_t and u_t respectively.

As described in Section 3.1, the error between the predicted interval \hat{X}_t and the actual interval X_t contains two parts; the position error ε_c and the length error ε_r . The former is the distance between the central points of two intervals, while the latter is the difference between the radii of two intervals. If the mean squared errors of the location and the scale are always summed up, it will be hard to discern the efficiencies of the forecasting method between the position and the length.

Consider the interval $X = [4, 7] = (5, 1.5)$, and the predicted intervals $\hat{X}_1 = [1, 8] = (4.5, 3.5)$ and $\hat{X}_2 = [8, 10] = (9, 1)$ obtained by two different forecasting methods. Although the radius of \hat{X}_1 is larger than that of \hat{X}_2 , the central point of \hat{X}_1 is closer to the central point of X . Besides, the range of \hat{X}_1 covers the range of the actual interval X , while the range of \hat{X}_2 does not. As a result, we still regard \hat{X}_1 as the better predicted interval. Consequently, while considering the efficiency of the interval forecasting, it is the most important whether the predicted interval does cover the actual interval. Explicitly speaking, forecasting result is better if the center \hat{c} of the predicted interval is closer to the center c of the actual interval and their overlap is larger. By combining the two factors of the center and the radius of interval, we have three decision conditions; (1) when $\frac{|c - \hat{c}|}{r + \hat{r}} < 1$, there is overlap of the predicted and the actual intervals, it means that the interval forecasting is better; (2) when $\frac{|c - \hat{c}|}{r + \hat{r}} \ll 1$, it means that there is more overlap so that the interval forecasting is much better; (3) while $\frac{|c - \hat{c}|}{r + \hat{r}} \geq 1$, the predicted interval and the actual interval are completely separated, so the interval forecasting is undesirable. Because $\|X_t \Theta \hat{X}_t\| = 2(r + \hat{r})$, $\frac{2|c_t - \hat{c}_t|}{\|X_t \Theta \hat{X}_t\|}$ can be the criterion for the evaluating the forecasting. Therefore, we propose the following definition to be another criterion for analyzing the integrated efficiency of interval forecasting.

Definition 3.4 (Mean relative interval error, MRIE)

If the interval time series is $\{X_t = (c_t, r_t)\}$, the predicted interval is $\hat{X}_t = (\hat{c}_t, \hat{r}_t)$, and $\varepsilon_t = \frac{2|c_t - \hat{c}_t|}{\|X_t \Theta \hat{X}_t\|}$ is the

relative error between the predicted interval \hat{X}_t and the actual interval X_t , then the mean relative interval error (MRIE) is given by

$$MRIE = \frac{1}{s} \sum_{t=l}^s [\varepsilon_{n+t}] = \frac{1}{s} \sum_{t=l}^s \frac{2|c_{n+t} - \hat{c}_{n+t}|}{\|X_{n+t} \Theta \hat{X}_{n+t}\|}$$

where n denotes the current of time, s is the number of the preceding intervals, and \hat{c}_t is the estimation of c_t .

If $\{X_t = (m_t, l_t, u_t)\}$ is the interval time series, $\hat{X}_t = (\hat{m}_t, \hat{l}_t, \hat{u}_t)$ is the predicted interval, and

$\varepsilon_t = \frac{2(m_t - \hat{m}_t)}{\|X_t \Theta \hat{X}_t\|^{(*)}}$ is the relative error between the predicted interval \hat{X}_t and the actual interval X_t , then

the mean relative interval error (MRIE) is given by

$$MRIE = \frac{1}{s} \sum_{t=l}^s [\varepsilon_{n+t}] = \frac{1}{s} \sum_{t=l}^s \frac{2|m_{n+t} - \hat{m}_{n+t}|}{\|X_{n+t} \Theta \hat{X}_{n+t}\|},$$

where n represents the current of time, s is the number of the preceding intervals, and \hat{m}_t is the estimation of m_t .

(*) Note: When $m_t \geq \hat{m}_t$, $\|X_t \Theta \hat{X}_t\| = 2(l_t + \hat{u}_t)$; when $m_t < \hat{m}_t$, $\|X_t \Theta \hat{X}_t\| = 2(u_t + \hat{l}_t)$.

Example 3.2 Assume as in Example 3.1, then the mean squared error of interval is given by
 $MSEI = MSEP + MSEL = 0.65 + 0.17 = 0.82$.

Example 3.3 Assume as in Example 3.1, then the mean relative interval error is given by

$$MRIE = \frac{1}{2} \left(\frac{2|5 - 4.1|}{2(1 + 1.3)} + \frac{2|6.5 - 5.8|}{2(1.5 + 2)} \right) = 0.34.$$

4. Testing Hypothesis with Fuzzy Data

It is a new research topic about the hypothesis testing of fuzzy mean with interval values. First of all, we will give a brief definition about the defuzzification. Then under the fuzzy significant level δ , we make a one side or two side testing. These methods are a little different from traditional significant level α . In order to get the robustic properties, we will set up the rejection area level F_δ , according to the fuzzy population.

4.1 Testing Hypothesis for Fuzzy Equal

Let U be the universal set (a discussion domain), $L = \{L_1, L_2, \dots, L_k\}$ a set of k -linguistic variables on U , and $A = \{A_1, A_2, \dots, A_m\}$ and $B = \{B_1, B_2, \dots, B_n\}$ be two sets of fuzzy sample drawn from categorical populations with numbers on U . For each sample we assign a linguistic variable L_j and a normalized membership m_{ij} ($\sum_{j=1}^k m_{ij} = 1$), and let $Fn_{Aj} = \sum_{i=1}^m Ln_{ij}$, $Fn_{Bj} = \sum_{i=1}^n Ln_{ij}$ $j = 1, 2, \dots, k$ be the total memberships for each data set.. The following statements are process for testing hypothesis

Testing hypothesis of fuzzy equal for discrete fuzzy mean

Consider a k -cell multinomial vector $n = \{n_1, n_2, \dots, n_k\}$ with $\sum_i n_i = n$. The *Pearson Chi-squared test* ($\chi^2 = \sum_i \sum_j \frac{n_{ij} - e_{ij}}{e_{ij}}$) is a well known statistical test for investigating the significance of the differences between observed data arranged in k classes and the theoretically expected frequencies in the k classes. It is clear that the large discrepancies between the observed data and expected cell counts will result in larger values of χ^2

However, a somewhat ambiguous question is whether (quantitative) discrete data can be considered categorical data, for which the traditional χ^2 -test can be used. For example, suppose a child is asked the following question: “how much do you love your sister?” If the responses is a fuzzy number (say, 70% of the time), it is certainly inappropriate to use the traditional χ^2 -test for the analysis. We will present a χ^2 -test for fuzzy data as follows:

Procedures for Testing hypothesis of fuzzy equal for discrete fuzzy mean

1. *Hypothesis: Two populations have the same distribution ratio.*

2. *Statistics: $\chi^2 = \sum_{i \in A, B} \sum_{j=1}^c \frac{([Fn_{ij}] - e_{ij})^2}{e_{ij}}$. (In order to perform the Chi-square test for fuzzy data, we transfer the decimal fractions of Fn_{ij} in each cell of fuzzy category into the integer $[Fn_{ij}]$ by counting 0.5 or higher fractions as 1 and discard the rest.)*

3. Decision rule : under significance level α , if $\chi^2 > \chi^2_{\alpha}(k-1)$, then we reject H_0 .

Testing hypothesis of fuzzy index equal for discrete fuzzy mean

Let \overline{FX} be the fuzzy sample mean, \overline{X}_f be the defuzzification of \overline{FX} . Under the fuzzy significant level F_δ , and the corresponding critical value F_δ , we want to test $H_0: \overline{FX} = F\mu$, where $F\mu$ is the fuzzy mean of the underlying population. Let μ is the defuzzification value of $F\mu$, then the above hypothesis becomes $H_0: \mu = \mu_0$.

1 Hypothesis: $H_0: F\mu = F\mu_0$ vs. $H_1: F\mu \neq F\mu_0$.

2. Statistics: find \overline{FX} from a random sample $\{S_i, i = 1, \dots, n\}$.

3. Decision rule: under the fuzzy significant level F_δ , if $|\overline{X}_f - \mu_0| > \delta$, then reject H_0 .

Note: for left side test $H_0: \mu \leq \mu_0$ vs. $H_1: \mu > \mu_0$ under the fuzzy significant level F_δ , if $\mu_0 - \overline{X}_f > \delta$, we reject H_0 . The right hand side testing is similar.

Testing hypothesis with continuous fuzzy mean

1 Hypothesis: $H_0: F\mu =_F [a, b]$ vs. $H_1: F\mu \neq_F [a, b]$.

2. Statistics: find $\overline{FX} = [x_l, x_u]$ from a random sample $\{S_i, i = 1, \dots, n\}$.

3. Decision rule: under the significant level F_δ , find $k = \delta r$ (where $r = b - a$), if $|x_l - a| > k$ or $|x_u - b| > k$ then reject H_0 .

4.2 Testing Hypothesis for fuzzy belongs to

Testing of fuzzy belongs to with bounded sample

1 Hypothesis: $H_0: F\mu \in_F [a, b]$ vs. $H_1: F\mu \notin_F [a, b]$.

2. Statistics: find $\overline{FX} = [x_l, x_u]$ from a random sample $\{S_i, i = 1, \dots, n\}$.

3. Decision rule: under the significant level F_δ , find $k = \delta r$ (where $r = b - a$), if $x_l < a - k$ or $x_u > b + k$, then reject H_0 .

Testing of fuzzy belongs to with unbounded below sample

1 Hypothesis: $H_0: F\mu \in_F (-\infty, b]$ vs. $H_1: F\mu \notin_F (-\infty, b]$.

2. Statistics: find $\overline{FX} = (-\infty, x_u]$ from a random sample $\{S_i, i = 1, \dots, n\}$.

3. Decision rule: under the significant level F_δ , find $k = \delta r$ (where r is a constant), if $x_u > b + k$, then reject H_0 .

Testing of fuzzy belongs to with unbounded above sample

1 Hypothesis: $H_0: F\mu \in_F [a, \infty)$ vs. $H_1: F\mu \notin_F [a, \infty)$

2. Statistics: find $\overline{FX} = [x_l, \infty)$ from a random sample $\{S_i, i = 1, \dots, n\}$.

3. Decision rule: under the significant level F_δ , find $k = \delta r$ (where r is a constant), if $x_l < a - k$, then reject H_0

4. Case Analysis

In this section we use a practical case to illustrate the forecasting methods. The interval time series comes from the data that the highest price (b_t), the lowest price (a_t) and the weighted average price (m_t) of monthly trading values of China Steel stock from January 1995 to October 2005. Then the proposed methods are applied to make the interval forecasting for the last 6 intervals (from May to October). This data comes from the report of monthly trading value of individual stock provided by Taiwan Stock Exchange Corporation. The

chart of the interval time series is shown in Figure 2.

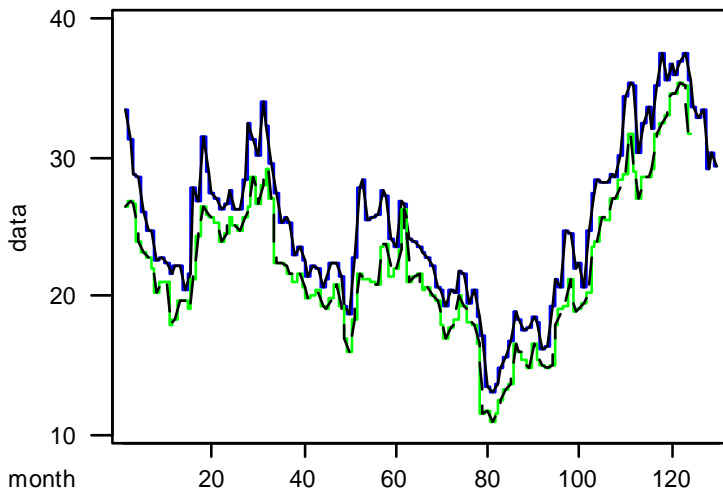


Figure 1 The chart of the monthly trading values of China Steel stock from 1/1995 to 4/2005.

Table 5.1 lists the monthly trading values of China Steel stock from May 2005 to October 2005 and their forecasts by the proposed interval forecasting methods described in Section 3.2. Table 5.2 shows the comparison of the efficiency of the interval forecasting methods.

Table 4.1 The monthly trading values of China Steel stock and their forecasts by the proposed interval forecasting methods for the latter 6 periods from 5/2005 to 10/2005.

Monthly Trading Value of China Steel 5/2005–10/2005 $(c_t, r_t) / (m_t, l_t, u_t)$	Interval Moving Average (\hat{c}_t, \hat{r}_t)	Weighted Moving Average (\hat{c}_t, \hat{r}_t)	Interval ARIMA of (c, r) Average $((2,0,0), (1,0,0))$ (\hat{c}_t, \hat{r}_t)	ARIMA of (m, l, u) $((2,0,0), (0,0,0), (2,0,0))$ $(\hat{m}_t, \hat{l}_t, \hat{u}_t)$
(32.08,1.58)/(31.94,1.71,1.44)	(23.57,1.46)	(24.29,1.79)	(32.60,1.62)	(32.59,1.39,1.54)
(31.50,1.35)/(31.51,1.34,1.36)	(23.52,1.44)	(24.17,1.75)	(32.03,1.51)	(32.06,1.39,1.60)
(30.03,1.35)/(31.02,2.43,4.42)	(23.48,1.44)	(24.11,1.75)	(31.60,1.48)	(31.65,1.39,1.56)
(27.70,1.40)/(28.16,0.94,1.86)	(23.44,1.44)	(24.09,1.75)	(31.23,1.47)	(31.29,1.39,1.56)
(29.38,0.98)/(29.33,1.02,0.93)	(23.42,1.43)	(24.06,1.74)	(30.89,1.47)	(30.95,1.39,1.55)
(27.58,1.88)/(27.17,2.28,1.47)	(23.41,1.43)	(24.06,1.74)	(30.58,1.47)	(30.62,1.39,1.55)

Table 4.2 The comparison of efficiency of the proposed interval forecasting methods for the interval time series of monthly trading values of China Steel stock.

Criteria	Interval Average	Weighted Moving Average	Interval ARIMA of (c, r) $((2,0,0), (1,0,0))$	ARIMA of (m, l, u) $((2,0,0), (0,0,0), (2,0,0))$
MSEP	41.65	33.82	4.48	4.24
MSEL	0.73	0.63	0.70	0.92
MSEI	42.38	34.45	5.18	5.16
MRIE	2.04	1.65	0.57	0.57

From the above analysis result, the forecast results of this interval time series are extremely similar with that of the model (4.3). It shows that the interval time series is nonstationary but steadily increasing or decreasing. So the interval moving average and the weighted interval moving average can only offer the prediction of the interval length. While $ARIMA((p_c, d_c, q_c), (p_r, d_r, q_r))$ of $(c; r)$ and $ARIMA((p_m, d_m, q_m), (p_l, d_l, q_l), (p_u, d_u, q_u))$ of $(m; l, u)$ can perform better forecasting concurrently for the position and the length of intervals. In addition, their mean relative interval errors are fairly small, which shows the overlapping range of the predicted intervals and the actual intervals is larger, so it is concluded that both are excellent forecasting methods. On the other hand, the mean relative interval errors of the interval moving average and the weighted interval moving average are relatively great, which shows that there are few overlapping part of the predicted intervals and the actual intervals, so these two forecasting methods are not suitable for this case.

5. Conclusions

In the progress of the scientific research and analysis, the uncertainty in the statistical numerical data is the crux of the problem that the traditional mathematical model is hard to be established. Manski (1990) has pointed out that the numerical data are over-demanded and over-explained. If we exploit this artificial accuracy to do causal analysis or measurement, it may lead to the deviation of the causal judgment, the misleading of the decision model, or the exaggerated difference between the predicted result and the actual data. This paper proposes to use the interval data to avoid such risks to happen. By means of the interval operations, this paper also discusses the model establishment and forecasting of interval time series. From the research results, it is found that the forecasting of ARIMA interval time series performs more accurate, no matter in the comparisons of the mean squared error of interval or the mean relative interval error, than the traditional forecasting methods such as the moving average and the weighted moving average, etc. Especially for the threshold time series, the forecasting results are much better.

In fact, using interval data to establish model and predict, we can find that the forecasting in each step is carried out by means of intervals, so as to increase the objectiveness of the forecasting results. In the general aspect, the "intervalization" seems to be a very normal phenomenon too. But on the contrary, if the concept of dealing with numerical data does not change and the forecasting method does not make a breakthrough, it often hampers the objectivity of measurements and the possibility of long-term forecasting. If we measure interval time series with the center and the length of intervals, it demonstrates clearly that interval time series has better forecasting ability than the traditional ARIMA method does. However, according to the interval operations and the ARIMA method, it is noteworthy that if we can establish the good model construction process, we can make a superior interval forecasting for the interval time series of stock trading values. For investors, it not only provides a new forecasting method, but also offers more flexible prediction result. Therefore, investors can make more objective judgment under correct information.

Although the approaches proposed in this paper perform interval forecasting effectively, there are some problems still remaining to be solved and some improvement can be done for further research, which are described respectively as follows.

- (1) There are so many unpredictable factors on the monthly trading values of stocks, such as trading volumes, exchange rates, interest rates or even the influence of the government policy etc. Consequently, in respect to interval time series proposed in this paper, we only consider the monthly highest price and the monthly lowest price as the range of the monthly trading price caused by all factors. If it needs to make the result more accurate, it can consider finding out the key factors of influencing the interval range.
- (2) With regard to interval time series, in order to achieve more accurate result, it needs to make the collected data stationary to make further analysis. But, how to judge if an interval time series is stationary? This paper considers the stationariness of interval location and interval length respectively. Is it possible to find another approach to judge the stationariness of interval time series by other interval operations?
- (3) Because the research of interval data forecasting is rare in the past, as generating interval time series data by simulation, we can consider other simulation approaches, such as Bootstrap, Bayesian, Bayesian, etc. The variety of the simulation methods will contribute to the improvement of the forecasting methods.
- (4) In the analysis and forecasting of interval time series, how to estimate the forecasting accuracy of interval data is an important issue. There are found four forecasting situations, the forecast interval is too wide, the forecast interval is too narrow, the forecast interval inclines to the right, and the forecast

interval inclines to the left. From the overlapping parts and the nonoverlapping parts of the actual intervals and the forecast intervals, it should be defined a criterion which is more sufficient to show the efficiency of interval forecasting.

References

- [1] Wu, B. (1995). *Introduction to Time Series Analysis*, HwaTai: Taipei.
- [2] Montgomery, D. C. and Johnson, L. A. (1976), *Forecasting Time Series Analysis*, New York: McGraw-Hill.
- [3] Abraham, B. and Ledolter, J. (1983), *Statistical Methods for Forecasting*. New York: John Wiley.
- [4] Chatfield, C. (1989), *The Analysis of Time Series (4th ed.)*, London: Chapman and Hall.
- [5] Granger, C.W.J., White, H. and Kamstra, M. (1989), Interval Forecasting: An analysis based upon ARCH-quantile estimators, *Journal of Econometrics*, 40, 87–96.
- [6] Chatfield, C. (1993), Calculating interval forecasts, *Journal of Business and Economic Statistics*, 11, 121–135.
- [7] Diebold, F. X. and Mariano, R. S. (1995), Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13, 253–265.
- [8] Christoffersen, P. F. (1998), Evaluating interval forecasts, *International Economic Review*, 39, 841–862.
- [9] Freedman, D. A. and Peter, S. C. (1984), Bootstrapping a Regression Equation: Some Empirical Results, *Journal of the American Statistical Association*, 79, 97–106.
- [10] Thombs, L. A. and Schucany, W. R. (1990), Bootstrap Prediction Intervals for Autoregression, *Journal of the American Statistical Association*, 85, 486–492.
- [11] McCullough, B. D. (1994), Bootstrapping forecast intervals: An application to AR(p) models, *Journal of Forecasting*, 13, 51–66.
- [12] McCullough, B. D. (1996), Consistent forecast intervals when the forecast-period exogenous variables are stochastic, *Journal of Forecasting*, 15, 293–304.
- [13] Winkler, R.L. (1972), A Decision-Theoretic Approach to Interval Estimation, *Journal of the American Statistical Association*, 67, 187–191.
- [14] Thompson, P. A. and Miller, R. B. (1986), Sampling the future: A Bayesian approach to forecasting from univariate time series models, *Journal of Business & Economic Statistics*, 4, 427–436.
- [15] Barnett, G., Kohn, R. and Sheather, S. (1996), Robust estimation of an autoregressive model using Markov chain Monte Carlo, *Journal of Econometrics*, 74, 237–254.
- [16] Chatfield, C. (1996), Model uncertainty and forecast accuracy, *Journal of Forecasting*, 15, 495–508.
- [17] Faraway, J. and Chatfield, C. (1998), Time-series forecasting with neural networks: A comparative study using the airline data, *Applied Statistics*, 47, 231–250.
- [18] Young, R. C. (1931), The algebra of many-valued quantities, *Mathematische Annalen*, 104, 260–290, Also at <http://www.cs.utep.edu/interval-comp/young.pdf>.
- [19] Dwyer, P. S. (1951), *Linear Computations*. New York: John Wiley and Sons.
- [20] Zamowitz, V. and Braun, P. (1993), Twenty-two years of the NBER-ASA Quarterly Economic Outlook surveys. J. H. Stock and M. Watson (eds.), *Business Cycles, Indicators and Forecasting*, University of Chicago Press, Chicago.
- [21] Granger, C.W.J. (1996), Can we Improving the Perceived Quality of Economic Forecasts? *Journal of Applied Econometrics*, 11, 455–473.
- [22] Wu, B. and Chen, M. (1999). Use fuzzy statistical methods in change periods detection. *Applied Mathematics and Computation*. **99**, 241–254.
- [23] Hayes, B. (2003), A Lucid Interval. *American Scientist*, 91, 484–488.
- [24] Akaike, H. (1969), Fitting autoregressive models for prediction, *Annals of the Institute of Statistical Mathematics*, 21, 243–247.
- [25] Manski, C. (1990), The Use of Intention Data to Predict Behavior: A Best Case Analysis. *Journal of the American Statistical Association*, 85, 934–940.
- [26] Box, G. E. P. and Jenkins, G. M. (1976), *Time Series Analysis : Forecasting and control*, San Francisco: Holden-Day.
- [27] Introduction to time series analysis. <http://www.itl.nist.gov>.
- [28] Introduction to interval computation. <http://www.cs.utep.edu/interval-comp/>

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

✓ 達成目標

未達成目標（請說明，以 100 字為限）

實驗失敗

因故實驗中斷

2. 研究成果在學術期刊發表或申請專利等情形：

■ 論文：☐已發表 ☐未發表之文稿 ☐撰寫中 ☐無

專利：☐已獲得 ☐申請中 ☐無

技轉：☐已技轉 ☐洽談中 ☐無

其他：（以 100 字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

The forecasting of ARIMA interval time series performs more accurate, no matter in the comparisons of the mean squared error of interval or the mean relative interval error, than the traditional forecasting methods such as the moving average and the weighted moving average, etc. Especially for the threshold time series, the forecasting results are more appropriate and realistic in applications.

Using interval data to establish model and predict, we can find that the forecasting in each step is carried out by means of intervals, so as to increase the objectiveness of the forecasting results. In the general aspect, the "intervalization" seems to be a very normal phenomenon too. But on the contrary, if the concept of dealing with numerical data does not change and the forecasting method does not make a breakthrough, it often hampers the objectivity of measurements and the possibility of long-term forecasting. If we measure interval time series with the center and the length of intervals, it demonstrates clearly that interval time series has better forecasting ability than the traditional ARIMA method does.

However, according to the interval operations and the ARIMA method, it is noteworthy that if we can establish the good model construction process, we can make a superior interval forecasting for the interval time series of stock trading values. For investors, it not only provides a new forecasting method, but also offers more flexible prediction result. Therefore, investors can make more objective judgment under correct information. Because the research of interval data forecasting is rare in the past, as generating interval time series data by simulation, we can consider other simulation approaches, such as Bootstrap, Bayesian, Bayesian, etc.

行政院國家科學委員會補助國內專家學者出席國際學術會議報告

2009 年 9 月 16 日

報告人姓名	吳柏林	服務機構及職稱	國立政治大學應用數學系教授
時間 會議 地點	2009/8/2~8/7 泰國清邁	本會核定 補助文號	計畫編號 NSC962416H004014-MY3
會議名稱	Symposium on Econometrics and Forecasting		
發表論文題目	Fuzzy Estimation Methods and their Application in Real Estate Evaluation		

一. 參加會議經過

The International Symposium on Econometrics and Forecasting, 於 2009 年 8 月 2 日至 8 月 7 日在泰國清邁舉行, 由泰國 Chainmai University 與 Thailand Econometrics Society 主辦, 來自各國之學者專家約有 40 餘人參加, 包括地主國泰國外, 台灣, Japan, USA, Canada, Hongkong, 澳洲, 馬來西亞, 等 10 餘國. The conference is held for promoting researches on the Econometrics and soft computing. In recent years, researches on logistics, service, supply chain and knowledge management have become prosperous. In addition, new applications of Mobile business derived from e-Commerce are developing fast and widely. New concepts and theories from above have promoted ubiquitous communication and collaboration. Thanks all participants to attend and present recent results and discuss on management engineering from the perspective of knowledge management and e-Commerce.

2 本次大會就以下幾項重點主題進行研討

1. New development in Econometrics 2. Soft computing 3. Intelligent computing 4. knowledge management. 除此之外, 尚有幾個序列的 Workshops 與 Panel discussions 討論有關 computing Economisc 等. 歐國家如泰國等教授在會中特別提到: Intelligent Computing 與 Soft Computing 之工程經濟觀點. Mangement challenges in times of global change and uncertainty. 以及一些新觀念如: Perception based reasoning, 區間計算, 模糊樣本分析, 知識經濟與創新計算等新看法。

二. 攜回的資料:

1. Symposium on Econometrics and Forecasting 研討會論文集 2. 與 Professor Songsak, Professor Sonpond, Professor SAATTakada 等學者討論有關最近著作與研究結果。

三. 建議與其他

目前筆者在政大應數研究所開設人工智慧與時間數列分析與預測課程多年, 深感學術研究發展日新月異、一日千里。國科會能給予補助出國出席國際學術研討會, 收穫相當大。希望將來能多利用課餘時間出國做短期研究, 吸收國外新知、及研究方向。回國後繼續開設模糊時間數列分析與預測課程, 指導博、碩士班研究生, 籌辦國際學術研討會, 推動國際經濟與管理學術研究工作。並於國際著名學術論文期刊, 發表學術論文。

行政院國家科學委員會補助國內專家學者出席國際學術會議報告

2009 年 12 月 16 日

報告人姓名	吳柏林	服務機構及職稱	國立政治大學應用數學系教授
時間 會議 地點	2009/12/12~12/15 中國桂林	本會核定 補助文號	計畫編號 NSC962416H004014-MY3
會議名稱	Intech 2009 Intelligent Technologies		
發表論文題目	Testing hypothesis with fuzzy sample mean and variance		

一. 參加會議經過

The International Symposium on Management Engineering, 於 2009 年 11 月 30 日至 12 月 2 日在日本北九州舉行, 由泰國 Assumption University 主辦, 來自各國之學者專家約有 70 餘人參加, 包括地主國中國外, 韓國 台灣, 美國, 加拿大, 澳洲, 南非, 馬來西亞, 捷克等 10 餘國. The conference is held for promoting researches on the Intelligent Technology. In recent years, researches on logistics, service, supply chain and knowledge management have become prosperous. In addition, new applications of Mobile business derived from e-Commerce are developing fast and widely. New concepts and theories from above have promoted ubiquitous communication and collaboration. Thanks all participants to attend and present recent results and discuss on management engineering from the perspective of knowledge management and e-Commerce.

2 本次大會就以下幾項重點主題進行研討

1. Management and control 2. Soft computing 3. Intelligent computing 4. knowledge management. 除此之外, 尚有幾個序列的 Workshops 與 Panel discussions 討論有關 Management Engineering 等. 東歐國家如捷克等教授在會中特別提到: Intelligent Computing 與 Soft Computing 之管理工程觀點. Management challenges in times of global change and uncertainty. 以及一些新觀念如: Perception based reasoning, 區間計算, 模糊樣本分析, 知識管理與創新計算等新看法。

二. 攜回的資料:

1. International Symposium on Management Engineering 研討會論文集 2. 與 Professor Watada, Professor Takada 等學者討論有關最近著作與研究結果。

三. 建議與其他

目前筆者在政大應數研究所開設人工智慧與時間數列分析與預測課程多年, 深感學術研究發展日新月異、一日千里。國科會能給予補助出國出席國際學術研討會, 收穫相當大。希望將來能多利用課餘時間出國做短期研究, 吸收國外新知、及研究方向。回國後繼續開設模糊時間數列分析與預測課程, 指導博、碩士班研究生, 籌辦國際學術研討會, 推動國際經濟與管理學術研究工作。並於國際著名學術論文期刊, 發表學術論文。

Fuzzy Estimation Methods and their Application in Real Estate Evaluation

Jia-chun Xie

*Faculty of Economics
Chiang Mai University,
Thailand*

Berlin Wu

*Department of Math. Sciences National Chengchi
University, Taiwan*

Songsak [Sriboonchita](#)

*Faculty of Economics
Chiang Mai University,
Thailand*

Abstract

The evaluation of real estate is complex, it involves many factors, such as peoples' utility (human subjective recognition) and the economic efficiency etc. In this paper we develop the traditional real estate appraisal approach the market comparison approach by integrating it with fuzzy estimation which reflects the imprecise and vague nature of human thinking. The primary main methods used are fuzzy statistical analysis, fuzzy-weighting and fuzzy ranking. This integrated procedure is aimed at yielding appropriate and reasonable rank and value of utility. We also give empirical examples to illustrate the techniques and how to evaluate the real estate. Result shows that fuzzy statistics with soft computing are more realistic and reasonable in the realistic evaluation. Finally certain comments are suggested for the further studies.

Keywords: Fuzzy estimation, Real estate valuation, Fuzzy statistics, Fuzzy-weight

6. 1. Introduction

Why do we think we have experienced such acceleration in prices of real estate in Shanghai, China? How are people recognizing the future values of a real estate? Should we depend on the banks to recognize the real value of the property we want to purchase?

Usually, brokers use licensed appraisers to estimate the value of a house. Moreover, the licensed appraisers use comparable sales and if there are not any, they have to improvise and use some comparable cases from similar areas. If the evaluation really gets skewed, what the licensed appraisers can do is use comparable cases from recent sales as close to the likeness of the property as possible. Because of these licensing appraisal practices the next properties sold could be over or under valued. How do they know the true value and how do they evaluate?

There have been numerous literatures in the methodology of real estate valuation. David Mackmin(1995) showed some implications of DCF(Discounted Cash Flow) for the property valuation. Mats Ekelid *et al.* (1997) have focused on the structure and arguments in appraisal reports and how these have changed over time and on the treatment of different types of uncertainty. Christian Janssen (1999) used the regression approach to estimate the market value of a townhouse. Monte Carlo simulations were used to incorporate the

uncertainty of valuation parameters by Martin Hoesli *et al.* (2005). Zhen Chen *et al.* (2009) introduced a novel decision-making approach to risk assessment in commercial real estate development based on the ANP (Analytic Network Process) model. Patrick McAllister *et al.* (2009) proposed an alternative approach that uses pricing signals from traded cash flows to appraise the various incomes and costs. Hasan Sahin *et al.* (2009) used the CART (Classification and Regression Tree) approach to analyze empirically major factors that affect housing price in Istanbul, Turkey.

All of these studies mentioned above are concerned with how to reduce the risk and uncertainty involved in the imprecise and vague information in the process of valuation. Risk and uncertainty are inherent parts of the valuation process as the price given is unable to be inclusive of all current and future influences on the value of the asset (Alastair Adair *et al.* 2009). In this paper, we are proposing an alternative method which uses fuzzy logic to improve the traditional way of valuation by more accurately reflecting the real way of human thinking.

In practice, in the real world, the price provided in the real estate valuation report supplied by the valuation agency is an exact number. However, the buyer and the seller are the different stakeholders in the commercial transaction. The more benefits the buyer obtains the less the seller can earn from his/her house. Different stakeholders have different expected prices and this price will not be the valuation price except in some rare situations. Actually in the real world, when we ask the holders how many the commercial houses worth, usually they would like to show an interval price. In order to work on the data, we need to provide an improved fuzzy-based methodology in the real estate valuation which involves in the attributes of impreciseness and vagueness.

There has been considerable and increasing attention paid to the idea of fuzzy logic since it was introduced by Zadeh as a modification of conventional mathematical Set Theory (Zadeh 1965). The basic aim is that vagueness and ambiguity can be described and distinguished mathematically. The concept of a linguistic variable was introduced as well by Bellman and Zadeh (1970). Fuzzy theory has been widely used a variety of authors. Arvind Verma(1997) used fuzzy logic to construct offender profiles since the police officers received descriptions of suspects that were fuzzy in nature. Nen Fon Tseng *et al.* (2002) applied fuzzy regression models to business cycle analysis. Berlin Wu *et al.* (2002) proposed new approaches on market research with fuzzy statistical analysis. Kostas Metaxiotis(2003) integrated fuzzy logic into a decision support system. Hong Zhang (2004) proposed the fuzzy discrete-event simulation to model the uncertain activity duration. Rajkumer Ohdar *et al.* (2004) introduced the fuzzy-based approach to measure and evaluate the performance of suppliers in the supply chain. Don Jyh-Fu Jeng(2006) applied fuzzy forecasting techniques in DNA computing. Shu Meei Ho(2006) used integrated fuzzy statistics analysis in the valuation of intellectual capital. Sharon M. Ordoobadi(2008) used fuzzy logic to evaluate advanced technologies for decision makers and provided a model based on fuzzy logic for decision makers to help them with selection of appropriate suppliers in 2009. Fuzzy logic was introduced in real estate appraisal by a great many of scholars. Fuzzy logic was introduced to deal with the risk and uncertainty which were poorly considered in real estate analysis by Peter Byrne(1995). Bagnoli and Simth(1998) introduced the theory of fuzzy logic in property appraisal and estimation of appropriate market value. Maurizio d'Amato(2002) applied fuzzy theory to appraise property and made a comparison between this method and the most common statistic instruments. Marco Aure'lio Stumpf Gonza'lez(2006) improved mass appraisal techniques using FRBS(Fuzzy Rule-Based System) to model the real estate market and compared it with traditional hedonic regression model. Elli Pagourtzi *et al.* (2006)

proposed a new methodology and discussed the architecture for a decision support system for real estate analysis based on GIS (Geographic Information System) integrated with fuzzy theory and spatial analysis. Eddie Chi Man Hui *et al.* (2009) integrated the housing indicator data such as RI (Rental Index) with fuzzy logic system to assist practitioners as well as investors on decision making in real estate investment. In this paper we focus on the valuation approach in real estate. We take the city of Kunming in Yunnan, China, as an empirical study. How to get a reasonable and realistic price of real estate is the major requirement for buyers, sellers and valuation practitioners.

7. 2. On Traditional Evaluation Techniques with Real Estate Appraisal

There are many traditional methods to estimate the value of real estate. These include the *Market Comparison Approach*, *Income Approach*, *Cost Approach*, *Hypothetical Development Method*, and so on which can be selected for different appraisal purposes and subject properties.

The Market Comparison Approach compares the subject property with a similar property whose trading date is close to the appraisal date. It then modifies some of influence factors in order to evaluate an objective and reasonable price or value of the real estate. The Income Approach evaluates the future incomes of the subject property and calculates the present value with the proper capitalization rate. It then takes the present value as the evaluated price. The Cost Approach uses the rebuilding price or reset price of the subject property and the deduction of the depreciation to compute objective and reasonable price or value of the subject property. The Hypothetical Development Method is on an assumption that the property is developed and then deducts the expected development costs, taxes and profits from the predicted value of the developed property.

All of these methods are applied widely in the valuation of real estate. One common feature of all of them is that the valuation result is a specific real number. However the trade price of the property is not the appraisal price. The appraisal price can be quite close to the final trade price but will not be the same except in some occasional situations.

Furthermore, from the seller's perspective, he maximizes his profit by raising the price of the property. Similarly the buyer maximizes his utility by reducing the price of the property simultaneously. Obviously, both of the expected prices of seller and buyer are highly unlikely to be the appraisal price which is a specific real number supplied by the appraisal report. There exists a range which includes the expected prices of the seller and buyer.

Therefore, we need to use fuzzy data to evaluate a reasonable range of the appraisal price which can satisfy both the seller and buyer.

Just like other business, real estate has various types of value. There is bank value, tax value, book value and asset value. In most cases this evaluation process is designed not only for financial purposes but also based on utility. Most people would be happy to pay the bank appraisal price because they know the real values of the property have not been considered and the property is worth more.

Note that in this paper we are not talking about residential raw housing real estate. We are talking about homes with acreage, lake and river frontage, elevated views as well as panoramic views. These properties offer something unique and beneficial to the owner or buyer and therefore command a premium price.

Nowadays it is not as time consuming as before because there are several websites available on the net that really help to know which location is suitable for investment in real estate and why. Contact to the property agents of that particular location through mail and send their enquires whatever in your mind or call them and

ask about different questions of property buying, selling and renting of that area. The information may give us enough knowledge about that particular area's real estate investment and we can start to get a good picture of the property business in that region. Before considering any location we need to know some things about that region. Such as: *Population, Tax and ownership law, Rents of property, Buyer and Sellers of that particular region, Future projects and Developments, Temperature and Environment, Traffic convenient, Management cost.*

8. 3. Evaluation Based on the Fuzzy Statistical Analysis

Traditional statistics deals with a single answer or certain range of answer through sampling survey, but it has difficulty in reflecting people's incomplete and uncertain thought. In other words, these processes often ignore the intriguing, complicated and yet sometimes conflicting human logic and feeling. If people can use the membership function to express the degree of their feelings based on their own concept, the result will be closer to their real thought. For instance, when people process a pollution assessment, they classify the distraction into two categories: pollution and non-pollution. This kind of classification is not realistic, since pollution is a fuzzy concept (degree) and can hardly be justified by true-false logic. Therefore, computing the information based on fuzzy logic would be more reasonable. Mode, mean and median are essential statistics in analyzing the sampling survey. Calvo and Mesiar (2001) proposed the generalized median by discussion aggregation operators closely related to medians and to propose new types of aggregation operators appropriate, both for the cardinal and ordinal types of information. But they didn't give a realistic example. In this paper we present the definitions of fuzzy mode, fuzzy mean and related properties. We hope via these new techniques peoples' thought can be extracted and known in a more precise way.

If people can use the membership function to express the degree of their feelings based on their own choices, the result presented will be closer to real human thinking. Therefore, collecting information based on the fuzzy mode should be the first step to take. Since a lot of times, the information itself is embedded with uncertainty and ambiguity, it is natural for us to propose fuzzy statistics, such as fuzzy mode and fuzzy median, to fit the modern requirement. In this and next section we demonstrate the definitions for fuzzy mode and fuzzy mean generalized from traditional statistics.

8.1. 3.1 How to get fuzzy number

Definition 3.1 Fuzzy Number

Let U denote a universal set, $\{A_i\}_{i=1}^n$ be a subset of discussion factors on U , and $\Lambda(A_i)$ be a level set of A_i for $i=1,2,\dots,n$. The fuzzy number of a statement or a term X over U is defined as:

$$\mu_U(X) = \sum_{i=1}^n \mu_i(X) I_{A_i}(X) \quad (3.1)$$

Where $\{\mu_i(X), 0 \leq \mu_i(X) \leq 1\}_{i=1}^n$ are set of membership functions for corresponding factor in $\{A_i\}_{i=1}^n$, and $I_{A_i}(x) = 1$ if $x \in A_i$; $I_{A_i}(x) = 0$ if $x \notin A_i$. If the domain of the universal set is continuous, then the fuzzy number can be written as : $\mu_U(X) = \int_{A_i \subseteq U} \mu_i(X) I_{A_i}(X) \circ$

In the research of social science, the sampling survey is always used to evaluate and understand public opinion on certain issues. The traditional survey forces people to choose one answer from the survey, but it

ignores the uncertainty of human thinking. For instance, when people needs to choose the answer from the survey which lists five choices including "Very satisfactory," "Satisfactory," "Normal," "Unsatisfactory," "Very unsatisfactory," traditional survey become quite exclusive.

The advantages of valuation with fuzzy numbers include: (i) The valuation process becomes robust and consistent by reducing the degree of subjectivity of the evaluator. (ii) Self-potentiality is highlighted by indicating individual distinctions. (iii) It provides the evaluators with an encouraging, stimulating and self-reliant guide that emphasizes individual characteristics. The drawback is that the calculation process will be a little more complex than the conventional one.

Example 3.1 *The use of fuzzy number in a sampling survey about favorite tourist attractions.*
Consider a fuzzy set of favorite tourist attractions for a person as shown in Table 3.1. Note that in the extreme cases when a degree is given as 1 or 0, that is “like” or “dislike”, a standard “yes” and “no” are in a complementary relationship, as in binary logic. Let A_1 represents for “favorite tourist attractions”, A_2 “dislike the tourist attractions”.

Table 3.1 Comparing fuzzy number with integral number favorite tourist attractions

<i>Favorite tourist attractions</i>	A_1	A_2	A_1	A_2
Degree of feelings	$\mu_{A_1}(X)$	$\mu_{A_2}(X)$	binary logic	
Jin Dian	1	0	X	
Xi Shan	0.4	0.6		X
Ye Ya hu	0.7	0.3	X	
Stone Forest	0.8	0.2	X	
National Museum	0.3	0.7		X

Based on the analysis of binary logic, we can find that he likes Jin Dian, Ye Ya hu and Stone Forest but dislikes Xi Shan and National Museum. On the other hand, the fuzzy statistical result can be represented as:

$$\mu_{A_1}(X) = 1I_{JinDian}(X) + 0.4I_{XiShan}(X) + 0.7I_{YeYahu}(X) + 0.8I_{StoneForest}(X) + 0.3I_{NationalMuseum}(X)$$

$$\mu_{A_2}(X) = 0I_{JinDian}(X) + 0.6I_{XiShan}(X) + 0.3I_{YeYahu}(X) + 0.2I_{StoneForest}(X) + 0.7I_{NationalMuseum}(X)$$

This means that the person likes Jin Dian 100%, YeYahu 70%, Stone Forest with 80%. He dislikes XiShan 60%, dislikes National Muesum70%, and dislikes National Museum 70% of degree. Therefore, based on the binary (like or dislike) logic, we can see only the superficial feeling about people’s favorite tourist attractions. With the information of fuzzy response we will see a more detailed data representation.

8.2. 3.2 How to compute fuzzy data

Definition 3.2 Fuzzy Mode (data with interval values)

Let U be the universal set (a discussion domain), $L = \{L_1, L_2, \dots, L_k\}$ a set of k -linguistic variables on U , and $\{FS_i = [a_i, b_i], a_i, b_i \in R, i = 1, 2, \dots, n\}$ be a sequence of random fuzzy sample on U . For each sample FS_i , if there is an interval $[a, b]$ which is covered by certain samples, we call these samples as a cluster. Let MS be the set of clusters which contains the maximum number of sample, then the fuzzy mode FM is defined as

$$FM = [a, b] = \{\cap[a_i, b_i] | [a_i, b_i] \subset MS\}.$$

If $[a, b]$ does not exist (i.e. $[a, b]$ is an empty set), we say this fuzzy sample does not have fuzzy mode.

Example 3.2 Suppose we have the following sample: (2,5), (2,4), (2,3), (3,10), (2,3), (3,4), (3,5) Then by definition, the set of clusters which contains the maximum number of sample is $\{(2,5),(2,4),(3,10),(3,4),(3,5)\}$. Hence FM= (3, 4)

Definition 3.3 Fuzzy Mean (data with interval values)

Let U be the universe set, and $FS_i=[a_i,b_i]$, $i=1, 2, \dots, n$ be a sequence of random fuzzy sample on U . Then the fuzzy mean is defined as $E(X)=[\frac{1}{n}\sum_{i=1}^n a_i, \frac{1}{n}\sum_{i=1}^n b_i]$.

8.3. 3.3 How to compute fuzzy weight

In the valuation process, people usually treat each factor with the equal weight. That is, we assume that the factors have the same contribution to the universe domain. However, in order to get a more accurate valuation, we had better use different weight, according to their contributions to the object, for different factor. Since then, the macro-performance valuation will reflect the real world situation.

To investigate the fuzzy weight of each factor, we may use the fuzzy set theory and sampling survey technique. Especially, using fuzzy memberships and multiple values assignment, we can get an appropriate fuzzy weight for the object. Hence, let's give a detailed process about how to get fuzzy weight.

Here, we propose the calculation process of entity fuzzy weight :

Step1: First, determine the effective factors $A = \{A_1, A_2, \dots, A_k\}$ for the real estate appraisal

Step2: Ask each interviewee i to give the importance of factors set with a membership m_{ij} $\sum_{j=1}^k m_{ij} = 1$. Let m_{ij} be the membership of importance of factor j for the i th interviewee

Step3: Calculate the fuzzy weight w_j of A_j by $w_j = \frac{\sum_{i=1}^n m_{ij}}{\sum_{j=1}^k \sum_{i=1}^n m_{ij}}$.

Example 3.4 Suppose there are five experts rank the influence factors of real estate valuation, see Table 3.4.

Table 3.4 Fuzzy weights of the influence factors of real estate valuation

experts \ factors	location	house type	community	quality
1	0.25	0.25	0.25	0.25
2	0.3	0.25	0.3	0.15
3	0.25	0.3	0.2	0.25
4	0.4	0.1	0.35	0.15
5	0.3	0.1	0.4	0.2
Total rank	1.5	1	1.5	1
weight	0.3	0.2	0.3	0.2

It is easy to compute that $w_1 = \frac{\sum_{i=1}^5 m_{i1}}{\sum_{k=1}^4 \sum_{i=1}^5 m_{ij}} = \frac{1.5}{5} = 0.3, \dots, w_4 = \frac{\sum_{i=1}^5 m_{i4}}{\sum_{k=1}^4 \sum_{i=1}^5 m_{ij}} = \frac{1}{5} = 0.2$.

9. 4. An integrated fuzzy valuation process

Since human thought and behavior are typically vague and uncertain, analysis by traditional methods usually involves the following weaknesses: (a) The use of arithmetic in traditional questionnaires is often an

over-explanation. (b) Experimental data are often overused just to cater to the need for apparent numerical accuracy. (c) For the sake of simplifying the evolutionary model, the relationship between the actual condition and dynamic characteristics are neglected. That is why we will prefer to apply fuzzy theory to handle the questions that involve human opinion.

9.1. 4.1 Determination about price fuzzy distribution

It is appropriate to apply the membership function, a more precise mathematical technique, in analyzing the fuzzy information. The value of the membership function, between 0 and 1, is derived from the characteristic function, to express the membership grade of each element in a set. There are many types of membership functions, such as Z- type, Λ - type, Π - type, S- type, etc, see Nguyen and Wu (2006). In this research we use Λ – type membership functions. It assesses the fuzzy interval of various valuations, and then calculates the fuzzy value of an enterprise according to appraiser's fuzzy weighting.

We also use the Λ - type to reflect the value for a commercial house price distribution. That is, we will give the price of commercial houses into different linguistic terms, such as, high-level, intermediate and unfurnished. Each term will correspond to a real value, which will be determined by the sampling survey and fuzzy statistical analysis.

9.2. 4.2 The highly correlated property for the discussion factors

After detailed discussion from the above sections, an integrated process of fuzzy valuation is started by fixing the crucial affection factors of the commercial house. We use the weighted arithmetic average instead of the geometric average. The reason is that the factors are highly correlated; any extreme value of a certain factor will influence the real price of the commercial house. Take evaluating the price of commodity house as an example.

Suppose the factors of the valuation price are $\{location, house\ type, community, quality\}$. If the factor location is valueless, no matter the *community* is high and/or the other factors are high too, the integrated valuation will be low.

Finally we make weighted arithmetic average to get a more appropriate valuation. That is, suppose the factor sets is $A = \{A_1, A_2, \dots, A_l\}$ and which corresponding to the with a weight set $w = \{w_1, w_2, \dots, w_l\}$, then the integrated Valuation Price will be

$$\text{Valuation Price} = \prod_{i=1}^l A_i^{w_i}$$

9.3. 4.3 The Mean Absolute Error for interval difference

Since an interval X_i can be represented by $X = (c; r)$, where c is the center of the interval and r is the radius of the interval. To consider the error of two intervals, we need to consider the difference of center as well as the difference of radius simultaneously. We propose the following definitions which are useful in determining the interval difference.

Definition 4.1 (Mean Absolute Error of Interval Location, MAEL)

Let $\{\hat{X}_t = (\hat{c}_t, \hat{r}_t); t = 1, 2, \dots, n\}$ be a sequence of predicted interval for $\{X_t = (c_t, r_t); t = 1, 2, \dots, n\}$. The mean absolute error of interval location (MAEL) is given by $MAEL = \frac{1}{n} \sum_{t=1}^n |c_t - \hat{c}_t|$.

Definition 4.2 (Mean Absolute Error of Interval Scale, MAES)

Let $\{\hat{X}_t = (\hat{c}_t, \hat{r}_t); t = 1, 2, \dots, n\}$ be a sequence of predicted interval for $\{X_t = (c_t, r_t); t = 1, 2, \dots, n\}$. The mean absolute error of interval scale (MAES) is given by $MAES = \frac{1}{n} \sum_{t=1}^n |r_t - \hat{r}_t|$.

Consider the interval $X=[4,7]=(5.5;1.5)$, and the forecast intervals $\hat{X}_1=[1,8]=(4.5;3.5)$ and $\hat{X}_2=[6,8]=(7;1)$ obtained by two different forecasting methods. The MAES of \hat{X}_1 , denoted $MAES_1$, is 2. The MAES of \hat{X}_2 , denoted $MAES_2$, is 0.5. Then \hat{X}_2 is good forecast interval than \hat{X}_1 from comparing $MAES_1$ and $MAES_2$. However, since the radius of \hat{X}_1 is larger than that of \hat{X}_2 , the central point of \hat{X}_1 is closer to the central point of X . Since the range of \hat{X}_1 covers the range of the actual interval X is more than the range of \hat{X}_2 does. As a result, we still regard \hat{X}_1 as the better forecast interval. Consequently, while considering the efficiency of the interval forecasting, both the two indexes should approach the real values. Hence, by combining the two factors of the center and the radius of interval, we need to integrate those above definitions.

Definition 4.3 (Mean Absolute Error of Interval, MAEI)

Let $\{\hat{X}_t = (\hat{c}_t, \hat{r}_t); t = 1, 2, \dots, n\}$ be a sequence of predicted interval for $\{X_t = (c_t, r_t); t = 1, 2, \dots, n\}$. The mean squared error of interval (MSEI) is given by

$$MAEI = \frac{1}{n} \sum_{t=1}^n |c_t - \hat{c}_t| + \frac{1}{n} \sum_{t=1}^n |r_t - \hat{r}_t| = MAEL + MAES$$

Note that for some special purposes, we may use the Weighted Mean Absolute Error of interval,

$$WMAEI = w_1 \frac{1}{n} \sum_{t=1}^n |c_t - \hat{c}_t| + w_2 \frac{1}{n} \sum_{t=1}^n |r_t - \hat{r}_t| = MAEL + MAES, w_1 + w_2 = 2. \text{ instead of the } MAEI.$$

Example 4.2 Let $X=[4,7]=(5.5;1.5)$ be the real interval, $\hat{Y}=[1,8]=(4.5;3.5)$ and $\hat{Z}=[6,8]=(7;1)$ be the forecasting intervals obtained by two different forecasting methods. It is easy to calculate

MAEL of $\hat{Y} = 1$, MRES of $\hat{Y} = 2$, MAEI of $\hat{Y} = 3$, and

MREL of $\hat{Z} = 1.5$, MRES of $\hat{Z} = 1.5$, MAEI of $\hat{Z} = 3$

That is \hat{Y} has more efficient forecasting location than \hat{Z} . While \hat{Z} has more efficient forecasting interval scale than \hat{Y} . However they have the same MAEI.

Example 4.3 Let the interval time series be $X_1=[4,6]=(5;1)$, $X_2=[5,8]=(6.5;1.5)$, the predicted intervals are

$\hat{X}_1=[2.8,5.4]=(4.1;1.3)$ and $\hat{X}_2=[3.8,7.8]=(5.8;2)$. Then

$$MAEL = (|5 - 4.1| + |6.5 - 5.8|) / 2 = 0.8.$$

$$MRES = (|1 - 1.3| + |1.5 - 2|) / 2 = 0.4$$

$$MAEI = 0.8 + 0.4 = 1.2$$

10. 5. A Case Study on the Commercial House in Kunming (China)

In this section we present a case which shows how to value the price of real estate in Kunming by fuzzy estimation method. The fuzzy valuation process is as follows:

- (1) Decide the influence factors which are of importance for the valuation price of commercial house.
- (2) Apply the fuzzy ordering method to calculate the weights $\{w_1, w_2, \dots, w_k\}$.
- (3) Collect information of the similar property case and then use the Marketing Comparison Approach to modify the value of location, house type, community and quality of the cases.
- (4) Investigate the transaction price of commercial house which located in different district. In generally, five cases are possible.
- (5) Compute the fuzzy mode and fuzzy mean in order to get the valuation price of our subject property.

As is mentioned before, the most important effective factors of the commercial house valuation price are location, house type, community and quality.

The weights of location, house type, community and quality have been calculated in the section of 3.3.

Table 5.1 is the result of the surveys for weights with respect to the effective factors

Table 5.1 the weights of Commercial house

factors	location	house type	community	quality
w	0.3	0.2	0.3	0.2

We use the Marketing Comparison Approach to choose some of the similar cases which depend on the most important effective factors which we will show later. Then the fuzzy estimation method will be applied to obtain the predicted value of our subject property. In this paper we suppose the appraisal purpose is to get the current market price of our subject property for the trader.

Generally, the effective factors of estimated value of a commercial home include: location, house type, community and quality. Take the commercial house in Panlong District as our subject property, so the value of location, house type, community and quality of the subject property can be standardized as 1. By using the Marketing Comparison Approach, we need to choose other similar cases as the reference object to compare. Here we take the district of Kaiyue Times, ShiGuangJunYuan, PropertyCenter, HeTangYueSe and JiangDongWorld as our comparable market cases, which are quite similar in the effective factors.

Take the district of Kaiyue Times for an example then modify the location effective factor. The standard value of location of our subject property is 1 and the evaluating rule of the location is that the less far from the center of the city the more valuable it is. Since the location of the district of Kaiyue Times is much a little far away from the center of the city than our subject property, so we can modify the location effective factor by 0.9.

Modify the house type. Our subject property in Panlong District is with one living room, two bed rooms, two bath rooms, one study room, one kitchen and two balconies which is a quite normal type whose area is 120.5 square meters. The case in the district of Kaiyue Times is also designed similar as our subject property, but be decorated much more reasonable, easier living and with more bright sun shines. The area is 135 square meters. So we can modify the house type effective factor by 1.3.

Modify the community. The community of our subject property in Panlong District is like the other excellent districts in Kunming. It is covered with green grass and trees in the forest belt. There are some body building

equipments and entertainment facilities for the residents.

The nearest bus station is about one kilometer away and it is convenient to go to the urban centre. There are many restaurants nearby that can supply lots of delicious dishes. However, the community in the case of the district of Kaiyue Times has more advantages. Besides, there is a modern shopping mall surrounding it and a hospital not far away. So we can modify the community effective factor by 1.1.

Modify the quality. The quality of our subject property in Panlong District is a common level and has 7 years history. However, the quality of the case in the district of Kaiyue Times has higher quality since the building only have been finished 2 years. So we can modify the quality effective factor by 1.2.

In order to get a more appropriate price we evaluate the price of our subject property by fuzzy estimation method.

The detailed valuation steps are as follows. The valuation price of commercial house is:

Valuation Price : { price, location, house type, community, quality }

The price initially used to get the valuation price is the transaction price of the cases. In this paper, we investigated the transaction price under the help of one real estate business agency which located in Kunming. We get the transaction price in the first market case of Kaiyue Times district is form 5151 to 6255 yuan per square meters. As is known to all, some of the commercial houses are decorated in a high grade level, some of them are of common level and some even are blank housing. Therefore, three typical cases in the district of Kaiyue Times are selected by the rank of decoration, and the interval is from 5151 to 6522 yuan per square meter.

Let

$$P_{t+1} = P_t \cdot L_t^\alpha \cdot H_t^\beta \cdot C_t^\gamma \cdot Q_t^\lambda$$

Where P_{t+1} = the valuation price at time $t+1$, P_t = the price at time t . L_t^α = location affection, H_t^β = house type affection, C_t^γ = community affection, Q_t^λ = quality affection, $\alpha, \beta, \gamma, \lambda$, stand for the multiplicative weight.

The Table 5.2 showed how we get the fuzzy valuation price of our subject property.

Table 5.2 the Valuation Price of The Commercial House in Pan long District Unit =RMB / m^2

Factors Cases	Price _t	Location	House type	Community	Quality	Price _{t+1}
Kaiyue Times	[5151,6522]	0.9	1.3	1.1	1.2	[5613,7107]
ShiGuangJunYuan	[4103,7338]	1.1	0.9	0.8	0.9	[3785,6771]
PropertyCenter	[6111,7792]	0.8	1.1	0.9	0.8	[5398,6883]
HeTangYueSe	[5250,6535]	0.6	0.8	1.2	1.1	[4637,5773]
JiangDongWorld	[5206,6263]	0.7	0.9	1.1	1.2	[4888,5880]
Fuzzy Mode	[5398,5773]					
Fuzzy Mean	[4864,6483]					

From Table 5.2 we can find that the fuzzy mode and fuzzy mean of the valuation price. Take the first case as an example, from Definition 3.2, we can find that the [5151, 6522] is the transaction price of the market case of Kaiyue Times district. Finally computed with the weight and the modified value of location, house type,

community and quality, we can get the fuzzy mean of the subject property, which is

$$[5151 \cdot 0.9^{0.3} \cdot 1.3^{0.2} \cdot 1.1^{0.3} \cdot 12^{0.2}, 6522 \cdot 0.9^{0.3} \cdot 1.3^{0.2} \cdot 1.1^{0.3} \cdot 12^{0.2}] = [5613, 7107]$$

As is shown in the Table 5.2 we get the fuzzy mode is [5398, 5773] and the fuzzy mean is [4864, 6483]. It is not difficult to find that the interval of fuzzy mode is short than the fuzzy mean. It implies that the probabilities of the valuation price of our subject property are more likely in the fuzzy mode interval. However, the fuzzy mean interval implies the lowest price which the seller can accept and the highest price which the buyer can supply.

According to section 4.3 we can calculate the MAEL, MAES and MAEI. The Table 5.3 shows them.

Table 5.3 the MAEL, MAES and MAEI of the valuation price, Unit =RMB / m²

cases \ factors	Price _{t+1} (c; r)	Fuzzy Mode (5585;187)			Fuzzy Mean (5673;809)		
		MAEL	MAES	MAEI	MAEL	MAES	MAEI
Kaiyue Times	(6360;747)	775	560	1335	687	62	749
ShiGuangJunYuan	(5278;1493)	307	1306	1613	395	684	1079
PropertyCenter	(6140;742)	555	555	1110	467	67	534
HeTangYueSe	(5205;568)	380	381	761	468	241	709
JiangDongWorld	(5384;496)	201	309	510	289	313	602
Result		444	622	1066	461	273	735

It is easy for us seeing the result from table 5.3 that the fuzzy mode has more efficient forecasting location than fuzzy mean, since 444 is smaller than 461. While fuzzy mean has more efficient forecasting interval scale than fuzzy mode, since 273 is smaller than 622.

11. 6. Conclusion

Estimating the value of real estate is a wide-ranging and complex area and its evaluation involves much dispute. The advantage of the fuzzy statistical analyzing techniques proposed in this article lies in the way it handles human thought and recognition, improving on vague measurement. The presented integrated procedure differs from the traditional assessment method, and establishes the membership grade of evaluator's weight to better capture real values. Moreover, suppose we are surveying real estate. No matter how carefully we read the measuring process, we can never be certain of the exact value, but we can answer with more confidence that the appropriate area lies within certain bounds. Though interval analysis and fuzzy set theory are areas of active research in mathematics, numerical analysis and computer science began in the late 1950s and early 1960s. The application to statistical evaluations in real estate is just beginning.

Using fuzzy statistical analysis we can get fuzzy data which can be applied in different areas. The methodology which integrates the traditional valuation approach with fuzzy logic shows us how the appraisals can value a commercial house in the form of a fuzzy interval which satisfies different components of real transactions.

Reference

- Calvo, T., and Mesiar, R. (2001), Generalized Median, *Fuzzy Sets and Systems*, 124, pp.59-61.
- Chen, T., and Wang, M. (1999), Forecasting methods using fuzzy concepts, *Fuzzy Sets and System*, 105, pp.339-352.
- Dubois, D., and Prade, H. (1991), Fuzzy sets in approximate reasoning, *Part 1: Inference with possibility*

- distributions, *Fuzzy Sets and Systems*, 40, pp.143-202.
- Hwang, Z., and Wu, B. (1995), Testing for the stability of Taiwan's short run money demand function: the application of fuzzy time series, *Annual Conference Proceedings of Taiwan Economic Association*.
- Lowen, R. (1990), A fuzzy language interpolation theorem, *Fuzzy Sets and Systems*, 34, pp.33-38.
- Nguyen, H., and Wu, B. (2006), *Fundamentals of Statistics with Fuzzy Data*, Springer-Verlag: Heidelberg.
- OECD., *Innovation(1996), Patent and Technology Strategy*, OECD, Paris,.
- Wu, B., and Hsu, Yu-Yun. (2004), A New approach of bivariate fuzzy time series: with applications to the stock index forecasting, *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*. 11(6), pp.671-690.
- Wu, B., and Hsu, Yu-Yun. (2004), The use of kernel set and sample memberships in the identification of nonlinear time series, *Soft Computing Journal*, 8, pp.207-216.
- Wu, B., and Sun, C. (2001), Interval-valued statistics, fuzzy logic, and their use in computational semantics, *Journal of Intelligent and Fuzzy Systems*., 11,1-7.
- Wu, B., and Tseng, N. (2002), A new approach to fuzzy regression models with application to business cycle analysis, *Fuzzy Sets and System*. 130, pp.33-42.
- Zadeh, L. A. (1965), Fuzzy Sets, *Information and Control*, Vol.8, pp.338-353.
- Elli Pagourtzi and Vassilis Assimakopoulos(2003), Real estate appraisal: a review of valuation methods, *Journal of Property Investment & Finance*, Vol.21 No.4,pp.383-401.
- Alastair Adair(2005),The reporting of risk in real estate appraisal property risk scoring, *Journal of Property Investment & Finance*, Vol.23 No.3,pp.254-268.
- Mats Ekelid, Hans Lind, Stellan Lundström and Erik Persson(1998), Treatment of uncertainty in appraisals of commercial properties, *Journal of Property Investment & Finance*, Vol.16 No.4,pp.386-396.
- Zhen Chen and Sukulpat Khumpaisai(2009), An analytic network process for risks assessment in commercial real estate development, *Journal of Property Investment & Finance*, Vol.27 No.3,pp.238-258.
- Patrick McAllister and Pavlos Loizou(2009), The appraisal of data centres: deconstructing the cash flow, *Journal of Property Investment & Finance*, Vol.27 No.1,pp.65-80.
- Martin Hoesli, Elion Jani and Andre' Bender(2006), Monte Carlo simulations for real estate valuation, *Journal of Property Investment & Finance*, Vol.24No.2,pp.102-122.
- Christian Janssen and Zan Yang(1999), Estimating the market value of a proposed townhouse development, *Journal of Property Investment & Finance*, Vol.17No.5,pp.501-516.
- Nick French (2004), The valuation of specialized property: A review of valuation methods, *Journal of Property Investment & Finance*, Vol.22No.6,pp.533-541.
- David Mackmin (1995), DCF discounted: further implications for the valuation surveyor arising from the over-rented property debate, *Journal of Property Valuation & Investment*, Vol.13No.2,pp.5-15.
- Onur O'zsoy and Hasan Sahin (2009), Housing price determinants in Istanbul, Turkey: An application of the classification and regression tree model, *International Journal of Housing Markets and Analysis*, Vol.2No.2,pp.167-178.
- Elli Pagourtzi, Konstantinos Nikolopoulos and Vassilios Assimakopoulos(2006), Housing price determinants in Istanbul, Turkey: An application of the classification and regression tree model, *Journal of Property Investment & Finance*, Vol.24No.1,pp. 68-78.
- Eddie Chi Man Hui, Otto Muk Fai Lau and Tony Kak Keung Lo(2009), Deciphering real estate investment

- decisions through fuzzy logic systems, *Property Management*, Vol.27No.3,pp. 164-177.
- Hong Yao, Li-Yin Shen, Jianli Hao and Chi-ho Michael Yam (2007), A fuzzy-analysis-based method for measuring contractors' environmental performance, *Management of Environmental Quality: An International Journal*, Vol.18No.4,pp. 442-458.
- Peter Byrne (1995), Fuzzy analysis A vague way of dealing with uncertainty in real estate analysis? , *Journal of Property Valuation &Investment*, Vol.13No.3,pp. 22-41.
- Marco Aure'lio Stumpf Gonza'lez UNISINOS, Sa' o Leopoldo, Brazil, and Carlos Torres Formoso(2006), Mass appraisal with genetic fuzzy rule-based systems , *Property Management*, Vol.24No.1,pp. 20-30.
- Maurizio d'Amato(2002), Appraising property with rough set theory , *Journal of Property Investment & Finance*, Vol.20No.4,pp. 406-418.
- Hong Zhang, Heng Li and C.M. Tam (2004), Fuzzy discrete-event simulation for modeling uncertain activity duration, *Engineering, Construction and Architectural Management*, Vol.11No.6,pp. 426-437.
- Kostas Metaxiotis, John Psarras and Emanuel Samouilidis (2003), Integrating fuzzy logic into decision support system: current research and future prospects, *Information Management & Computer Security*, Vol.11No.2, pp. 53-59.
- Sharson M. Ordoobadi (2009), Development of a supplier selection model using fuzzy logic, *Supply Chain Management: An International Journal*, Vol.14No.4, pp. 314-327.
- Arvind Verma (1997), Construction of offender profiles using fuzzy logic, *Policing: An International Journal of Police Strategies & Management*, Vol.22No.2, pp. 408-418.
- Rajkumar Ohdar and Pradip Kumar Ray(2004), Performance measurement and evaluation of suppliers in supply chain: an evolutionary fuzzy-based approach, *Journal of Manufacturing Technology Management*, Vol.15No.8, pp. 723-734.
- Sharon M. Ordoobadi (2008), Fuzzy logic and evaluation of advanced technologies, *Industrial Management & Data Systems*, Vol.108No.7, pp. 928-946.

行政院國家科學委員會補助國內專家學者出席國際學術會議報告

2009 年 9 月 16 日

報告人姓名	吳柏林	服務機構及職稱	國立政治大學應用數學系教授
時間 會議 地點	2009/8/2~8/7 泰國清邁	本會核定 補助文號	計畫編號 NSC962416H004014-MY3
會議名稱	Symposium on Econometrics and Forecasting		
發表論文題目	Fuzzy Estimation Methods and their Application in Real Estate Evaluation		

一. 參加會議經過

The International Symposium on Econometrics and Forecasting, 於 2009 年 8 月 2 日至 8 月 7 日在泰國清邁舉行, 由泰國 Chainmai University 與 Thailand Econometrics Society 主辦. 來自各國之學者專家約有 40 餘人參加, 包括地主國泰國外, 台灣, Japan, USA, Canada, Hongkong, 澳洲, 馬來西亞, 等 10 餘國. The conference is held for promoting researches on the Econometrics and soft computing. In recent years, researches on logistics, service, supply chain and knowledge management have become prosperous. In addition, new applications of Mobile business derived from e-Commerce are developing fast and widely. New concepts and theories from above have promoted ubiquitous communication and collaboration. Thanks all participants to attend and present recent results and discuss on management engineering from the perspective of knowledge management and e-Commerce.

2 本次大會就以下幾項重點主題進行研討

1. New development in Econometrics 2. Soft computing 3. Intelligent computing 4. knowledge management. 除此之外, 尚有幾個序列的 Workshops 與 Panel discussions 討論有關 computing Economisc 等. 歐國家如泰國等教授在會中特別提到: Intelligent Computing 與 Soft Computing 之工程經濟觀點. Mangement challenges in times of global change and uncertainty. 以及一些新觀念如: Perception based reasoning, 區間計算, 模糊樣本分析, 知識經濟與創新計算等新看法。

二. 攜回的資料：

1. Symposium on Econometrics and Forecasting 研討會論文集 2. 與 Professor Songsak, Professor Sonpond, Professor SAATTakada 等學者討論有關最近著作與研究結果。

三. 建議與其他

目前筆者在政大應數研究所開設人工智慧與時間數列分析與預測課程多年, 深感學術研究發展日新月異、一日千里。國科會能給予補助出國出席國際學術研討會, 收穫相當大。希望將來能多利用課餘時間出國做短期研究, 吸收國外新知、及研究方向。回國後繼續開設模糊時間數列分析與預測課程, 指導博、碩士班研究生, 籌辦國際學術研討會, 推動國際經濟與管理學術研究工作。並於國際著名學術論文期刊, 發表學術論文。

行政院國家科學委員會補助國內專家學者出席國際學術會議報告

2009 年 12 月 16 日

報告人姓名	吳柏林	服務機構及職稱	國立政治大學應用數學系教授
時間 會議 地點	2009/12/12~12/15 中國桂林	本會核定 補助文號	計畫編號 NSC962416H004014-MY3
會議名稱	Intech 2009 Intelligent Technologies		
發表論文題目	Testing hypothesis with fuzzy sample mean and variance		

一. 參加會議經過

The International Symposium on Management Engineering, 於 2009 年 11 月 30 日至 12 月 2 日在日本北九州舉行, 由泰國 Assumption University 主辦, 來自各國之學者專家約有 70 餘人參加, 包括地主國中國外, 韓國 台灣, 美國, 加拿大, 澳洲, 南非, 馬來西亞, 捷克等 10 餘國. The conference is held for promoting researches on the Intelligent Technology. In recent years, researches on logistics, service, supply chain and knowledge management have become prosperous. In addition, new applications of Mobile business derived from e-Commerce are developing fast and widely. New concepts and theories from above have promoted ubiquitous communication and collaboration. Thanks all participants to attend and present recent results and discuss on management engineering from the perspective of knowledge management and e-Commerce.

2 本次大會就以下幾項重點主題進行研討

1. Management and control 2. Soft computing 3. Intelligent computing 4. knowledge management. 除此之外, 尚有幾個序列的 Workshops 與 Panel discussions 討論有關 Management Engineering 等. 東歐國家如捷克等教授在會中特別提到: Intelligent Computing 與 Soft Computing 之管理工程觀點. Management challenges in times of global change and uncertainty. 以及一些新觀念如: Perception based reasoning, 區間計算, 模糊樣本分析, 知識管理與創新計算等新看法。

二. 攜回的資料：

1. International Symposium on Management Engineering 研討會論文集 2. 與 Professor Watada, Professor Takada 等學者討論有關最近著作與研究結果。

三. 建議與其他

目前筆者在政大應數研究所開設人工智慧與時間數列分析與預測課程多年, 深感學術研究發展日新月異、一日千里。國科會能給予補助出國出席國際學術研討會, 收穫相當大。希望將來能多利用課餘時間出國做短期研究, 吸收國外新知、及研究方向。回國後繼續開設模糊時間數列分析與預測課程, 指導博、碩士班研究生, 籌辦國際學術研討會, 推動國際經濟與管理學術研究工作。並於國際著名學術論文期刊, 發表學術論文。

Fuzzy Estimation Methods and their Application in Real Estate Evaluation

Jia-chun Xie

*Faculty of Economics
Chiang Mai University,
Thailand*

Berlin Wu

*Department of Math. Sciences National
Chengchi University, Taiwan*

Songsak [Sriboonchita](#)

*Faculty of Economics
Chiang Mai University,
Thailand*

Abstract

The evaluation of real estate is complex, it involves many factors, such as peoples' utility (human subjective recognition) and the economic efficiency etc. In this paper we develop the traditional real estate appraisal approach the market comparison approach by integrating it with fuzzy estimation which reflects the imprecise and vague nature of human thinking. The primary main methods used are fuzzy statistical analysis, fuzzy-weighting and fuzzy ranking. This integrated procedure is aimed at yielding appropriate and reasonable rank and value of utility. We also give empirical examples to illustrate the techniques and how to evaluate the real estate. Result shows that fuzzy statistics with soft computing are more realistic and reasonable in the realistic evaluation. Finally certain comments are suggested for the further studies.

Keywords: Fuzzy estimation, Real estate valuation, Fuzzy statistics, Fuzzy-weight

1. Introduction

Why do we think we have experienced such acceleration in prices of real estate in Shanghai, China? How are people recognizing the future values of a real estate? Should we depend on the banks to recognize the real value of the property we want to purchase?

Usually, brokers use licensed appraisers to estimate the value of a house. Moreover, the licensed appraisers use comparable sales and if there are not any, they have to improvise and use some comparable cases from similar areas. If the evaluation really gets skewed, what the licensed appraisers can do is use comparable cases from recent sales as close to the likeness of the property as possible. Because of these licensing appraisal practices the next properties sold could be over or under valued. How do they know the true value and how do they evaluate?

There have been numerous literatures in the methodology of real estate valuation. David Mackmin(1995) showed some implications of DCF(Discounted Cash Flow) for the property valuation. Mats Ekelid *et al.* (1997) have focused on the structure and arguments in appraisal reports and how these have changed over time and on the treatment of different types of uncertainty. Christian Janssen (1999) used the regression approach to estimate the market value of a townhouse. Monte Carlo simulations were used to incorporate the uncertainty of valuation

parameters by Martin Hoesli *et al.* (2005). Zhen Chen *et al.* (2009) introduced a novel decision-making approach to risk assessment in commercial real estate development based on the ANP (Analytic Network Process) model. Patrick McAllister *et al.* (2009) proposed an alternative approach that uses pricing signals from traded cash flows to appraise the various incomes and costs. Hasan Sahin *et al.* (2009) used the CART (Classification and Regression Tree) approach to analyze empirically major factors that affect housing price in Istanbul, Turkey.

All of these studies mentioned above are concerned with how to reduce the risk and uncertainty involved in the imprecise and vague information in the process of valuation. Risk and uncertainty are inherent parts of the valuation process as the price given is unable to be inclusive of all current and future influences on the value of the asset (Alastair Adair *et al.* 2009). In this paper, we are proposing an alternative method which uses fuzzy logic to improve the traditional way of valuation by more accurately reflecting the real way of human thinking.

In practice, in the real world, the price provided in the real estate valuation report supplied by the valuation agency is an exact number. However, the buyer and the seller are the different stakeholders in the commercial transaction. The more benefits the buyer obtains the less the seller can earn from his/her house. Different stakeholders have different expected prices and this price will not be the valuation price except in some rare situations. Actually in the real world, when we ask the holders how many the commercial houses worth, usually they would like to show an interval price. In order to work on the data, we need to provide an improved fuzzy-based methodology in the real estate valuation which involves in the attributes of impreciseness and vagueness.

There has been considerable and increasing attention paid to the idea of fuzzy logic since it was introduced by Zadeh as a modification of conventional mathematical Set Theory (Zadeh 1965). The basic aim is that vagueness and ambiguity can be described and distinguished mathematically. The concept of a linguistic variable was introduced as well by Bellman and Zadeh (1970). Fuzzy theory has been widely used a variety of authors. Arvind Verma(1997) used fuzzy logic to construct offender profiles since the police officers received descriptions of suspects that were fuzzy in nature. Nen Fon Tseng *et al.* (2002) applied fuzzy regression models to business cycle analysis. Berlin Wu *et al.* (2002) proposed new approaches on market research with fuzzy statistical analysis. Kostas Metaxiotis(2003) integrated fuzzy logic into a decision support system. Hong Zhang (2004) proposed the fuzzy discrete-event simulation to model the uncertain activity duration. Rajkumer Ohdar *et al.* (2004) introduced the fuzzy-based approach to measure and evaluate the performance of suppliers in the supply chain. Don Jyh-Fu Jeng(2006) applied fuzzy forecasting techniques in DNA computing. Shu Meei Ho(2006) used integrated fuzzy statistics analysis in the valuation of intellectual capital. Sharon M. Ordoobadi(2008) used fuzzy logic to evaluate advanced technologies for decision makers and provided a model based on fuzzy logic for decision makers to help them with selection of appropriate suppliers in 2009.

Fuzzy logic was introduced in real estate appraisal by a great many of scholars. Fuzzy logic was introduced to deal with the risk and uncertainty which were poorly considered in real estate analysis by Peter Byrne(1995). Bagnoli and Simth(1998) introduced the theory of fuzzy logic in property appraisal and estimation of appropriate market value. Maurizio d'Amato(2002) applied

fuzzy theory to appraise property and made a comparison between this method and the most common statistic instruments. Marco Aure'lio Stumpf Gonza'lez(2006) improved mass appraisal techniques using FRBS(Fuzzy Rule-Based System) to model the real estate market and compared it with traditional hedonic regression model. Elli Pagourtzi *et al.* (2006) proposed a new methodology and discussed the architecture for a decision support system for real estate analysis based on GIS (Geographic Information System) integrated with fuzzy theory and spatial analysis. Eddie Chi Man Hui *et al.* (2009) integrated the housing indicator data such as RI (Rental Index) with fuzzy logic system to assist practitioners as well as investors on decision making in real estate investment.

In this paper we focus on the valuation approach in real estate. We take the city of Kunming in Yunnan, China, as an empirical study. How to get a reasonable and realistic price of real estate is the major requirement for buyers, sellers and valuation practitioners.

2. On Traditional Evaluation Techniques with Real Estate Appraisal

There are many traditional methods to estimate the value of real estate. These include the *Market Comparison Approach*, *Income Approach*, *Cost Approach*, *Hypothetical Development Method*, and so on which can be selected for different appraisal purposes and subject properties.

The Market Comparison Approach compares the subject property with a similar property whose trading date is close to the appraisal date. It then modifies some of influence factors in order to evaluate an objective and reasonable price or value of the real estate. The Income Approach evaluates the future incomes of the subject property and calculates the present value with the proper capitalization rate. It then takes the present value as the evaluated price. The Cost Approach uses the rebuilding price or reset price of the subject property and the deduction of the depreciation to compute objective and reasonable price or value of the subject property. The Hypothetical Development Method is on an assumption that the property is developed and then deducts the expected development costs, taxes and profits from the predicted value of the developed property.

All of these methods are applied widely in the valuation of real estate. One common feature of all of them is that the valuation result is a specific real number. However the trade price of the property is not the appraisal price. The appraisal price can be quite close to the final trade price but will not be the same except in some occasional situations.

Furthermore, from the seller's perspective, he maximizes his profit by raising the price of the property. Similarly the buyer maximizes his utility by reducing the price of the property simultaneously. Obviously, both of the expected prices of seller and buyer are highly unlikely to be the appraisal price which is a specific real number supplied by the appraisal report. There exists a range which includes the expected prices of the seller and buyer.

Therefore, we need to use fuzzy data to evaluate a reasonable range of the appraisal price which can satisfy both the seller and buyer.

Just like other business, real estate has various types of value. There is bank value, tax value, book value and asset value. In most cases this evaluation process is designed not only for

financial purposes but also based on utility. Most people would be happy to pay the bank appraisal price because they know the real values of the property have not been considered and the property is worth more.

Note that in this paper we are not talking about residential raw housing real estate. We are talking about homes with acreage, lake and river frontage, elevated views as well as panoramic views. These properties offer something unique and beneficial to the owner or buyer and therefore command a premium price. Nowadays it is not as time consuming as before because there are several websites available on the net that really help to know which location is suitable for investment in real estate and why. Contact to the property agents of that particular location through mail and send their enquires whatever in your mind or call them and ask about different questions of property buying, selling and renting of that area. The information may give us enough knowledge about that particular area's real estate investment and we can start to get a good picture of the property business in that region. Before considering any location we need to know some things about that region. Such as: *Population, Tax and ownership law, Rents of property, Buyer and Sellers of that particular region, Future projects and Developments, Temperature and Environment, Traffic convenient, Management cost.*

3. Evaluation Based on the Fuzzy Statistical Analysis

Traditional statistics deals with a single answer or certain range of answer through sampling survey, but it has difficulty in reflecting people's incomplete and uncertain thought. In other words, these processes often ignore the intriguing, complicated and yet sometimes conflicting human logic and feeling. If people can use the membership function to express the degree of their feelings based on their own concept, the result will be closer to their real thought. For instance, when people process a pollution assessment, they classify the distraction into two categories: pollution and non-pollution. This kind of classification is not realistic, since pollution is a fuzzy concept (degree) and can hardly be justified by true-false logic. Therefore, computing the information based on fuzzy logic would be more reasonable. Mode, mean and median are essential statistics in analyzing the sampling survey. Calvo and Mesiar (2001) proposed the generalized median by discussion aggregation operators closely related to medians and to propose new types of aggregation operators appropriate, both for the cardinal and ordinal types of information. But they didn't give a realistic example. In this paper we present the definitions of fuzzy mode, fuzzy mean and related properties. We hope via these new techniques peoples' thought can be extracted and known in a more precise way. If people can use the membership function to express the degree of their feelings based on their own choices, the result presented will be closer to real human thinking. Therefore, collecting information based on the fuzzy mode should be the first step to take. Since a lot of times, the information itself is embedded with uncertainty and ambiguity, it is natural for us to propose fuzzy statistics, such as fuzzy mode and fuzzy median, to fit the modern requirement. In this and next section we demonstrate the definitions for fuzzy mode and fuzzy mean generalized from traditional statistics.

3.1 How to get fuzzy number

Definition 3.1 Fuzzy Number

Let U denote a universal set, $\{A_i\}_{i=1}^n$ be a subset of discussion factors on U , and $\lambda(A_i)$ be a level set of A_i for $i=1,2,\dots,n$. The fuzzy number of a statement or a term X over U is defined as:

$$\mu_U(X) = \sum_{i=1}^n \mu_i(X) I_{A_i}(X) \quad (3.1)$$

Where $\{\mu_i(X), 0 \leq \mu_i(X) \leq 1\}_{i=1}^n$ are set of membership functions for corresponding factor in $\{A_i\}_{i=1}^n$, and. $I_{A_i}(x) = 1$ if $x \in A_i$; $I_{A_i}(x) = 0$ if $x \notin A_i$. If the domain of the universal set is continuous, then the fuzzy number can be written as : $\mu_U(X) = \int_{A_i \subseteq U} \mu_i(X) I_{A_i}(X) \circ$

In the research of social science, the sampling survey is always used to evaluate and understand public opinion on certain issues. The traditional survey forces people to choose one answer from the survey, but it ignores the uncertainty of human thinking. For instance, when people needs to choose the answer from the survey which lists five choices including "Very satisfactory," "Satisfactory," "Normal," "Unsatisfactory," "Very unsatisfactory," traditional survey become quite exclusive.

The advantages of valuation with fuzzy numbers include: (i) The valuation process becomes robust and consistent by reducing the degree of subjectivity of the evaluator. (ii) Self-potentiality is highlighted by indicating individual distinctions. (iii) It provides the evaluators with an encouraging, stimulating and self-reliant guide that emphasizes individual characteristics. The drawback is that the calculation process will be a little more complex than the conventional one.

Example 3.1 The use of fuzzy number in a sampling survey about favorite tourist attractions.

Consider a fuzzy set of favorite tourist attractions for a person as shown in Table 3.1. Note that in the extreme cases when a degree is given as 1 or 0, that is "like" or "dislike", a standard "yes" and "no" are in a complementary relationship, as in binary logic. Let A_1 represents for "favorite tourist attractions", A_2 "dislike the tourist attractions".

Table 3.1 Comparing fuzzy number with integral number favorite tourist attractions

Favorite tourist attractions	A_1	A_2	A_1	A_2
Degree of feelings	$\mu_{A_1}(X)$	$\mu_{A_2}(X)$	binary logic	
Jin Dian	1	0	X	
Xi Shan	0.4	0.6		X
Ye Ya hu	0.7	0.3	X	
Stone Forest	0.8	0.2	X	
National Museum	0.3	0.7		X

Based on the analysis of binary logic, we can find that he likes Jin Dian, Ye Ya hu and Stone Forest but dislikes Xi Shan and National Museum. On the other hand, the fuzzy statistical result can be represented as:

$$\mu_{A_1}(X) = 1I_{JinDian}(X) + 0.4I_{XiShan}(X) + 0.7I_{YeYahu}(X) + 0.8I_{StoneForest}(X) + 0.3I_{NationalMuseum}(X)$$

$$\mu_{A2}(X) = 0I_{JinDian}(X) + 0.6I_{XiShan}(X) + 0.3I_{YeYahu}(X) + 0.2I_{StoneForest}(X) + 0.7I_{NationalMuseum}(X)$$

This means that the person likes Jin Dian 100%, YeYahu 70%, Stone Forest with 80%. He dislikes XiShan 60%, dislikes National Muesum70%, and dislikes National Museum 70% of degree.

Therefore, based on the binary (like or dislike) logic, we can see only the superficial feeling about people's favorite tourist attractions. With the information of fuzzy response we will see a more detailed data representation.

3.2 How to compute fuzzy data

Definition 3.2 Fuzzy Mode (data with interval values)

Let U be the universal set (a discussion domain), $L = \{L_1, L_2, \dots, L_k\}$ a set of k -linguistic variables on U , and $\{FS_i = [a_i, b_i], a_i, b_i \in R, i = 1, 2, \dots, n\}$ be a sequence of random fuzzy sample on U . For each sample FS_i , if there is an interval $[a, b]$ which is covered by certain samples, we call these samples as a cluster. Let MS be the set of clusters which contains the maximum number of sample, then the fuzzy mode FM is defined as

$$FM = [a, b] = \{\cap [a_i, b_i] | [a_i, b_i] \subset MS\}.$$

If $[a, b]$ does not exist (i.e. $[a, b]$ is an empty set), we say this fuzzy sample does not have fuzzy mode.

Example 3.2 Suppose we have the following sample: (2,5), (2,4), (2,3), (3,10), (2,3), (3,4), (3,5) Then by definition, the set of clusters which contains the maximum number of sample is $\{(2,5), (2,4), (3,10), (3,4), (3,5)\}$. Hence $FM = (3, 4)$

Definition 3.3 Fuzzy Mean (data with interval values)

Let U be the universe set, and $FS_i = [a_i, b_i], i = 1, 2, \dots, n$ be a sequence of random fuzzy sample on U . Then the fuzzy mean is defined as $E(X) = [\frac{1}{n} \sum_{i=1}^n a_i, \frac{1}{n} \sum_{i=1}^n b_i]$.

3.3 How to compute fuzzy weight

In the valuation process, people usually treat each factor with the equal weight. That is, we assume that the factors have the same contribution to the universe domain. However, in order to get a more accurate valuation, we had better use different weight, according to their contributions to the object, for different factor. Since then, the macro-performance valuation will reflect the real world situation.

To investigate the fuzzy weight of each factor, we may use the fuzzy set theory and sampling survey technique. Especially, using fuzzy memberships and multiple values assignment, we can get an appropriate fuzzy weight for the object. Hence, let's give a detailed process about how to

get fuzzy weight.

Here, we propose the calculation process of entity fuzzy weight :

Step1: First, determine the effective factors $A = \{A_1, A_2, \dots, A_k\}$ for the real estate appraisal

Step2: Ask each interviewee i to give the importance of factors set with a membership m_{ij}

$\sum_{j=1}^k m_{ij} = 1$. Let m_{ij} be the membership of importance of factor j for the i th interviewee

Step3: Calculate the fuzzy weight w_j of A_j by $w_j = \frac{\sum_{i=1}^n m_{ij}}{\sum_{j=1}^k \sum_{i=1}^n m_{ij}}$.

Example 3.4 Suppose there are five experts rank the influence factors of real estate valuation, see Table 3.4.

Table 3.4 Fuzzy weights of the influence factors of real estate valuation

<i>factors</i> <i>experts</i>	<i>location</i>	<i>house type</i>	<i>community</i>	<i>quality</i>
<i>1</i>	<i>0.25</i>	<i>0.25</i>	<i>0.25</i>	<i>0.25</i>
<i>2</i>	<i>0.3</i>	<i>0.25</i>	<i>0.3</i>	<i>0.15</i>
<i>3</i>	<i>0.25</i>	<i>0.3</i>	<i>0.2</i>	<i>0.25</i>
<i>4</i>	<i>0.4</i>	<i>0.1</i>	<i>0.35</i>	<i>0.15</i>
<i>5</i>	<i>0.3</i>	<i>0.1</i>	<i>0.4</i>	<i>0.2</i>
<i>Total rank</i>	<i>1.5</i>	<i>1</i>	<i>1.5</i>	<i>1</i>
<i>weight</i>	<i>0.3</i>	<i>0.2</i>	<i>0.3</i>	<i>0.2</i>

It is easy to compute that $w_1 = \frac{\sum_{i=1}^5 m_{i1}}{\sum_{k=1}^4 \sum_{i=1}^5 m_{ij}} = \frac{1.5}{5} = 0.3, \dots, w_4 = \frac{\sum_{i=1}^5 m_{i4}}{\sum_{k=1}^4 \sum_{i=1}^5 m_{ij}} = \frac{1}{5} = 0.2$.

4. An integrated fuzzy valuation process

Since human thought and behavior are typically vague and uncertain, analysis by traditional methods usually involves the following weaknesses: (a) The use of arithmetic in traditional questionnaires is often an over-explanation. (b) Experimental data are often overused just to cater to the need for apparent numerical accuracy. (c) For the sake of simplifying the evolutionary model, the relationship between the actual condition and dynamic characteristics are neglected. That is why we will prefer to apply fuzzy theory to handle the questions that involve human opinion.

4.1 Determination about price fuzzy distribution

It is appropriate to apply the membership function, a more precise mathematical technique, in analyzing the fuzzy information. The value of the membership function, between 0 and 1, is derived from the characteristic function, to express the membership grade of each element in a set. There are many types of membership functions, such as Z- type, Λ - type, Π - type, S- type, etc, see Nguyen and Wu (2006). In this research we use Λ – type membership functions. It assesses the fuzzy interval of various valuations, and then calculates the fuzzy value of an enterprise according to appraiser's fuzzy weighting.

We also use the Λ - type to reflect the value for a commercial house price distribution. That is, we will give the price of commercial houses into different linguistic terms, such as, high-level, intermediate and unfurnished. Each term will correspond to a real value, which will be determined by the sampling survey and fuzzy statistical analysis.

4.2 The highly correlated property for the discussion factors

After detailed discussion from the above sections, an integrated process of fuzzy valuation is started by fixing the crucial affection factors of the commercial house. We use the weighted arithmetic average instead of the geometric average. The reason is that the factors are highly correlated; any extreme value of a certain factor will influence the real price of the commercial house. Take evaluating the price of commodity house as an example.

Suppose the factors of the valuation price are $\{location, house\ type, community, quality\}$. If the factor location is valueless, no matter the *community* is high and/or the other factors are high too, the integrated valuation will be low.

Finally we make weighted arithmetic average to get a more appropriate valuation. That is, suppose the factor sets is $A = \{A_1, A_2, \dots, A_l\}$ and which corresponding to the with a weight set $w = \{w_1, w_2, \dots, w_l\}$, then the integrated Valuation Price will be

$$\text{Valuation Price} = \prod_{i=1}^l A_i^{w_i}$$

4.3 The Mean Absolute Error for interval difference

Since an interval X_t can be represented by $X = (c; r)$, where c is the center of the interval and r is the radius of the interval. To consider the error of two intervals, we need to consider the difference of center as well as the difference of radius simultaneously. We propose the following definitions which are useful in determining the interval difference.

Definition 4.1 (Mean Absolute Error of Interval Location, MAEL)

Let $\{\hat{X}_t = (\hat{c}_t, \hat{r}_t); t = 1, 2, \dots, n\}$ be a sequence of predicted interval for $\{X_t = (c_t, r_t); t = 1, 2, \dots, n\}$. The mean absolute error of interval location (MAEL) is given by $MAEL = \frac{1}{n} \sum_{t=1}^n |c_t - \hat{c}_t|$.

Definition 4.2 (Mean Absolute Error of Interval Scale, MAES)

Let $\{\hat{X}_t = (\hat{c}_t, \hat{r}_t); t = 1, 2, \dots, n\}$ be a sequence of predicted interval for $\{X_t = (c_t, r_t); t = 1, 2, \dots, n\}$. The mean absolute error of interval scale (MAES) is given by $MAES = \frac{1}{n} \sum_{t=1}^n |r_t - \hat{r}_t|$.

Consider the interval $X = [4, 7] = (5.5; 1.5)$, and the forecast intervals $\hat{X}_1 = [1, 8] = (4.5; 3.5)$ and $\hat{X}_2 = [6, 8] = (7; 1)$ obtained by two different forecasting methods. The MAES of \hat{X}_1 , denoted $MAES_1$, is 2. The MAES of \hat{X}_2 , denoted $MAES_2$, is 0.5. Then \hat{X}_2 is good forecast interval than \hat{X}_1 from comparing $MAES_1$ an $MAES_2$. However, since the radius of \hat{X}_1 is larger than that of \hat{X}_2 , the central point of \hat{X}_1 is closer to the central point of X . Since the range of \hat{X}_1 covers the range of the actual interval X is more than the range of \hat{X}_2 does. As a result, we still regard \hat{X}_1 as the better forecast interval. Consequently, while considering the efficiency of the interval forecasting, both the two indexes should approach the real values. Hence, by combining the two

factors of the center and the radius of interval, we need to integrate those above definitions.

Definition 4.3 (Mean Absolute Error of Interval, MAEI)

Let $\{\hat{X}_t = (\hat{c}_t, \hat{r}_t); t = 1, 2, \dots, n\}$ be a sequence of predicted interval for $\{X_t = (c_t, r_t); t = 1, 2, \dots, n\}$. The mean squared error of interval (MSEI) is given by

$$MAEI = \frac{1}{n} \sum_{t=1}^n |c_t - \hat{c}_t| + \frac{1}{n} \sum_{t=1}^n |r_t - \hat{r}_t| = MAEL + MAES$$

Note that for some special purposes, we may use the Weighted Mean Absolute Error of interval,

$$WMAEI = w_1 \frac{1}{n} \sum_{t=1}^n |c_t - \hat{c}_t| + w_2 \frac{1}{n} \sum_{t=1}^n |r_t - \hat{r}_t| = MAEL + MAES, w_1 + w_2 = 2. \text{ instead of the } MAEI.$$

Example 4.2 Let $X=[4,7]=(5.5;1.5)$ be the real interval, $\hat{Y}=[1,8]=(4.5;3.5)$ and $\hat{Z}=[6,8]=(7;1)$ be the forecasting intervals obtained by two different forecasting methods. It is easy to calculate

$MAEL$ of $\hat{Y} = 1$, $MRES$ of $\hat{Y} = 2$, $MAEI$ of $\hat{Y} = 3$, and

$MREL$ of $\hat{Z} = 1.5$, $MRES$ of $\hat{Z} = 1.5$, $MAEI$ of $\hat{Z} = 3$

That is \hat{Y} has more efficient forecasting location than \hat{Z} . While \hat{Z} has more efficient forecasting interval scale than \hat{Y} . However they have the same $MAEI$.

Example 4.3 Let the interval time series be $X_1=[4,6]=(5;1)$, $X_2=[5,8]=(6.5;1.5)$, the predicted intervals are $\hat{X}_1=[2.8,5.4]=(4.1;1.3)$ and $\hat{X}_2=[3.8,7.8]=(5.8;2)$. Then

$$MAEL = (|5 - 4.1| + |6.5 - 5.8|) / 2 = 0.8.$$

$$MRES = (|1 - 1.3| + |1.5 - 2|) / 2 = 0.4$$

$$MAEI = 0.8 + 0.4 = 1.2$$

5. A Case Study on the Commercial House in Kunming (China)

In this section we present a case which shows how to value the price of real estate in Kunming by fuzzy estimation method. The fuzzy valuation process is as follows:

- (1) Decide the influence factors which are of importance for the valuation price of commercial house.
- (2) Apply the fuzzy ordering method to calculate the weights $\{w_1, w_2, \dots, w_k\}$.
- (3) Collect information of the similar property case and then use the Marketing Comparison Approach to modify the value of location, house type, community and quality of the cases.
- (4) Investigate the transaction price of commercial house which located in different district. In generally, five cases are possible.
- (5) Compute the fuzzy mode and fuzzy mean in order to get the valuation price of our subject property.

As is mentioned before, the most important effective factors of the commercial house valuation price are location, house type, community and quality.

The weights of location, house type, community and quality have been calculated in the section of 3.3.

Table 5.1 is the result of the surveys for weights with respect to the effective factors

Table 5.1 the weights of Commercial house

factors	location	house type	community	quality
w	0.3	0.2	0.3	0.2

We use the Marketing Comparison Approach to choose some of the similar cases which depend on the most important effective factors which we will show later. Then the fuzzy estimation method will be applied to obtain the predicted value of our subject property. In this paper we suppose the appraisal purpose is to get the current market price of our subject property for the trader.

Generally, the effective factors of estimated value of a commercial home include: location, house type, community and quality. Take the commercial house in Panlong District as our subject property, so the value of location, house type, community and quality of the subject property can be standardized as 1. By using the Marketing Comparison Approach, we need to choose other similar cases as the reference object to compare. Here we take the district of Kaiyue Times, ShiGuangJunYuan, PropertyCenter, HeTangYueSe and JiangDongWorld as our comparable market cases, which are quite similar in the effective factors.

Take the district of Kaiyue Times for an example then modify the location effective factor. The standard value of location of our subject property is 1 and the evaluating rule of the location is that the less far from the center of the city the more valuable it is. Since the location of the district of Kaiyue Times is much a little far away from the center of the city than our subject property, so we can modify the location effective factor by 0.9.

Modify the house type. Our subject property in Panlong District is with one living room, two bed rooms, two bath rooms, one study room, one kitchen and two balconies which is a quite normal type whose area is 120.5 square meters. The case in the district of Kaiyue Times is also designed similar as our subject property, but be decorated much more reasonable, easier living and with more bright sun shines. The area is 135 square meters. So we can modify the house type effective factor by 1.3.

Modify the community. The community of our subject property in Panlong District is like the other excellent districts in Kunming. It is covered with green grass and trees in the forest belt. There are some body building equipments and entertainment facilities for the residents. The nearest bus station is about one kilometer away and it is convenient to go to the urban centre. There are many restaurants nearby that can supply lots of delicious dishes. However, the community in the case of the district of Kaiyue Times has more advantages. Besides, there is a modern shopping mall surrounding it and a hospital not far away. So we can modify the community effective factor by 1.1.

Modify the quality. The quality of our subject property in Panlong District is a common level and has 7 years history. However, the quality of the case in the district of Kaiyue Times has higher quality since the building only have been finished 2 years. So we can modify the quality effective

factor by 1.2.

In order to get a more appropriate price we evaluate the price of our subject property by fuzzy estimation method.

The detailed valuation steps are as follows. The valuation price of commercial house is:

Valuation Price : { price, location, house type, community, quality }

The price initially used to get the valuation price is the transaction price of the cases. In this paper, we investigated the transaction price under the help of one real estate business agency which located in Kunming. We get the transaction price in the first market case of Kaiyue Times district is form 5151 to 6255 yuan per square meters. As is known to all, some of the commercial houses are decorated in a high grade level, some of them are of common level and some even are blank housing. Therefore, three typical cases in the district of Kaiyue Times are selected by the rank of decoration, and the interval is from 5151 to 6522 yuan per square meter.

Let

$$P_{t+1} = P_t \cdot L_t^\alpha \cdot H_t^\beta \cdot C_t^\gamma \cdot Q_t^\lambda$$

Where P_{t+1} = the valuation price at time $t+1$, P_t = the price at time t . L_t^α = location affection, H_t^β = house type affection, C_t^γ = community affection, Q_t^λ = quality affection, $\alpha, \beta, \gamma, \lambda$, stand for the multiplicative weight.

The Table 5.2 showed how we get the fuzzy valuation price of our subject property.

Table 5.2 the Valuation Price of The Commercial House in Pan long District Unit =RMB / m^2

Factors Cases	Price _t	Location	House type	Community	Quality	Price _{t+1}
Kaiyue Times	[5151,6522]	0.9	1.3	1.1	1.2	[5613,7107]
ShiGuangJunYuan	[4103,7338]	1.1	0.9	0.8	0.9	[3785,6771]
PropertyCenter	[6111,7792]	0.8	1.1	0.9	0.8	[5398,6883]
HeTangYueSe	[5250,6535]	0.6	0.8	1.2	1.1	[4637,5773]
JiangDongWorld	[5206,6263]	0.7	0.9	1.1	1.2	[4888,5880]
Fuzzy Mode	[5398,5773]					
Fuzzy Mean	[4864,6483]					

From Table 5.2 we can find that the fuzzy mode and fuzzy mean of the valuation price. Take the first case as an example, from Definition 3.2, we can find that the [5151, 6522] is the transaction price of the market case of Kaiyue Times district. Finally computed with the weight and the modified value of location, house type, community and quality, we can get the fuzzy mean of the subject property, which is

$$[5151 \cdot 0.9^{0.3} \cdot 1.3^{0.2} \cdot 1.1^{0.3} \cdot 1.2^{0.2}, 6522 \cdot 0.9^{0.3} \cdot 1.3^{0.2} \cdot 1.1^{0.3} \cdot 1.2^{0.2}] = [5613, 7107]$$

As is shown in the Table 5.2 we get the fuzzy mode is [5398, 5773] and the fuzzy mean is [4864, 6483]. It is not difficult to find that the interval of fuzzy mode is short than the fuzzy mean. It implies that the probabilities of the valuation price of our subject property are more likely in the

fuzzy mode interval. However, the fuzzy mean interval implies the lowest price which the seller can accept and the highest price which the buyer can supply.

According to section 4.3 we can calculate the MAEL, MAES and MAEI. The Table 5.3 shows them.

Table 5.3 the MAEL, MAES and MAEI of the valuation price, Unit =RMB / m²

cases \ factors	Price _{t+1} (c; r)	Fuzzy Mode (5585;187)			Fuzzy Mean (5673;809)		
		MAEL	MAES	MAEI	MAEL	MAES	MAEI
Kaiyue Times	(6360;747)	775	560	1335	687	62	749
ShiGuangJunYuan	(5278;1493)	307	1306	1613	395	684	1079
PropertyCenter	(6140;742)	555	555	1110	467	67	534
HeTangYueSe	(5205;568)	380	381	761	468	241	709
JiangDongWorld	(5384;496)	201	309	510	289	313	602
Result		444	622	1066	461	273	735

It is easy for us seeing the result from table 5.3 that the fuzzy mode has more efficient forecasting location than fuzzy mean, since 444 is smaller than 461. While fuzzy mean has more efficient forecasting interval scale than fuzzy mode, since 273 is smaller than 622.

6. Conclusion

Estimating the value of real estate is a wide-ranging and complex area and its evaluation involves much dispute. The advantage of the fuzzy statistical analyzing techniques proposed in this article lies in the way it handles human thought and recognition, improving on vague measurement. The presented integrated procedure differs from the traditional assessment method, and establishes the membership grade of evaluator's weight to better capture real values. Moreover, suppose we are surveying real estate. No matter how carefully we read the measuring process, we can never be certain of the exact value, but we can answer with more confidence that the appropriate area lies within certain bounds. Though interval analysis and fuzzy set theory are areas of active research in mathematics, numerical analysis and computer science began in the late 1950s and early 1960s. The application to statistical evaluations in real estate is just beginning.

Using fuzzy statistical analysis we can get fuzzy data which can be applied in different areas. The methodology which integrates the traditional valuation approach with fuzzy logic shows us how the appraisals can value a commercial house in the form of a fuzzy interval which satisfies different components of real transactions.

Reference

- Calvo, T., and Mesiar, R. (2001), Generalized Median, *Fuzzy Sets and Systems*, 124, pp.59-61.
- Chen, T., and Wang, M. (1999), Forecasting methods using fuzzy concepts, *Fuzzy Sets and System*, 105, pp.339-352.
- Dubois, D., and Prade, H. (1991), Fuzzy sets in approximate reasoning, *Part 1: Inference with possibility distributions*, *Fuzzy Sets and Systems*, 40, pp.143-202.
- Hwang, Z., and Wu, B. (1995), Testing for the stability of Taiwan's short run money demand function: the application of fuzzy time series, *Annual Conference Proceedings of Taiwan*

- Lowen, R. (1990), A fuzzy language interpolation theorem, *Fuzzy Sets and Systems*, 34, pp.33-38.
- Nguyen, H., and Wu, B. (2006), *Fundamentals of Statistics with Fuzzy Data*, Springer-Verlag: Heidelberg.
- OECD., *Innovation(1996), Patent and Technology Strategy*, OECD, Paris,.
- Wu, B., and Hsu, Yu-Yun. (2004), A New approach of bivariate fuzzy time series: with applications to the stock index forecasting, *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*. 11(6), pp.671-690.
- Wu, B., and Hsu, Yu-Yun. (2004), The use of kernel set and sample memberships in the identification of nonlinear time series, *Soft Computing Journal*, 8, pp.207-216.
- Wu, B., and Sun, C. (2001), Interval-valued statistics, fuzzy logic, and their use in computational semantics, *Journal of Intelligent and Fuzzy Systems*.,11,1-7.
- Wu, B., and Tseng, N. (2002), A new approach to fuzzy regression models with application to business cycle analysis, *Fuzzy Sets and System*. 130, pp.33-42.
- Zadeh, L. A. (1965), Fuzzy Sets, *Information and Control*, Vol.8, pp.338-353.
- Elli Pagourtzi and Vassilis Assimakopoulos(2003), Real estate appraisal: a review of valuation methods, *Journal of Property Investment & Finance*, Vol.21 No.4,pp.383-401.
- Alastair Adair(2005),The reporting of risk in real estate appraisal property risk scoring, *Journal of Property Investment & Finance*, Vol.23 No.3,pp.254-268.
- Mats Ekelid, Hans Lind, Stellan Lundström and Erik Persson(1998), Treatment of uncertainty in appraisals of commercial properties, *Journal of Property Investment & Finance*, Vol.16 No.4,pp.386-396.
- Zhen Chen and Sukulpat Khumpaisal(2009), An analytic network process for risks assessment in commercial real estate development, *Journal of Property Investment & Finance*, Vol.27 No.3,pp.238-258.
- Patrick McAllister and Pavlos Loizou(2009), The appraisal of data centres: deconstructing the cash flow, *Journal of Property Investment & Finance*, Vol.27 No.1,pp.65-80.
- Martin Hoesli, Elion Jani and Andre' Bender(2006), Monte Carlo simulations for real estate valuation, *Journal of Property Investment & Finance*, Vol.24No.2,pp.102-122.
- Christian Janssen and Zan Yang(1999), Estimating the market value of a proposed townhouse development, *Journal of Property Investment & Finance*, Vol.17No.5,pp.501-516.
- Nick French (2004), The valuation of specialized property: A review of valuation methods, *Journal of Property Investment & Finance*, Vol.22No.6,pp.533-541.
- David Mackmin (1995), DCF discounted: further implications for the valuation surveyor arising from the over-rented property debate, *Journal of Property Valuation & Investment*, Vol.13No.2,pp.5-15.
- Onur Özsoy and Hasan Sahin (2009), Housing price determinants in Istanbul, Turkey: An application of the classification and regression tree model, *International Journal of Housing Markets and Analysis*, Vol.2No.2,pp.167-178.
- Elli Pagourtzi, Konstantinos Nikolopoulos andVassilios Assimakopoulos(2006), Housing price determinants in Istanbul, Turkey: An application of the classification and regression tree

- model, *Journal of Property Investment & Finance*, Vol.24No.1,pp. 68-78.
- Eddie Chi Man Hui, Otto Muk Fai Lau and Tony Kak Keung Lo(2009), Deciphering real estate investment decisions through fuzzy logic systems, *Property Management*, Vol.27No.3,pp. 164-177.
- Hong Yao, Li-Yin Shen, Jianli Hao and Chi-ho Michael Yam (2007), A fuzzy-analysis-based method for measuring contractors' environmental performance, *Management of Environmental Quality: An International Journal*, Vol.18No.4,pp. 442-458.
- Peter Byrne (1995), Fuzzy analysis A vague way of dealing with uncertainty in real estate analysis? , *Journal of Property Valuation &Investment*, Vol.13No.3,pp. 22-41.
- Marco Aure'lio Stumpf Gonza'lez UNISINOS, Sa~ o Leopoldo, Brazil, and Carlos Torres Formoso(2006), Mass appraisal with genetic fuzzy rule-based systems , *Property Management*, Vol.24No.1,pp. 20-30.
- Maurizio d'Amato(2002), Appraising property with rough set theory , *Journal of Property Investment & Finance*, Vol.20No.4,pp. 406-418.
- Hong Zhang, Heng Li and C.M. Tam (2004), Fuzzy discrete-event simulation for modeling uncertain activity duration, *Engineering, Construction and Architectural Management*, Vol.11No.6,pp. 426-437.
- Kostas Metaxiotis, John Psarras and Emanuel Samouilidis (2003), Integrating fuzzy logic into decision support system: current research and future prospects, *Information Management & Computer Security*, Vol.11No.2, pp. 53-59.
- Sharson M. Ordoobadi (2009), Development of a supplier selection model using fuzzy logic, *Supply Chain Management: An International Journal*, Vol.14No.4, pp. 314-327.
- Arvind Verma (1997), Construction of offender profiles using fuzzy logic, *Policing: An International Journal of Police Strategies & Management*, Vol.22No.2, pp. 408-418.
- Rajkumar Ohdar and Pradip Kumar Ray(2004), Performance measurement and evaluation of suppliers in supply chain: an evolutionary fuzzy-based approach, *Journal of Manufacturing Technology Management*, Vol.15No.8, pp. 723-734.
- Sharon M. Ordoobadi (2008), Fuzzy logic and evaluation of advanced technologies, *Industrial Management & Data Systems*, Vol.108No.7, pp. 928-946.

無研發成果推廣資料

96 年度專題研究計畫研究成果彙整表

計畫主持人：吳柏林			計畫編號：96-2416-H-004-014-MY3				
計畫名稱：模糊樣本資料之新統計檢定程序與決策							
成果項目			量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
			實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比		
國內	論文著作	期刊論文	1	1	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	3	3	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	2	2	100%	人次	
		博士生	2	2	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	2	2	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	2	2	100%	人次	
		博士生	2	2	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	Invited speeches 30 November 2008, 29 November 2009 Waseda University, japan Invited speech, January 6, 2010 Economics Society, Chiangmai University
--	---

	成果項目	量化	名稱或內容性質簡述
科 教 處 計 畫 加 填 項 目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

☒ 達成目標

☐ 未達成目標（請說明，以 100 字為限）

☐ 實驗失敗

☐ 因故實驗中斷

☐ 其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文：☒ 已發表 ☐ 未發表之文稿 ☐ 撰寫中 ☐ 無

專利：☐ 已獲得 ☐ 申請中 ☒ 無

技轉：☐ 已技轉 ☐ 洽談中 ☒ 無

其他：（以 100 字為限）

Kolmogorov-Smirnov Two Sample Test with Continuous Fuzzy Data. Advances in Soft Computing, Springer Berlin / Heidelberg (2010), 175-186.

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

In this research, we studied the use of the nonparametric test (K-S two-sample test) with small samples of continuous fuzzy data. In order to identify the statistical pivot, we defined a new

function, the weight function, which includes both central point and radius. The weight function can be used to classify all continuous fuzzy data. With this rule, the cumulative distribution function can be found out. Therefore, we could obtain the statistical pivot of K-S test with continuous fuzzy data.