Multidisciplinary Review Journal

Journal of Critical Reviews

ISSN- 2394-5125 Vol 7, Issue 8, 2020

MEAN-VARIANCE MODEL WITH FUZZY RANDOM DATA

Mohammad Haris Haikal Othman[1], Nureize Arbaiy[2], Muhammad Shukri Che Lah[3], Pei Chun Lin[4]

[1,2,3]Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia, BatuPahat, Malaysia
[4]Department of Information Engineering and Computer Science,
Feng Chia University, No. 100, Wenhwa Rd., Taichung, Taiwan

Received: 20.05.2020 *Revised: 17.06.2020* *Accepted: 06.07.2020*

Abstract
This paper recommends discourse feeling acknowledgment from discourse signal dependent on highlights examination and PNN-classifier. The arrangement of acknowledgment incorporates discovery of discourse feelings, extraction and determination of highlights, lastly characterization. These highlights are valued to segregate the greatest number of tests precisely and the PNN classifier dependent on discriminant investigation is utilized to characterize the six distinctive articulations. The reproduced outcomes will be indicated that the channel occupied component extortion with utilized distribution presents much better exactness with less algorithmic unpredictability than other discourse feeling articulation acknowledgment draws near.

Keywords--- Fuzzy Random Data,Data Pre-processing, Mean-Variance, Standard Deviation

DOI: http://dx.doi.org/10.31838/jcr.07.08.272

INTRODUCTION

Processingon developing a prediction model using data is closely related to data quality. Furthermore, the element of uncertainty in the information also increases the risk of error in predictions, which further complicates the decision-making process (Henry et al. ,2017). The uncertainty usually caused bythedecision parameters related to real-world problems. Therefore, it is crucial to focus on how to treat the uncertainties inherent in the data before they are used to form a prediction model (Bakar & Yi, 2016). Study conduct by Wu (Wu et al., 2016) has found that such a higher level of uncertainty, such as policy uncertainty and government intervention in the economy and markets affect the company investment and stock price. Such uncertainty will affect the volatile of the stock market. In Wieczorek's study (Wieczorek, 2018),the data reference ambiguity and inaccuracy need to overcome because the uncertainty should be evident and the imprecision accurate to make the inference process.

An established portfolio optimization or Portfolio selection is an essentialapproach for practitioners and researchers in finance. Researcher such Hasuike(Hasuike & Mehlawat, 2018)andWang et al., (B. Wang et al., 2018)implemented a portfolio selection model in finance where Haisuke focused on a portfolio selection model that reflected current market trends. In contrast, Wang's (2018) study focused more on historical data. The construction of such portfolio selection models requires sufficient historical data in the forecasts of security returns. Data scientists are currently concentrating on contrasting multiple risk situations when choosing a portfolio(Righi and Borenstein, 2018).

The classic model of portfolio selection developed using a mean-variance method (Markowitz, 1952) and the principle of probabilities to address ambiguous circumstances (Shapiro et al., 2009), (Hussain et al., 2016a). Markowitz approach to portfolio selection assists investors to create investments that give investors the best possible risk. The confidence interval approach used to transform single data into interval data, where it defines the minimum and maximum data to extract the fuzzy data from the stock data. Based onImbens & Spady (Imbens & Spady, 2002), the confidence interval will establish the minimum and maximum value at the confidence area boundary based on the likelihood for the entire parameter vector. The confidence interval is used as the expected value to identify the expected return is a valid return.

The idea of random variables typically used when evaluating uncertainties in the real-life model. Previously, the probability theory used to cope with random events and fuzzy theory offered approaches for dealing with ambiguous data. Both of these theories play an essential role in resolving real-world uncertainties.

Furthermore, existing models should handle random events separately from fuzzy data, even if random events correlate with fuzzy data in the real world. Yoshida (Yoshida, 2018) finds that uncertainty represents certain factors such as random factors and fuzzy factors in linguistic information in economics. Yoshida also found that random variables that were blurred were assessed based on those expectations and criteria with optimistic and probable or needed parameters in the stock market. This study also introduces portfolio decision making to evaluate randomness and the variability of random variables.

Although previous studies used fuzzy data to resolve volatility, few have structured methods that process actual data into fuzzy and random data. It is crucial to have a pre-processing method for data containing fuzzy and random features to treat uncertainties to avoid more errorsthat have included in the prediction model built for the problem (Zhong et al., 2010) and (Hussain et al., 2016b).

This research, therefore, proposes a technique based on the fuzzy random variable to overcome the fuzzy random uncertainty inherent in the data for the construction of a model of portfolio selection.

Lin et al. proposed a series definition and proves of weight function, for fuzzy numbers to overcome the uncertianiess (Lin et al., 2011 - 2013), which gives us more information about the original fuzzy data.The pre-processing of the fuzzy random data performs a method requiring the handling of fuzzy data and a random number. The improvement in pre-processingof data proposed in this study to deal with the fuzzy random uncertainty anticipated to help improve prediction model accuracy, i.e. reducing errors in the initial stages of modelling. Motivated by the ambiguity principle(Chuliá et al., 2017) and (Islam et al., 2016), pre-processing data should use a good definition of a random variable to overcome the problem of simultaneous uncertainties.

The rest of the paper structured this sort of thing follows. Section 2 describes prior knowledge of the related research. Section 3 explains the primary method, and by using stock market statistics, Section 4 demonstrates the numerical illustration.The conclusion concluded in Section 5.

LITERATURE REVIEW

This section contains prior information of the portfolio selection model by based on mean-variance and fuzzy random numbers and risk estimator.

Mean-Variance Model

Portfolio selection is a vital process in decision making, especially for investment purposes. Selecting a portfolio includes the process of finding the best portfolio and requires proper methods. Markowitz proposed a method for determining the best portfolio with the highest returns and the lowest possible risks.

The portfolio selection model (Aouni et al., 2018) is defined as follows:

$$\min\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{x_i\sigma_{ij}x_j}\right\}$$
$$\boldsymbol{s.t.}\sum_{i=1}^{n}\boldsymbol{E(r_i)x_i=\rho} \qquad (1)$$
$$\sum_{i=1}^{n}\boldsymbol{x_i=1}$$
$$\boldsymbol{x_i\geq 0\,,\, i=1,\dots,n}$$

where r_i is the random variable for the return, $E(r_i)$ is the expected value for r_i, σ_{ii} is the variance of r_i, σ_{ij} is the covariance of r_i and r_j, x_i is the proportion of capital to be invested in security *i*and ρ is the parameter of expected return.

Fuzzy Set

Definition 1: Let U denote the universal set of the discourse. Then a fuzzy set A on U, is defined in terms of the membership function m_A that assigns to each element of U a real value from the interval [0, 1]. A fuzzy set A in U written as a set of ordered pairs in the form $A=\{(x,m_A(x)): x\in U\}$, where$m_A: U\to[0,1]$. The value m_A (x), known as the degree of membership (or grade) of x in$A$, expresses the degree to which x proves the property attribute of$A$.

Fuzzy number (FN) is a unique form of fuzzy set anof the real number set at R. In fuzzy maths, FNs play an essential role, comparable to the role ordinary numbers play in classical mathematics. The fuzzy number of x written as follows:

$$\boldsymbol{m_A(X)}=\sum_{i=1}^{n}\boldsymbol{m_i(X)I_{A_i}(X)} \qquad (2)$$

where $I_{A_i}(X)=1$ when $x\in A_i$and $I_{A_i}(X)=0$ if $x\notin A_i$.

Definition 2: A fuzzy set A on U with membership function $y=m(x)$ is said to be normal, if there exists x in U, such that $m(x)=1$.

Fuzzy Random Variables

Fuzzy sets (L.A. Zadeh, 1965)contribute to data hybridization where fuzzy data and randomness are mixed. Kwakernaak(Kwakernaak, 1978)suggested the notation of Fuzzy random variables while Madan L Puri& Dan A Ralescu(Madan L Puri, Dan A Ralescu, 1993)implemented a Fuzzy random variable with a new definition of exception to generalize a set-value function integral.

Theorem 1 Central Limit Theorem (Arnold, 1990)
Let $X_1,\dots,X_n$be the random variables with mean m and variance σ^2. Let

$$\boldsymbol{Z_n}=\frac{\boldsymbol{n^{\frac{1}{2}}(\bar{X}_n-m)}}{\boldsymbol{\sigma}}=\frac{\boldsymbol{W_n-nm}}{\boldsymbol{n^{\frac{1}{2}}\sigma}} \qquad (3)$$

Then, $Z_n$converges in distribution to Z as $n\to\infty$. We denote
$Zn\to Z\sim N(0,1)\ as\ n\to\infty$,
where Z distribute according to a standard normal distribution function *N(0,1).*

Theorem 2 T-test (Lin et al., 2013)
For samples $X_1,X_2,\dots,X_n$of a Normal distributed stochastic quantity $X\sim N(m,\sigma^2)$, a statistic for the hypothesis $H_0: m=m_0$is

$$T=\frac{\boldsymbol{(\bar{X}_n-m_0)}}{\frac{\boldsymbol{S_n}}{\sqrt{\boldsymbol{n}}}}$$

and the acceptance region Ā for T is under probability such

$$\bar{A}=\{t\in R\ |t|=\left|\frac{\boldsymbol{(\bar{X}_n-m_0)}}{\frac{\boldsymbol{S_n}}{\sqrt{\boldsymbol{n}}}}\right|\leq t_n-1:1-\frac{\alpha}{2}\} \qquad (4)$$

Based on Equation (4), $\boldsymbol{\bar{X}_n}$and $\boldsymbol{S_n}$are the mean and standard deviation of the data. $\boldsymbol{m_0}$is a specific value, n is the sample sizeand $t_n-1:1-\frac{\alpha}{2}$as the fractile of a probability distribution.

Probability Distribution Function

The Probability Distribution (PDF) is alist of all possible outcomes of a random variable along with their corresponding probability values.The probability distribution method is optimized to describe the uncertain variables in the analysed data. Based on (Vinnem & Røed, 2020)a pdf is a distribution of probabilities of the value of a dependent variable as a function of theimportance of an independent variable (often time). In the context of health risk assessment, the pdf represents the probability that a given health problem will occur at or before a specified time.

A probability density function (PDF) describes a continuous probability distribution in terms of integrals. Based onKarney(Karney, 2016), Normal distribution used to produce deviations that satisfy the ideal condition that the algorithm is equivalent to sampling and rounding a real number from the normal distribution to the nearest representable floating-point number. The formula to construct Normal distribution based on Karneyis as follows:

$$N(x)=\frac{\exp(-x^2/2)}{\sqrt{2\pi}} \qquad (5)$$

On the other hand,Wang & Wu(B. X. Wang & Wu, 2018) found that the Gamma constructed an approximate confidence interval for the quantile of the two-parameter gamma distribution. In other words, that this distribution can use in the fuzzy environment, especially on fuzzy interval data where the interval data is two-parameter. The formula of Gamma distribution based on Wang & Wuis as follows:

$$\mathrm{F(t|\theta L(t))\ =\ 1-\beta/2, F(t|\theta U(t))\ =\ \beta/2} \qquad (6)$$

The Weibull Distribution is unquestionably one of the popular and broadly used in statistical distributions. Based on Mohammadi(Mohammadi et al., 2016), Weibull distribution has its flexibility, simplicity and adaptability, so it is an advantage for researchers to use Weibull distribution as distribution to treat the randomness. Formula based on Mohammadi such;

$$f(v) = \frac{k}{c}\left(\frac{v}{c}\right)^{k-1} exp\left(-\left(\frac{v}{c}\right)^{k}\right) \quad (7)$$

The Logistic distribution based on (Agrawal & Ganesh, 2018) is essential in the analysis of survival data, the study of incomedistribution and modelling of the spread of an innovation. In Agrawal & Ganesh research, they use mean and standard deviation as the parameter of the random variable to the Logistic Distribution Function. This type of PDF used on fuzzy interval data where the formula of the distribution function such;

$$f_x(x) = \frac{e^{-\left(\frac{x-\mu}{s}\right)}}{s\left[1 + e^{-\left(\frac{x-\mu}{s}\right)}\right]^2} - \infty < x < \infty \quad (8)$$

The distribution function has to predicted under any given condition. The distribution of probability using a standard statistical method is challenging to obtain for data which contains uncertainties. So it takes techniques to manage these uncertain data.

Standard deviation as a risk estimator

Standard deviation is the first mathematical term that calculates the variance of the market or the average amount by which data points differ from the mean. The standard deviation is calculated in this paper using the formula based on the function of a probability distribution, and it only involves such distribution as Normal, Gamma, Weibull, and Logistic.The outcome will produce different types of standard deviation by using many forms ofa probability distribution. Standard deviation is used to calculate market volatility, and hence risk, in investing.

The more unpredictable the price action is, and the wider the range, the greater the risk.(Eschenbach & Lewis, 2019) statesthat the significant differences in standard deviations are greater than the minor differences in the estimated values. The standard deviation means that the level of risk used to calculate the expected value was assessed.

Researchers such(Dave et al., 2010) applying the standard deviation as the risk estimator where when increasing the risk(standard deviation), the expected value can be obtained.Standard deviation is used as a risk estimator in this analysis to determine the risk level for the portfolio selected.

After the specific information is required, the portfolio selection problem can be a deal. After the specific information is required, the portfolio selection problem can be a deal. The next section will clarify the pre-processing of fuzzy random data preparation, where such uncertainties are present in the data used for analysis. Then, the pre-processed data use for the analysis of mean-variance.

Fuzzy Random Data Pre-Processing for Mean-Variance Model

This section explains the proposed methodology for constructing a Fuzzy RandomData pre-processing procedure for the Mean-Variance method. This methodology also gives a risk estimation strategy for the built model. The model has two critical phases. The first step is to pre-process the data to solve the uncertainty using random data. The second phase shows the selection method for the portfolio model by using the mean-variance model with pre-processed random data. Thestandard deviation used as estimators of model selection risk. The methodology described in the following subsection.

Fuzzy Random Data Pre-processing

In the first phase, the data should be identified as interval data or single data. When data is in the form of intervals, calculate the midpoint and width to form the fuzzy data format.If the data is in single point form or crisp, then it needs to be fuzzified using methods such as measurement error (Lah et al., 2019). For example, the 5% percentage of measurement error method is used to fuzzify the data.

A larger 5% error is considered in this study to include more data in the predictions.The fuzzy data includes the maximum, and the minimum value depends on the estimation of the measurement error. Let us say that single data is inS_i, and the maximum and minimum value intervals are F_i . Eq. (5) demonstrates the formulation in which the single value data is produced in the fuzzy data.

$$F_i = \left[S_i - \left(S_i * \frac{5}{100}\right), S_i + \left(S_i * \frac{5}{100}\right)\right], \forall \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \dots \boldsymbol{n} \quad (9)$$

Fuzzified datais presented in an interval form of $F_i = [A_i, B_i]$where $A_i$is the minimum value and $B_i$is the maximum value. For fuzzified results, let the central point and width be a $X_i = [c_i, w_i]$. $X_i$is a fuzzy parameter, c is the centre of the fuzzy number, and w is the width of the fuzzy number.Following is the formula for computing the central point and width given.

$$\boldsymbol{c_i} = \frac{\boldsymbol{A_i} + \boldsymbol{B_i}}{\mathbf{2}} \forall \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \dots \boldsymbol{n} \quad (10)$$
$$\boldsymbol{w_i} = \frac{\boldsymbol{B_i} - \boldsymbol{A_i}}{\mathbf{2}} \forall \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \dots \boldsymbol{n}$$

The randomness in data is dealt with using probability distribution method in this study. The functions used for the probability distribution are Normal Distribution, Weibull, Gamma, and Logistic.The values are compare based on p-values from four different distributions. For the next step, a higher distribution of the p value will be chosen.The fuzzy pre-processing random data generated an interval type of$Y_i = [C_i, W_i]$where C_i is central point and W_i is the width where $Y_i$is a fuzzy interval number parameter from the resulting probability distribution.

Fuzzy Random Data Pre-processingfor Mean-Variance Model

Let us assume $Y_i = (C_i, W_i)$ is the set of an interval fuzzy number where $\forall i = 1,2, \dots, n$. Expected value and variance of fuzzy interval data are determined as follows:

Expected value: (11)

$$\boldsymbol{E(A_i)} = (\boldsymbol{E(c_i)}, \boldsymbol{E(w_i)})$$

Variance: (12)

$$\boldsymbol{\sigma^2(A_i)} = (\boldsymbol{\sigma^2(c_i)}, \boldsymbol{\sigma^2(w_i)})$$

where E is the expected value of interval fuzzy number and σ^2as the variance. Portfolio formulation such as follows:

$$\left.\begin{array}{l} \max \sum_{i=1}^{n} \boldsymbol{E(c_i)x_i} \\ \min \sum_{i=1}^{n} \boldsymbol{E(w_i)x_i} \\ \boldsymbol{s.t.} \sum_{i=1}^{n} \boldsymbol{\sigma\ (c_i)x_i} \\ \sum_{i=1}^{n} \boldsymbol{\sigma\ (w_i)x_i} \\ \sum_{i=1}^{n} \boldsymbol{x_i} \leq \mathbf{1} \\ \sum_{i=1}^{n} \boldsymbol{\sigma\ (w_i)x_i} \\ \boldsymbol{x_i} \geq \mathbf{0} \quad \forall \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \dots, \boldsymbol{n} \end{array}\right\} \quad (13)$$

In (9) $max \sum_{i=1}^{n} E(c_i)x_i$ and $min \sum_{i=1}^{n} E(w_i)x_i$ uses expected value in centre and width.

The phase-in this proposed method may be simplified as follows, based on 3.1, 3.2 and 3.3:

Phase 1: Data Pre-processing

Step 1 : Collect and identify the data format. Convert the data into fuzzy data if the data in a single form of interval F_i parameter.

Step 2 : Identify the central point and width of X_i Fuzzy data interval.

Step 3 : Identify the underlying centre and width distribution by using the method of the probability density

Step 4 : Use data from the parameter Y_i measure the expected value and variance.

Phase 2: Model Building (Portfolio Selection Model)

Step 5 : Construct the variance model according to Equation (13), and apply the expected values and uncertainties in the portfolio selection model.

Phase 3: Risk Assessment

Step 6 : Use the list of standard deviation based on centre point variance as k value to obtain the optimal solution. The optimal solution is indicated with only one portfolio and $\sum_{i=1}^{n} x_i = 1$.

Step 7 : Identify the acceptance region based on the t-test.

The suggested method aims to provide a list of the best options for decision-makers to construct selection model while also addressing data uncertainties.

NUMERICAL EXPERIMENT

This section describes some numerical experiments based on the stock market. Data stock (ISRG, GOOGL, LEN, PCLN, and NFLX) derived from https://www.kaggle.com/camnugent/sandp500. The duration of the data is thehalf-year of 2014, from July 1, 2014, to December 31, 2014. The objective is to identify which portfolio is best for investing.The experimental steps proceed, as set out in Section 3.

Phase 1: Fuzzy Random Data Preprocessing

In step 1, the data should first define its original format, either as a single form of data or in the form of intervals.Table 1 shows the data contain minimum and maximum value each of the currency. The identified data is in interval form and it already fuzzy data.The opening price for each of the currency is 137.28,587.65,41.08,1217.70 and 65.18 respectively which did not change each day because it considers that the investor will invest each currency for half year which from 1 July 2014 to 31 December 2014. Table 2 shows the calculation where the data is $a_i = A_i - opening\ price$ and $b_i = B_i - opening\ price$ where A_i is minimum price and $B_i$asthe maximum price.

Table 1. Interval data F_i

CURRENCY	1 JUL	2 JAN	3 JAN	...	31 DEC
ISRG	[135.24, 138.12]	[134.02, 136.40]	[133.62, 134.97]		[175.14, 178.67]
GOOGL	[586.29, 593.66]	[589.11, 594.15]	[589.00, 594.25]		[530.20, 538.40]
LEN	[41.03, 41.83]	[40.91, 41.78]	[40.82, 41.25]		[43.47, 44.45]
PCLN	[1216.75, 1249.00]	[1233.35, 1252.61]	[1238.07, 1247.44]		[1140.21, 1154.00]
NFLX	[65.04, 67.67]	[66.58, 67.98]	[66.98, 67.52]		[48.79, 49.39]

Table 2. Interval data $[a_i, b_i]$

CURRENCY	1 JUL	2 JAN	3 JAN	...	31 DEC
ISRG	[-2.04, 0.84]	[-3.25, -0.87]	[-3.65, -2.30]		[37.86, 41.39]
GOOGL	[-1.36, 6.01]	[1.46, 6.50]	[1.35, 6.60]		[-57.45, -49.25]
LEN	[-0.05, 0.75]	[-0.17, 0.71]	[-0.25, 0.18]		[2.39, 3.37]
PCLN	[-0.95, 31.30]	[15.65, 34.91]	[20.37, 29.74]		[-77.49, -63.70]
NFLX	[-0.13, 2.49]	[1.40, 2.81]	[1.81, 2.34]		[-16.39, -15.78]

After adjusting the fuzzy data based on the opening price, Step 2 begins. The next step shows where the pre-processing of fuzzy random data is performed.

When dealing with such randomness, fuzzy data values are obtained using Equation (10), where the middle and width values of fuzzy data are defined. The$[a_i, b_i]$fuzzy data are translated using centre and width values, $[c_i, w_i]$form. Table 3 tabulates the fuzzy interval data.

Table 3. Fuzzy centre point and width $[c_i, w_i]$

CURRENCY	1 JUL	2 JAN	3 JAN	...	31 DEC
ISRG	[-0.60, 1.44]	[-2.06, 1.19]	[-2.98, 0.67]		[39.63, 1.76]
GOOGL	[2.32, 3.69]	[3.98, 2.52]	[3.98, 2.63]		[-53.35, 4.10]
LEN	[0.35, 0.40]	[0.27, 0.44]	[-0.04, 0.22]		[2.88, 0.49]
PCLN	[15.18, 16.13]	[25.28, 9.63]	[25.05, 4.69]		[-70.60, 6.89]
NFLX	[1.18, 1.31]	[2.10, 0.70]	[2.07, 0.27]		[-16.09, 0.30]

The centre point and the width of the fuzzy data points are defined using Equation (12) from the data chosen for analysis. The centre point and width of the resulting fuzzy data display in Table 2. The function of the distribution of probability is to handle randomness. The Probability Distribution Function will produce four data types for each currency data that uses the distributions Normal, Logistic, Gamma and Weibull. That distribution will provide the p-value during the selection process.The highest p values in this paper represent the best outcomes for random evaluation. Table 4 displays the chosen function of probability distribution dependent on the most significant p-value. LOG signifies distribution of Logistic, W as distribution of Weibull, and γ as distribution of Gamma. The moment estimator is used to estimate the expected value and variance of each result. Table 5 shows the results and complete the data pre-processing steps. The pre-processed data used to identify the best portfolio in the portfolio selection model.

Table 4. Interval number of the probability distribution function

	centre, C	width, W
ISRG	Log (21.71, 6.97)	Log (1.65, 0.36)
GOOGL	Log (-17.90, 15.30)	γ (5.80, 0.76)
LEN	Log (-0.77, 1.81)	Log (0.41, 0.10)
PCLN	Log (-39.46, 38.43)	γ (5.35, 2.07)

NFLX	Log (-5.46, 4.39)	Log (0.70, 0.16)

Table 5. Expected value and variance

		centre, C		**width, W**	
		expected value	**variance**	**expected value**	**variance**
ISRG	x_1	21.71	159.94	1.65	0.42
GOOGL	x_2	-17.9	770.09	4.39	3.33
LEN	x_3	-0.77	10.75	0.41	0.03
PCLN	x_4	-39.46,	4858.3	11.08	22.95
NFLX	x_5	-5.46	63.5	0.70	0.09

Phase 2: Model building (Portfolio Selection Model)

The Equation (13) portfolio selection model is used to describe the formula in (14);

$$
\begin{aligned}
&max && 21.71x_1 + -17.90_{x2} + -0.77_{x_3} + -39.46x_4 + -5.46_{x5} \\
&min && 1.65x_1 + 4.39x_2 + 0.41x_3 + 11.08x_4 + 0.70x_5 \\
&s.t. && \sqrt{159.94x_1} + \sqrt{770.09x_2} + \sqrt{10.75\,x_3} + \sqrt{4858.3x_4} + \sqrt{63.5x_5} = k \\
& && \sqrt{0.42} + \sqrt{3.33x_2} + \sqrt{0.03x_3} + \sqrt{22.95x_4} + \sqrt{0.09x_5} \le k \\
& && x_1 + x_2 + x_3 + x_4 + x_5 \le 1 \\
& && x_i \ge 0 \quad \forall i = 1,2,\dots,n
\end{aligned}
\tag{14}
$$

A linear programming method is used to solve Model (14). k is the risk level and x_n = (1,0,0,0,0)' or x_n = (0,1,0,0,0)' or x_n = (0,0,1,0,0)' or x_n = (0,0,0,1,0)' or x_n = (0,0,0,0,1)' is the optimal solution. When the optimal solution is obtained the calculation to solve the model is stopped. Values need to be tested based on standard deviation of the midpoint to identify the optimal level of risk k.

The equation to test risk level k is $\sqrt{S_{O_{x^n}}}$ n = 1,2,3, … where $S_{O_{x^n}}$is the variance for each of the currency.

Phase 3: Risk Assessment

Table 6 shows the results of the optimal solution. The probability of$k$ = 7.97derived from the results table with an optimized solution where the expected return is (0.69, -4.89) and x_n = (0,0,0,0,1).

Table 6. The result with an optimal solution

k	**x^nMAX**	**x^nMIN**	**E[R(x)]**	**E[R(N)]**	**S_k**	**Accepted Region**
3.28	(0,0,0,0,0.41)	(0,0,0,0,0.41)	(-2.25,0.29)	(-2.28, 0.31)	(2.88,0.14)	[-2.747, -1.733] x [0.2854, 0.3346]
7.97	(0,0,0,0,1)	(0,0,0,0,1)	(-5.46, 0.75)	(-5.90, 0.75)	(7.16, 0.33)	[-7.16, -4.64] x [0.69, 0.81]
12.65	(0,0.24,0,0,0.76)	(0,0,0,0.08,0.92)	(-8.40, 1.49)	-	-	-
27.75	(0,1,0,0,0)	(0,0,0,0.32,0.68)	(-17.9,4.03)	-	-	-
69.70	(0,0,0,1,0)	(0,0,0,1,0)	(-39.46,11.08)	-	-	-

x^nindicate each of the currency whiles$x_n$as the optimal result for the stock market. E[R(x)]is the expected value for minimum objective and E[R(x)]as the maximum value for the objective. Both minimum and maximum value are based on Min and Max objective in Model (13).

Based on Table 6, when k = 3.28, the x_n = (0,0,0,0,0.41). The equation to calculate its return is as follows:

$$J = \sum_{i=1}^{n} j_i x_i = j_1x_1 + j_2x_2 + j_3x_3 + j_4x_4 + j_5x_5 \qquad (15)$$

where the equation simplified for the risk level such as$J = 0.41 * x_4$. It will generate 124 new data based on model (14). Expected value and the standard deviation is calculated based on the new data points, so E[R(N)]is the new expected value and $S_k$as the new standard deviation. The results show E[R(N)] = $(-2.28, 0.31)$and S_k = (2.88,0.14). The hypothesis was H_0: E[R(x)] = $(-2.25, 0.29)$ while the statistic vector such $T \equiv (T_0, T_1)$ where

$$T_0 = \left|\frac{-2.28+2.25}{\frac{2.88}{\sqrt{124}}}\right| = -0.11 \text{and } T_1 = \left|\frac{0.31-0.29}{\frac{0.14}{\sqrt{124}}}\right| = 1.59.$$

Accepted regionpresented in Table 5 asthe 95% confidence interval of the confidence interval where the value of the new data points confidence interval is $[-2.747, -1.733]x[0.2854, 0.3346]$. The hypothesis is accepted where

$$(-2.25, 0.29) \in [-2.747, -1.733]x(0.2854, 0.3346]$$

because $\bar{A} \in$ E[R(x)] so the expected return can be accepted.

CONCLUSIONS

The development of a prediction model depends on the quality of the data used for the analysis. Therefore, pre-processing of data is crucial, as it minimizes errors in the data, which may lead to reduced model accuracy. Errors in the data may be due to uncertainty. Therefore, this study provides a procedure for developing a portfolio selection model presented by pre-processing fuzzy random data that addresses fuzzy and random data. The proposed fuzzy random data pre-processing is to use fuzzy data with probability distributions to address random elements. This paper also introduces the standard deviation method for estimating risk level in model selection. The results reveal that the stock market potential with optimal returns has a significant impact which is determined based on the level of risk.

ACKNOWLEDGEMENTS

The activity presented in the paper is part of the research grant of Fundamental Research Grant Scheme (FRGS) Vote No K208 and GeranPenyelidikanPascasiswazah (GPPS) grant (Vot No H332). This research work also supported by the Ministry of Education, R.O.C., under the grants of TEEP@AsiaPlus, and also supported by the Ministry of Science and Technology under Grant No. MOST 108-2221-E-035-065. The author thanks the anonymous viewers for the feedback.

REFERENCE

1. Arnold,S.F. (1990). Mathematical Statistics, Prentice-Hall, Englewood Cliffs, NJ.
2. Agrawal, P., & Ganesh, T. (2018). Multi-Choice Stochastic Transportation Problem Involving Logistic Distribution.

Advances and Applications in Mathematical Sciences, 18(1), 45–58.
3. Aouni, B., Doumpos, M., Pérez-Gladish, B., & Steuer, R. E. (2018). On the increasing importance of multiple criteria decision aid methods for portfolio selection. Journal of the Operational Research Society, 5682, 1–18. https://doi.org/10.1080/01605682.2018.1475118
4. Chuliá, H., Guillén, M., & Uribe, J. M. (2017). Measuring uncertainty in the stock market. International Review of Economics & Finance, 48, 18–33. https://doi.org/10.1016/j.iref.2016.11.003
5. Dave, C., Eckel, C. C., Johnson, C. A., & Rojas, C. (2010). Eliciting risk preferences: When is simple better? Journal of Risk and Uncertainty, 41(3), 219–243. https://doi.org/10.1007/s11166-010-9103-z
6. Eschenbach, T. G., & Lewis, N. A. (2019). Risk , standard deviation , and expected value : when should an individual start social security ? The Engineering Economist, 64(1), 24–39. https://doi.org/10.1080/0013791X.2018.1532543
7. Hasuike, T., & Mehlawat, M. K. (2018). Investor-friendly and robust portfolio selection model integrating forecasts for financial tendency and risk-averse. Annals of Operations Research, 269(1–2), 205–221. https://doi.org/10.1007/s10479-017-2458-7
8. Henry, W., Espinoza, I., Fernandez, A., & Torres, D. (2017). The Semantic Web as a Platform Against Risk and Uncertainty in Agriculture. 753–760. https://doi.org/10.1007/978-3-319-65151-4
9. Hussain, A., Mkpojiogu, E.O.C., Kamal, F.M. (2016a). Mobile Video Streaming Applications: A Systematic Review Of Test Metrics In Usability Evaluation. Journal of Telecommunication, Electronic and Computer Engineering, 8 (10), pp. 35-39.
10. Hussain, A., Mkpojiogu, E.O.C., Kamal, F.M. (2016b). The Role Of Requirements In The Success Or Failure Of Software Projects. International Review of Management and Marketing, 6 (7Special Issue), pp. 305-310.
11. Imbens, G. W., & Spady, R. (2002). Confidence intervals in generalized method of moments models. Journal of Econometrics, 107(1–2), 87–98. https://doi.org/10.1016/S0304-4076(01)00114-2
12. Islam, R., Ghani, A.B.A., Kusuma, B., Theseira, B.B. (2016). Education And Human Capital Effect On Malaysian Economic Growth. International Journal of Economics and Financial Issues, 6 (4), pp. 1722-1728.
13. Karney, C. F. F. (2016). Sampling exactly from the normal distribution. ACM Transactions on Mathematical Software, 42(1). https://doi.org/10.1145/2710016
14. Kwakernaak, H. (1978). Jlefiitious and 'Ilworems. 29, 1–29.
15. L.A. Zadeh. (1965). Fuzzy Sets. Information and Control, 8(3), 338–353.
16. Lah, M. S. C., Arbaiy, N., & Efendi, R. (2019). Stock Market Forecasting Model Based on AR(1) with Adjusted Triangular Fuzzy Number Using Standard Deviation Approach for ASEAN Countries. In V. Piuri, V. E. Balas, S. Borah, & S. S. Syed Ahmad (Eds.), Intelligent and Interactive Computing (pp. 103–114). Springer Singapore.
17. Lin, P.-C., Watada, J., & Wu, B. (2013). Risk Assessment of a Portfolio Selection Model Based on a Fuzzy Statistical Test. IEICE Transactions on Information and Systems, E96.D(3), 579–588. https://doi.org/10.1587/transinf.E96.D.579
18. Lin, P.-C., Watada, J., & Wu, B. (2013). Identifying the Distribution Difference between Two Populations of Fuzzy Data Based on a Nonparametric Statistical Method. IEEJ Transactions on Electronics, Information and Systems, 8(6), 591-598.
19. Lin, P.-C., Watada, J., & Wu, B. (2012).Portfolio Selection Model with Interval Values Base on Fuzzy Probability Distribution Functions. International Journal of Innovative Computing, Information and Control, 8(8), 5935-5944.
20. Lin, P.-C., Wu, B., & Watada, J. (2012). Goodness-of-Fit Test for Membership Functions with Fuzzy Data. International Journal of Innovative Computing, Information and Control, 8(10), 7437-7450.
21. Lin, P.-C., Watada, J., & Wu, B. (2011).A Database For A New Fuzzy Probability Distribution Function and Its Application. International Journal of Innovative Management, Information and Production, 2(2), 1-7.Madan L Puri, Dan A Ralescu, L. Z. (1993). Fuzzy random variables. 409–422. https://doi.org/978-3-642-13959-8
22. Markowitz, H. M. (1952). Portfolio selection. The Journal of Finance, 7(60), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x
23. Mohammadi, K., Alavi, O., Mostafaeipour, A., Goudarzi, N., & Jalilvand, M. (2016). Assessing different parameters estimation methods of Weibull distribution to compute wind power density. Energy Conversion and Management, 108, 322–335. https://doi.org/10.1016/j.enconman.2015.11.015
24. Righi, M. B., & Borenstein, D. (2018). A simulation comparison of risk measures for portfolio optimization. Finance Research Letters, 24, 199–220. https://doi.org/10.1016/j.frl.2017.07.013
25. Shapiro, A., Dentcheva, D., & Ruszczyǹski, A. (2009). Lectures on stochastic programming: modeling and theory. Technology, 447. https://doi.org/http://dx.doi.org/10.1137/1.9780898718751
26. Vinnem, J. E., & Røed, W. (2020). Environmental Risk Analysis. In Springer Series in Reliability Engineering. https://doi.org/10.1007/978-1-4471-7448-6_18
27. Wang, B., Li, Y., Wang, S., & Watada, J. (2018). A Multi-Objective Portfolio Selection Model with Fuzzy Value-At-Risk Ratio. IEEE Transactions on Fuzzy Systems, 26(6), 3673–3687. https://doi.org/10.1109/TFUZZ.2018.2842752
28. Wang, B. X., & Wu, F. (2018). Inference on the Gamma Distribution. Technometrics, 60(2), 235–244. https://doi.org/10.1080/00401706.2017.1328377
29. Wieczorek, D. (2018). Fuzzy risk assessment in the life cycle of building object - Selection of the right defuzzification method. AIP Conference Proceedings, 1978. https://doi.org/10.1063/1.5043866
30. Wu, T. P., Liu, S. B., & Hsueh, S. J. (2016). The Causal Relationship between Economic Policy Uncertainty and Stock Market: A Panel Data Analysis. International Economic Journal, 30(1), 109–122. https://doi.org/10.1080/10168737.2015.1136668
31. Yoshida, Y. (2018). Dynamic Fuzzy Asset Management for Worst Scenarios with Average Value-at-Risks. Colloquium in Information Science and Technology, CIST, 2018-Octob, 437–442. https://doi.org/10.1109/CIST.2018.8596615
32. Zhong, L., Li, M., & Zhang, D. L. (2010). How do uncertainties in hurricane model forecasts affect storm surge predictions in a semi-enclosed bay? Estuarine, Coastal and Shelf Science, 90(2), 61–72. https://doi.org/10.1016/j.ecss.2010.07.001