**Noname manuscript No.**

(will be inserted by the editor)

**Bootstrapping for Fuzzy Mediation and Moderated-Mediation Analysis**

**Da Jeong Kang1,†· Hyeon Gu Kang1,†· Sung Wook Baik3 · and Jin Hee Yoon\***

Received: date / Accepted: date

**Abstract**

1 Department of Mathematics and Statistics, Sejong University, Seoul 05006, South Korea ; {doramisister, mjsk1145}@naver.com

3 Department of Software, Sejong University, Seoul 05006,

Korea;[sbaik@sejong.ac.kr](mailto:sbaik@sejong.ac.kr)

**\***Correspondence: [jin9135@sejong.ac.kr](mailto:jin9135@sejong.ac.kr); Tel.: +82-10-9118-3135

**†** These two authors contributed equally to this work.

**Keywords:** Causal relationship; Mediation Analysis; Fuzzy Mediation Analysis; Covariates; Solar Power Data

# 1 Introduction

When describing human behavior, social scientists and behavioral scientists hold that when people are exposed to particular stimuli, they do not instantly react but rather do so through internal organic body.  We have been curious in the process by which one phenomenon influences another because of this. By adding a third variable, they attempted to investigate the causal relationship between the independent and dependent variables and better comprehend their relationship, and it was discovered that the mediator and moderator components of this variable are separated.

A mediator variable is a variable that logically intervenes between independent and dependent variables in the causal relationship and is required to explain why or how.

For instance, client satisfaction will increase as a result of a company's satisfied products. In other words, consumers who are happy with the product will also be satisfied with the company; on the other hand, customers who are unhappy with the product will typically be less satisfied. The variable that describes how this relationship impacts is the mediator variable in this instance where there is a static correlation between product satisfaction and customer satisfaction. In this relationship, it may be inferred that as consumer trust in the the product's manufacturer increases, so does the positive correlation between product satisfaction and customer satisfaction. The purpose of the mediation effect study is to identify a variable that may more effectively explain the relationship that exists in the middle by determining if there is a meaningful influence between these two variables. A moderator variable is a variable that influences the amount and direction of the relationship between independent and dependent variables. The purpose of a moderation effect study is to determine how moderator factors affect the intensity or direction of the relationship between independent and dependent variables. The purpose is to determine if, and under what circumstances, when, or from whom, the relationship between the two variables is weaker or stronger. Researchers in several fields have researched this mediation impact and moderation effect. [] Additionally, studies have been conducted in the past to confirm the relevance of combining mediation effects and moderation effects for qualitative comprehension between variables. [] One example of this is the moderated mediation effect, which will be discussed in this article. The term "moderated mediation effect," was first introduced in 1984 by James & Bret [], refers to the regulation of a particular variable or the intensity of the mediation effect. In other words, the mediation effect is either reinforced or diminished as the value of the moderator variable increases (Jame & Bret, 1984). In Fig. 1, the mediation effect, moderation effect, and moderated mediation effect are represented as a simple model.

The regression-based assessment techniques developed by Sobel (1982), Baron and Kenny (1986), and Aroian and Goodman test methods have all been extensively utilized in recent thesis to conduct mediation analyses. However, the examination of the mediation effect using Baron and Kenny (1986) only establishes if the mediation effect exists or not; it does not establish its statistical significance. Furthermore, in the analytical sequence and judgment process in the case of statistical significance of the mediation effect, the other Sobel, Aroian, and Goodman methodes are not simple , and these methodes have weak statistical power and do not account for measurement mistakes in the study model. The method employing bootstrap has lately been utilized in several studies as a way to address this since it is thought to have limitations in terms of confirming the mediation model and that it is not accurate.

In the meantime, the study of these mediation models was carried out using "precise numbers." However, there are certain statistics that are difficult to convey with precise numbers in reality because they contain ambiguous phrasing. It is simple to communicate language connotations like "some" and "moderate," but it might be challenging to work with precise numerical data. Particularly in the area of social science that deals with psychology, we often come across such ambiguous facts, and in trying to describe them in exact figures, we not only risk losing knowledge but also run into issues. It is true that a precise number cannot accurately capture a person's mental aspect, for instance, when a person's degree of stress is assessed as a variable. Additionally, even though this is stated numerically, each person's assessment scale is unique, so even if the data value is the same, it could really be a different value. As a result, if it is coded as it is, information loss is unavoidable. As a result, it makes sense to describe it as a soft number, like the fuzzy number that Zadeh initially proposed.

Yoon carried out a mediation study based on fuzzy theory in 2020 []. However, there has been no research done on the bootstrap paper using fuzzy mediation and fuzzy moderated-mediation. Bootstrap has been prevalently cited as a method for verifying the mediating effect. However, the Baron & Kenny method, an inaccurate statistical method with errors in statistical inference, and the sampling distribution, which shows the biased distribution with the assumption that the sampling distribution forms the normal distribution, are the alternatives to Sobel’s method, which encounters the criticism for not being able to verify statistically significant mediating effects. However, it is a beneficial method with strengths in that it does not require any assumptions for the variable and sampling distribution.)

In example, the bootstrap method, which requires millions of resampling operations, has been more popular lately as access has become simpler and computer speed has increased through the statistical software like AMOS. Therefore, in this study, we suggest utilizing the bootstrapping method to examine the fuzzy mediation model and the fuzzy moderated mediation model.

**2. Fuzzy Mediation and Moderated-Mediation Analysis**

**2.1 기존의 매개효과 검정 방법**

**2.1.1 Baron & Kenny**

Baron & Kenny’s (1986) research has made a clear definition of mediating effect and controlling factors and explained the logic of verification on mediating effect readily intelligibly and intuitively. It is the most widely cited method in the papers as a testing method of the mediating effect by verifying how the mediating effect can be proven.

The method has recently encountered criticisms due to a bevy of problems. When estimating the size of the mediating effect, the conclusion on the mediating effect has been made indirectly by verifying with different figures in order not by verifying from the statistical reasoning to determine if the size has a significant meaning. An error can occur at in anytime, especially when examining a hypothesis. The probability of an error is inevitably getting higher as the number of hypotheses to be proven increases simultaneously. Hence, it has turned out that the reliability of the testing is weak due to excessive errors that occurred from the sequential testing of multiple hypotheses (e.g., Fritz & MacKinnon, 2007; Hayes & Schaarkow, 2013). In addition, it is widely known that Baron & Kenny’s testing method analyzes the mediating effect based on the assumption that the effect of independent variables on the dependent variables should be statistically significant. However, it is not valid. The verification method of the mediating effect is under the criticism that it is not an accurate statistical method rather than it is not statistically close.

**2.1.2 Sobel Test**

The core problem of Baron & Kenny’s verification method occurs indirectly in the verification process of mediating effect. Sabel’s (1982) method can be considered an advanced approach in that the method calculates the magnitude of the effect directly. Researchers frequently cite the Sobel test since the method can be utilized comparatively simply than other methods in verifying the mediating effect. However, it is found that there are defects in Sobel’s verification method. When verifying the significance of the mediating effect with Sobel’s testing method, the assumption is that the sample distribution of the value forms the normal distribution. Unlike the assumption, however, the sampling distribution used widely by most researchers in mediating effect verification is mostly deflective, not showing the normal distribution (Bollen & Stein, 1990; Shrout & Bolger, 2002). Therefore, it can be deduced that Sobel’s method has limitation in telling the statistical significance of mediating effect (Fritz & MacKinnon, 2007; Hayes & Scharkow, 2013).

**2.2 Bootstrapping**

Generally, the confidence interval of the mediating effect has been calculated based on the assumption that the sampling distribution follows the normal distribution or a t distribution. However, the cases where the extracted samples do not follow a normal or t distribution are found. If the confidence interval is calculated with a normal distribution or a t distribution when the sampling distributions are not symmetrical, it can provide an approximation of the correct confidence interval. However, it cannot always provide a reliable approximate value as always. As an alternative to the weakness of the method, the bootstrapping method has recently become pervasive for researchers.

The bootstrapping method is a statistical method that estimates the sample distribution based on the empirical distribution utilizing sample data while the sample distribution is not informed. Namely, given the fact that the method can calculate the approximate standard error, confidence interval, and significance probability of the estimated sample distribution by repeatedly projecting after extracting and restoring the same-sized sample randomly without making any assumptions on the distribution of variables or sample distribution (Figure 0)

Two methods have been suggested for verifying the mediating effect with bootstrapping. First, one is to determine whether zero is included in the confidence interval of the re-extracted sample distribution, and the second is to identify the effect with the significant probability of total indirect effect through the decomposition of testing mediating method. The specific methods that calculate the confidence interval can be categorized into Percentile and Bias-corrected.

The method to examine the mediating effect by utilizing the bootstrapping method has been introduced by several scholars since the 1990s (Bolen & Stein, 1990). Despite the strengths the method has, the reason why the bootstrapping method has not been widely accepted was derived from the difficulty in executing a considerable amount of calculation without using a computer, and there were inevitable limitations in application due to its complexity in programming. However, along with the significant advancement in computer development and the simplified procedures in using bootstrapping through the various statistic packages, the utilization ratio of the method is getting higher in different academic fields. Therefore, in this paper, the statistical significance of the fuzzy mediation model was explained using bootstrap.

# 

**2.2.1 Percentile bootstrap**

When the magnitude of the influence of the independent variable on the parameter is set as a, and the magnitude of the effect of the parameter on the dependent variable, controlling the influence of the independent variable, is set as b, the indirect effect can be defined as ab. Among the reasoning methods that do not require assumptions on the sampling distribution of ab that refers to the magnitude of the indirect effect, there is a typical method that test the indirect effect by utilizing the confidence interval of a bootstrap. One of the procedures to set the confidence interval (95%) by the percentile bootstrap method is as follows (Shrout & Bolger 2002). All procedures are automatically carried out in PROCESS macro, a computer program developed by Hayes.

Reinforcement is extracted from the original sample with sample size N extracted from the population, and a bootstrap sample with the same size N as the original sample is extracted.

1. Using the bootstrap sample obtained in step 1, estimate the statistics of indirect effects in the resampling.
2. Repeat steps 1 and 2 k times to generate k bootstrap samples and estimate and store k indirect effects using them.
3. Sort the k indirect effect estimates from lowest to highest.
4. In the case of using a 95% confidence interval, the lower limit is defined as the statistic value corresponding to the 0.5th (100-95)th percentile of the distribution of the previously obtained statistic value. The upper bound is defined as the statistic corresponding to the [100-0.5 (100-95)]th percentile from the distribution of k statistics arranged in ascending order. The lower and upper bound values are determined as the endpoints of the 95% confidence interval.

If 0 is not included in this 95% confidence interval, the indirect effect is said to be statistically significant.

**2.2.2 Bias-corrected bootstrap**

A bias-corrected bootstrap compliments the potential bias in percentile bootstrap confidence intervals was suggested by Efron and Tibshirani (1986). The Bias-corrected approach shares the same grounds with the percentile confidence interval. However, it is different from the percentile bootstrap confidence interval in that the bias constant is calculated by utilizing the ratio of the point numbers more minor than the point estimate value of the indirect effect of the original sample among the k indirect effect estimates calculated from the k bootstrap samples. It is a revised confidence interval that equals the error rates of both ends of the percentile bootstrap confidence interval. It determines the upper and lower bounds of the confidence interval by closely reflecting the asymmetry of the bootstrap estimate distribution. Therefore, when the sampling distribution of the estimate is not symmetrical, the Bias-corrected method is more suitable for obtaining more accurate results. However, recently, several reports have mentioned that bias-corrected bootstrapping may not be a proper testing method since it causes type I error despite its high proving capability (Biesanz et al.,2010, Hayes & Scharkow, 2013, Falk & Biesanz, 2015, Tofighi & Kelly, 2020).

**3. Fuzzy Mediation and Moderated-Mediation Analysis**

In this section, referring to the basic concepts in [1], we introduce the definition of fuzzy numbers by Zadeh [2], and simple fuzzy mediation models with mediators and fuzzy moderated-mediation model introduced by Yoon [3,4].

**3.1 Fuzzy number**

The definition of a fuzzy number in real numbers R, which implies when normalized and convex, is a fuzzy set. An element of a fuzzy set is one that accepts a real value between -1 and 1 as a measure of belonging according to a function known as the membership function. There are no standard rules since the membership function's form might be described in terms of either objective or subjective possibilities. As a result, the LR-fuzzy numbers parametric class of fuzzy numbers is used in this particular situation. A fuzzy number A is referred to be an LR fuzzy number if it meets the conditions listed below.

where L and R are reference functions called left and right shape functions of X and have the following properties : L,R :R→[0,1] are left-continuous and decreasing function with R(0) = L(0) = 1, R(1) = L(1) = 0. And ‘m’ means the mode of the LR-fuzzy number A. ‘l’ and ‘r’ are greater than 0 and mean the width of the left and right sides. We abbreviate the LR-fuzzy number as . And LR-fuzzy number, one of the triangular numbers, has the following two operations.

= (,

.

**3.2 Simple Fuzzy Mediation Model**

Through simple regression analysis, the Baron and Kenny's Simple Mediation Model's mediation analysis method analyzes the causal relationship step-by-step through. A statistical technique called mediation analysis explores ideas on how a causal antecedent variable (X) influences an outcome variable (Y). We suggest the following derivation of the three regression equations of Baron and Kenny: regression between independent variables and dependent variables, regression between independent variables and mediators, and regression between mediators and dependent variables.

The regression constants and are significant in this model. Here, X's estimated "direct effect" on Y is represented by the number "", while X's estimated "indirect effect" on Y through M is represented by the number " " and the number " ", which is the product of the two. Additionally, X's direct and indirect effects are added together to get , which is known as the "total effect" and equal to . It demonstrates that the direct effect () is less significant than the total effect (.

It makes more sense to describe ambiguous concepts as variables with fuzzy numbers than crisp numbers, such as "more," "less," and "happy." The following is the suggested fuzzy mediation model.

In the model as above, is the total effect, is the indirect effect, and is the direct effect. Note that it is easily checked that

**텍스트, 시계, 손목시계이(가) 표시된 사진

자동 생성된 설명**

**Fig**

**3.3 Fuzzy Mediation Model for Multiple Mediators**

It is often advisable to utilize a basic fuzzy mediation for various mediators rather than a simple fuzzy mediation model since the real world is much more complicated and multi-causal.

The following is a simple fuzzy mediation model with parameters k (k>1).

where

In the model as above, is the total effect, is the indirect effect through and on , and is the direct effect. In other words, there are k indirect effects.



**Fig**

**3.4 Fuzzy Mediated-Moderation Model**

**3.4.1 Mediatied-Moderation Model**

Moderated-mediation is the mechanism by which the moderating variable (W), the fourth variable in the causal relationship, may adjust the indirect effect from the independent variable (X) to the dependent variable (Y) through the parameter (M). The terms "Adjusted mediating effect" and "conditional indirect effect" are presently used interchangeably and have the same meaning in statistics. (Preacher, Rucker & Hayes, 2007)

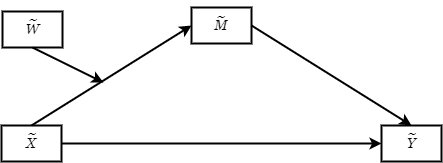
Understanding the conditional character of the mechanism by which one variable effect another variable and testing hypotheses about these conditional effects are the purpose of conditional process analysis, which combines conditioning analysis with mediation analysis. As a result, if the direct and indirect effects of the independent variable (X), which is important in the moderation analysis, are calculated and the influence of the independent variable (X) on the parameter (M), which is organized by the moderating variable (W), then X is not a single number but the effect on M is a function of W.

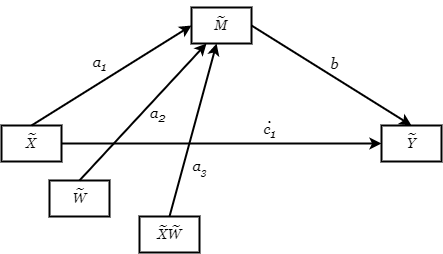
**3.4.2 Fuzzy Mediated-Moderation Model**

Let's say is a fuzzy predictor variable, is a fuzzy response variable, is a fuzzy mediator, and is a fuzzy moderator in a causal relationship involving ambiguous variables. The model shown in the Fig. below is a moderated mediation model, in which controls the route from to but does not influence any other pathways.

Because controls it, the eﬀect of is dependent on . Fig. 1 expresses the following model.

Here, is defined as a function of = and represents the "fuzzy conditional effect" of in . Because of this, it is possible to estimate the "fuzzy conditional indirect impact" from (X) to Y though M ̃ can be estimated as , the effect of the path from to , , which is the product of b and .

****

****

**Fig**

Yoon [3,4] has so far put up a number of fuzzy mediation models . However, no study has used bootstrapping to evaluate this fuzzy model. The initial proposals for bootstrapping-based analysis of parameters are made in Chapter 4 for simple fuzzy parameter analysis and chapter 5's proposals are for analyses of different data types and climatic data.

**4. Bootstrapping for Fuzzy Mediation and Moderated-Mediation Analysis**

**4.1 Estimation for Fuzzy Mediation and Moderated-Mediation Analysis**

When using least squares estimation with fuzzy data, it is necessary to have a suitable metric in the fuzzy set spaces. A helpful type of metric can be established through the use of support functions. The support function of any compact, convex set can be represented as which is determined by the following formula for all :

where is the (d-1)-dimensional unit sphere in and represents the scalar product in . It should be noted that for compact and convex sets the support function is uniquely defined. A metric in a fuzzy number set is established through the use of the *-* metric in the space of Lebesgue integrable functions, represented as:

This leads to the definition of an *-* metric for fuzzy numbers as:

A fuzzy regression model was previously introduced in the author's studies [24,25] and is expressed as follows:

.

The variables are represented by and for It is assumed that are the fuzzy random errors that account for the fuzziness. It is worth mentioning that all cases can be covered by defining and as follows:

where represent the left and right spreads of respectively.

The estimators are obtained by minimizing the following objective function:

where *q* is the number of the regression model in this fuzzy mediation analysis and *k=1,2,…,q*,. The objective function is based on the *-*metric, and the *-* distance can be calculated as:

To minimize the above equation, we obtain the normal equation applying

The normal equation has as its solution, and for each value of , the normal equation can be written as follows:

To determine the solution vector, we introduce a *triangular fuzzy matrix* *(t.f.m.)* which is expressed as

,

and abbreviated as , where is a triangular fuzzy number forandAdditionally, we define a triangular fuzzy vector

.

To minimize the objective function mentioned above, we apply the fuzzy operations, fuzzy numbers and estimators defined in our previous studies [26-29]. The fuzzy operations are as follows:

*.*

The following operations are defined for two triangular fuzzy matrices, , , and a crisp matrix :

,

*,*

*,* .

.

The solutions to the normal equation fuzzy estimators are derived for each by using the above operations and algebraic properties, with

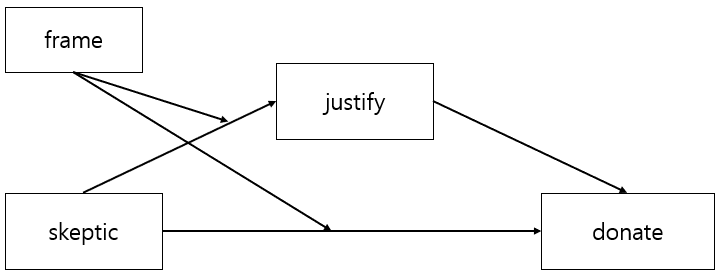
where

and , for Note that the solution (16) exists only if .

**5 Data Analysis**

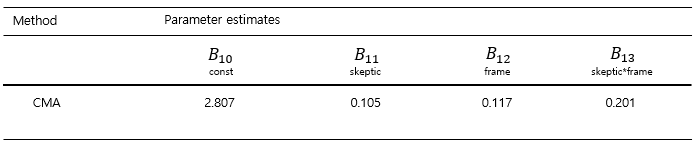
**5.1 Disaster Data**

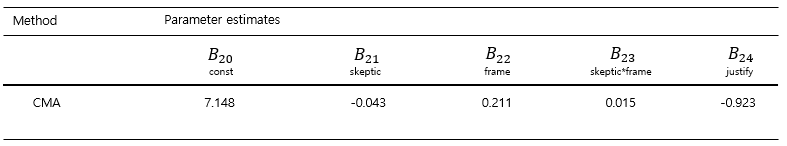
Studies have been conducted that beliefs about climate change affect individual decision-making on environmental issues. 211 participants read an article about famine in sub-Saharan Africa due to severe drought. 101 of them read a framed article that the cause of drought was human-induced climate change. The “donation” variable was measured using five questions. All questions have a 7-point scale (1 = *strongly disagree*, 7 = *strongly agree*), and the higher the score, the more positive the attitude toward donation. The “justify” variable means justification for donation. Higher scores indicated greater justifications to withhold donations. The “skeptic” variable was measured using five questions. All questions have a 9-point scale (1 = *strongly disagree*, 9 = *strongly agree*). And higher scores represent greater skepticism and less concern about climate change. The “frame” variable consists of 0 and 1. 0 means people who read general articles and 1 means people who read framed articles.



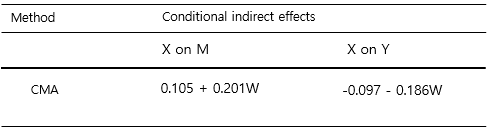
**Fig. (숫자)**

**Table1** Parameter estimates between skeptic and justify

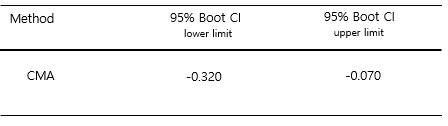


**Table2** Parameter estimates between skeptic + justify and donate

**Table3** Conditional effects of X and M on Y

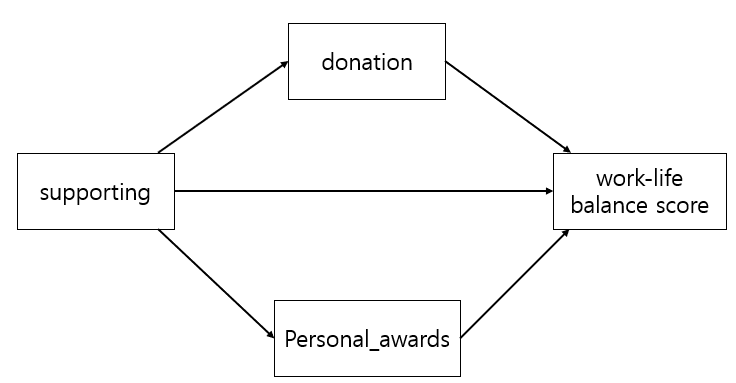
****

**Table4**

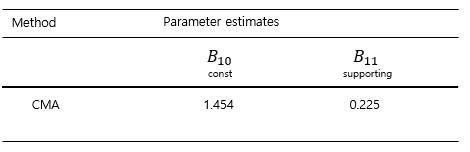
****

**5.2 Work-b Data**

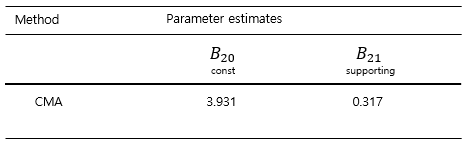
This data was collected by authentic-happiness.com. Participants take an online survey to measure the work-life balance score. The "supporting" variable means the number of people that you have been helping for more than a year(0: None, 10: 10 people or more). The "donation" variable means how many times you donate your money or time more than a year(0: None, 5: 5 times or more). The “personal\_awards” variable means how many achievements you are proud of in the past year.(0: None, 10: 10 achievements or more). The “wlbs” variable means the work-life balance score. The relation between these variables is shown in Fig.

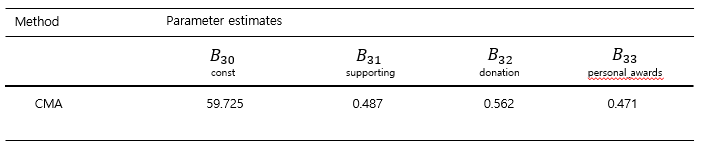
****

**Table1** Parameter estimates between supporting and donation

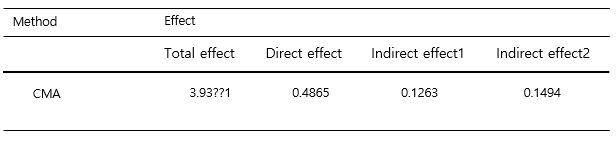


**Table2** Parameter estimates between supporting and personal\_awards

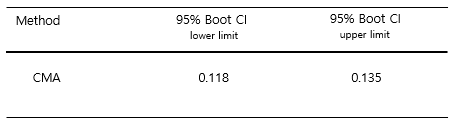


**Table3** Parameter estimates between supporting + donation + personal\_awards and work-life balance score

**Table4** Effects of the supporting on work-life balance score under donation and personal\_awards



**Table5**



**Acknowledgement** This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2020R1A2C1A01011131; No. 2019M3F2A1073179).

**References**