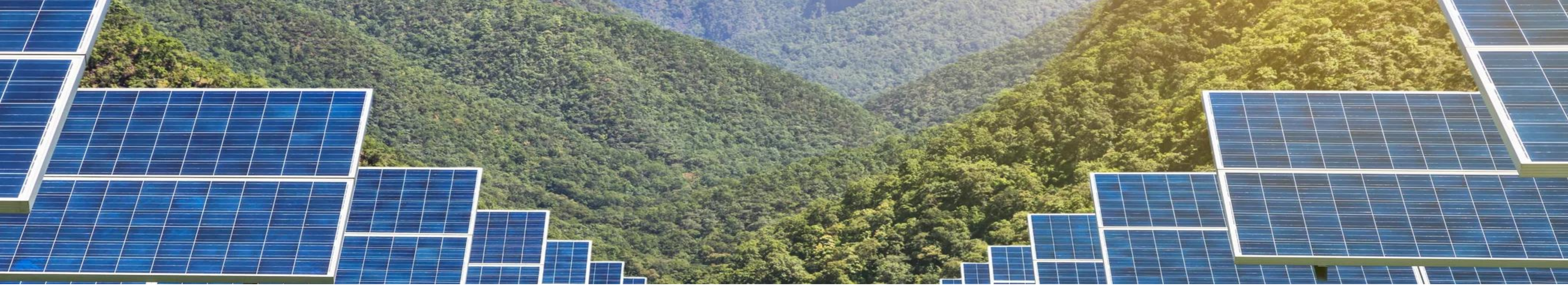
Mediation and Fuzzy Mediation Analysis for Multiple Covariates and Its Applications to Solar Power Data

Department of Mathematics and Statistics, Sejong University
Department of Digital Content, Sejong University

Jungeun Lee, Yeonkyung Lee, Sung Wook Baik and Jin Hee Yoon



01 Introduction

Background

Need for study

02 Preliminaries

Mediation Analysis

Fuzzification

Fuzzy Mediation Analysis

03 Main Contribution

Data Preprocessing

Modeling relationship of data

04 Results

Comparison of effects of CMA & FMA

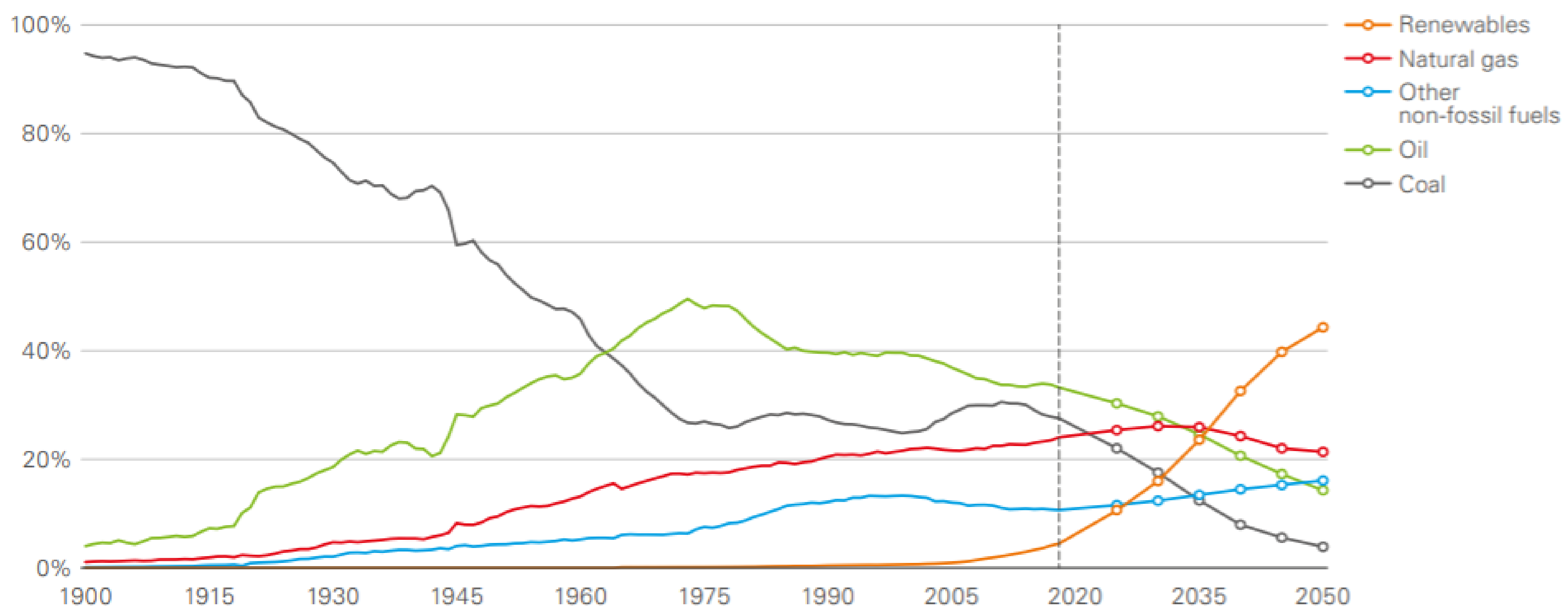
Significance Test of effects of CMA & FMA

05 Conclusion

Summary

Background

BP, Energy Outlook 2020 edition



PROS

Safe
Infinite resource
Restriction-free site
Long life

☞ Easy to recess

CONS

Highly fluctuating

☞ Hard to predict

It is important to predict accurate amount of solar power.

http://energytransitionkorea.org/sites/default/files/2020-09/bp-energy-outlook-2020.pdf

Need for research

01 The focus of prior research with solar power data

Lack of research on the relationships between variables
- ARMA, ARIMA
- SVM, ANN, LSTM

02 The focus of prior mediation study

Absence of mediation model with multiple covariates

03 Characteristics of solar power data

Ambiguity, the key characteristic of climate data

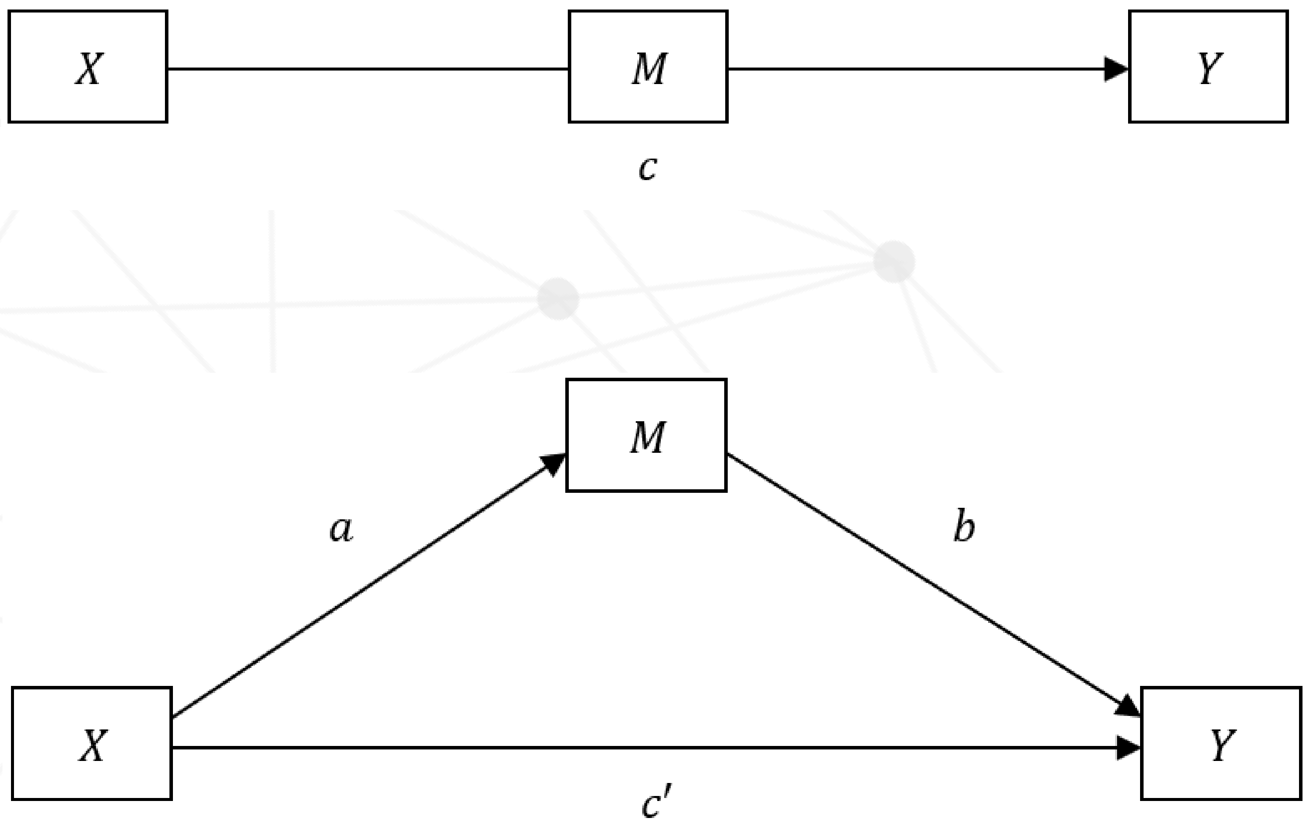
Need for extend independent variable to multiple version

Need for reflect ambiguity of data

Fuzzy Mediation Analysis with multiple covariates

Mediation Analysis

Simple Mediation Analysis by Baron and Kenny(1986)



Step 01 $Y = \beta_{10} + \beta_{11}X + \varepsilon_1$

Step 02 $M = \beta_{20} + \beta_{21}X + \varepsilon_2$

Step 03 $Y = \beta_{30} + \beta_{31}X + \beta_{32}M + \varepsilon_3$

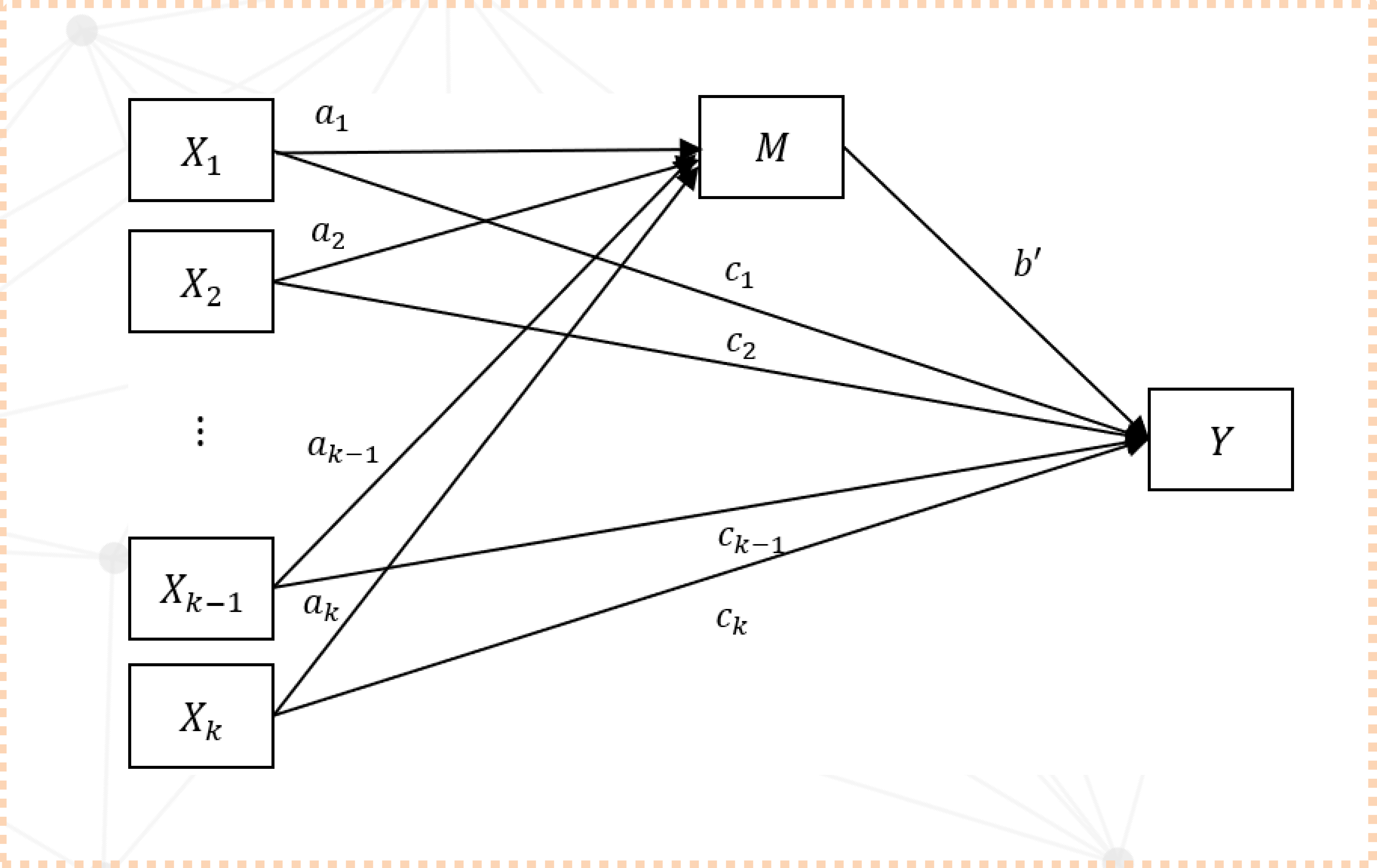
Total effect = Direct effect + Indirect effect

$$\beta_{11} = \beta_{31} + \beta_{21} \cdot \beta_{32}$$

https://psycnet.apa.org/buy/1987-13085-001

Mediation Analysis

Mediation Analysis for multiple covariates with one mediator



Step 01 $Y = \beta_{10} + \beta_{11}X_1 + \beta_{12}X_2 + \cdots + \beta_{1p}X_p + \varepsilon_1$

Step 02 $M = \beta_{20} + \beta_{21}X_1 + \beta_{22}X_2 + \cdots + \beta_{2p}X_p + \varepsilon_2$

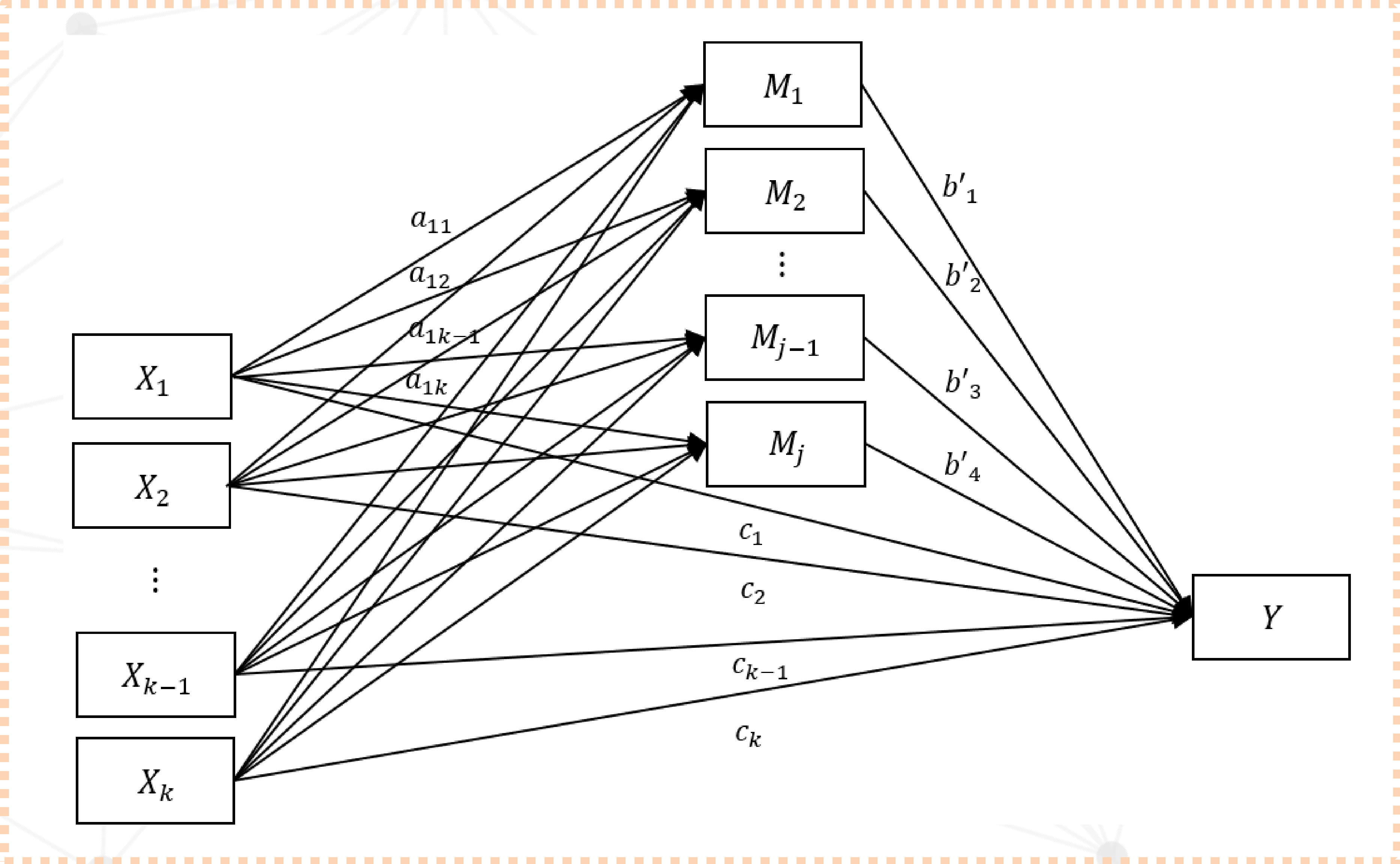
Step 03 $Y = \beta_{30} + \sum_{j=1}^{p} \beta_{31}^{j} X_j + \beta_{32}M + \varepsilon_3$

Total effect = Direct effect + Indirect effect

$$\boldsymbol{\beta_{1j} = \beta_{31}^{j} + \beta_{2j}\beta_{32}}$$

Mediation Analysis

Mediation Analysis for multiple covariates with multiple mediators



Step 01 $Y = \beta_{10} + \beta_{11}X_1 + \beta_{12}X_2 + \cdots + \beta_{1p}X_p + \varepsilon_1$

Step 02 $M_h = \beta_{20}^h + \beta_{21}^h X_1 + \beta_{22}^h X_2 + \cdots + \beta_{2p}^h X_p + \varepsilon_2^h$

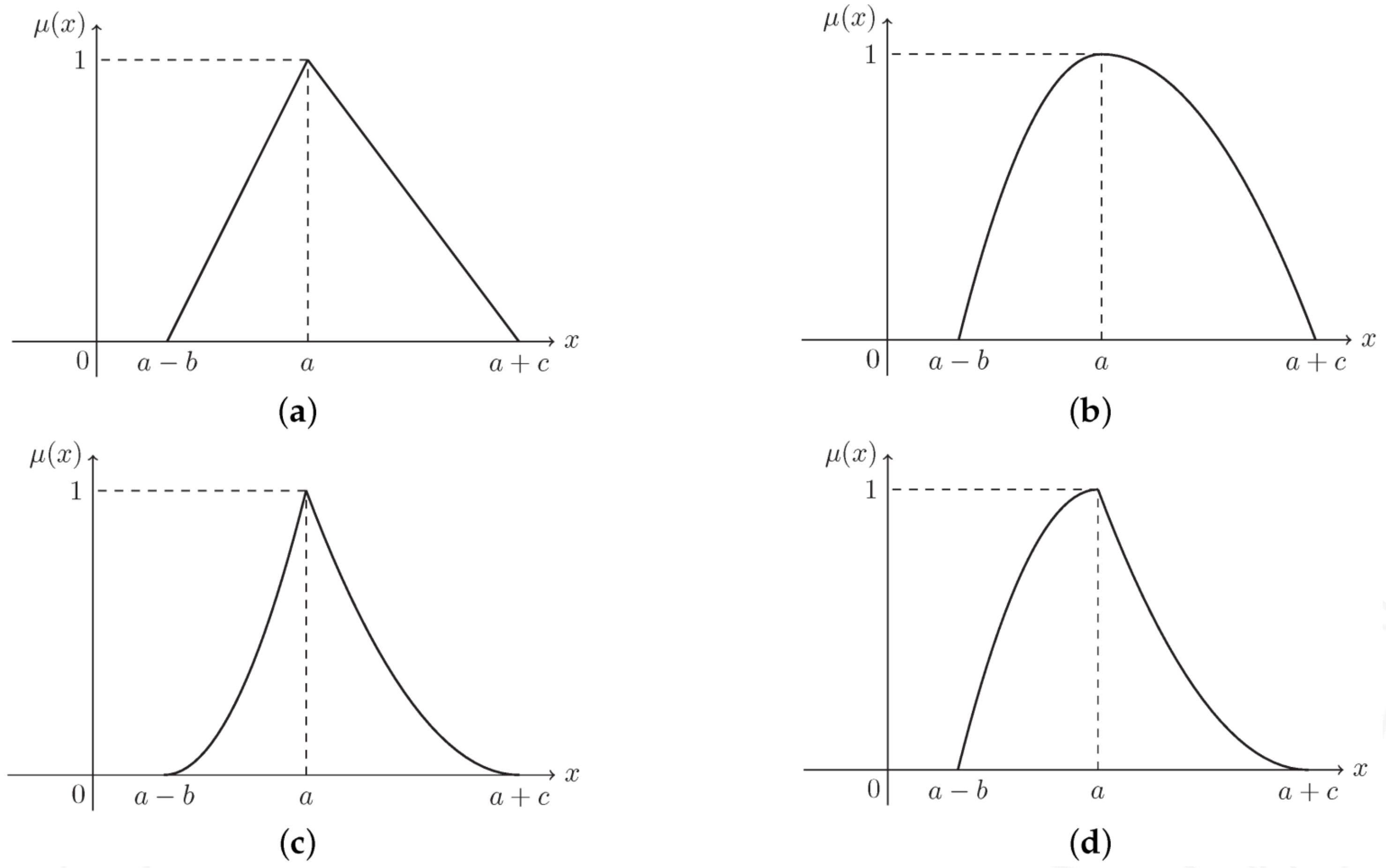
Step 03 $Y = \beta_{30} + \sum_{j=1}^{p} \beta_{31}^j X_j + \sum_{h=1}^{k} \beta_{32}^h M_h + \varepsilon_3$

Total effect = Direct effect + Indirect effect

$$\boldsymbol{\beta_{1j}} = \beta_{31}^j + \beta_{2j}^h \beta_{32}^h$$

Fuzzification

Fuzzy numbers by Zadeh (1986)



L-R fuzzy numbers

$$\mu_A(x) = \begin{cases} L\left(\frac{m-x}{l}\right) & if\ x \le m \\ R\left(\frac{x-m}{l}\right) & if\ x > m \end{cases}$$

m : mode
l : width of left
r : width of right

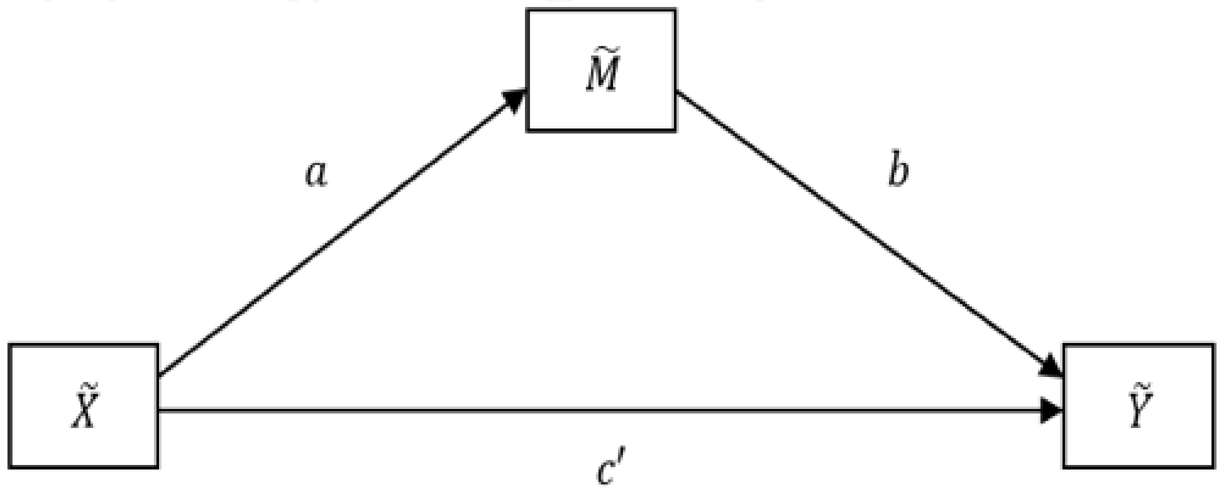
$X = (l_x, x, r_x), Y = (l_y, y, r_y) \in F_T$ for $k \in \mathbf{R}$

$$X \oplus Y = (l_x + l_y, x + y, r_x + r_y)$$

$$kX = \begin{cases} (kl_x, kx, kr_x) & if\ k \ge 0 \\ (kr_x, kx, kl_x) & if\ k < 0 \end{cases}$$

Fuzzy Mediation Analysis

Simple fuzzy mediation by Yoon (2020)



Step 01 $\tilde{Y} = \beta_{10} \oplus \beta_{11}\tilde{X} \oplus \tilde{E}_1$

Step 02 $\tilde{M} = \beta_{20} \oplus \beta_{21}\tilde{X} \oplus \tilde{E}_2$

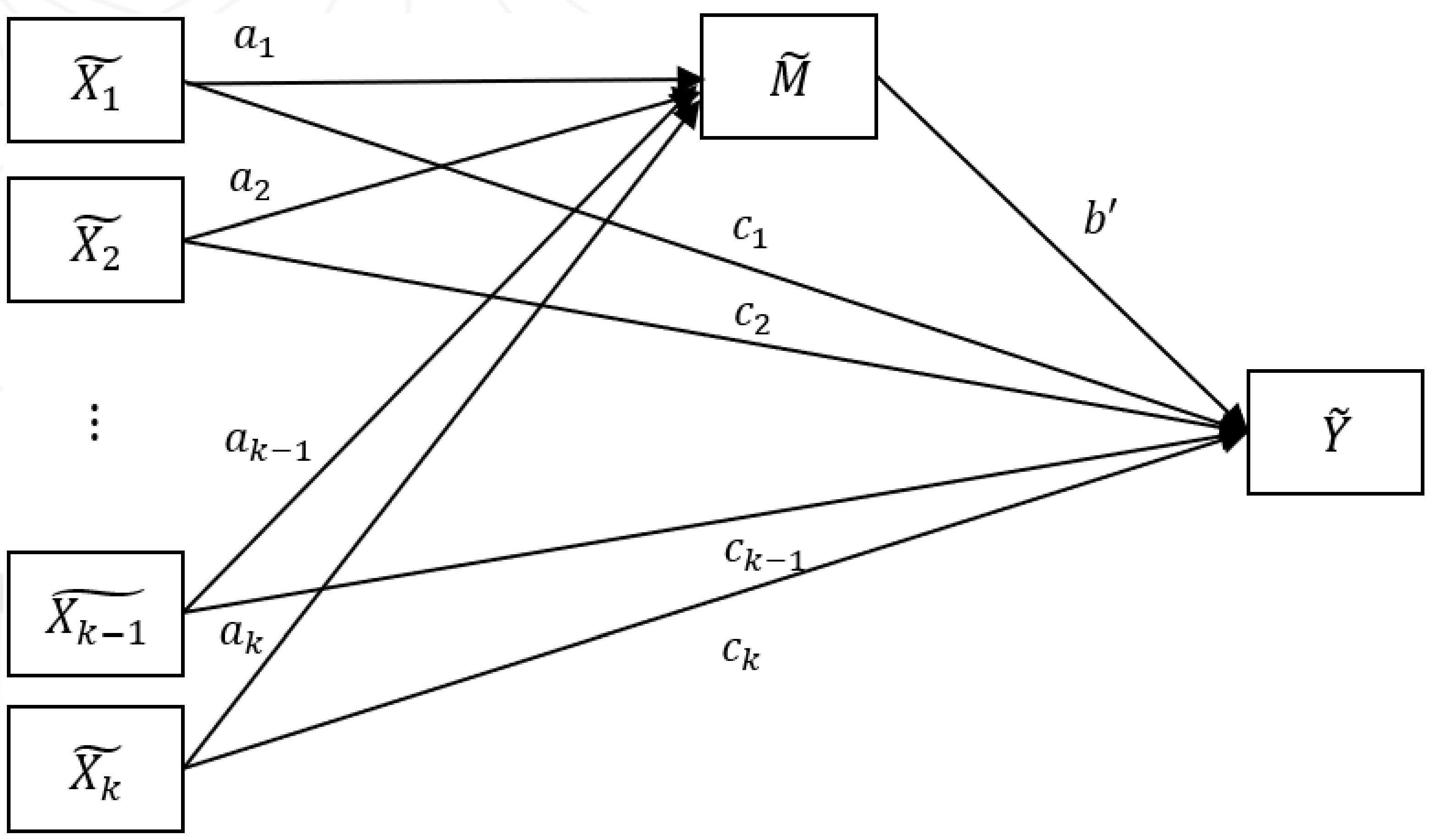
Step 03 $\tilde{Y} = \beta_{30} \oplus \beta_{31}\tilde{X} \oplus \beta_{32}\tilde{M} \oplus \tilde{E}_3$

Total effect = Direct effect + Indirect effect

$$\beta_{11} = \beta_{31} + \beta_{21} \cdot \beta_{32}$$

Fuzzy Mediation Analysis

Fuzzy Mediation Analysis for multiple covariates with one mediator



Step 01 $\tilde{Y} = \beta_{10} \oplus \beta_{11}\tilde{X}_1 \oplus \cdots \oplus \beta_{1p}\tilde{X}_p \oplus \tilde{E}_1,$

Step 02 $\tilde{M} = \beta_{20} \oplus \beta_{21}\tilde{X}_1 \oplus \cdots \oplus \beta_{2p}\tilde{X}_p \oplus \tilde{E}_2$

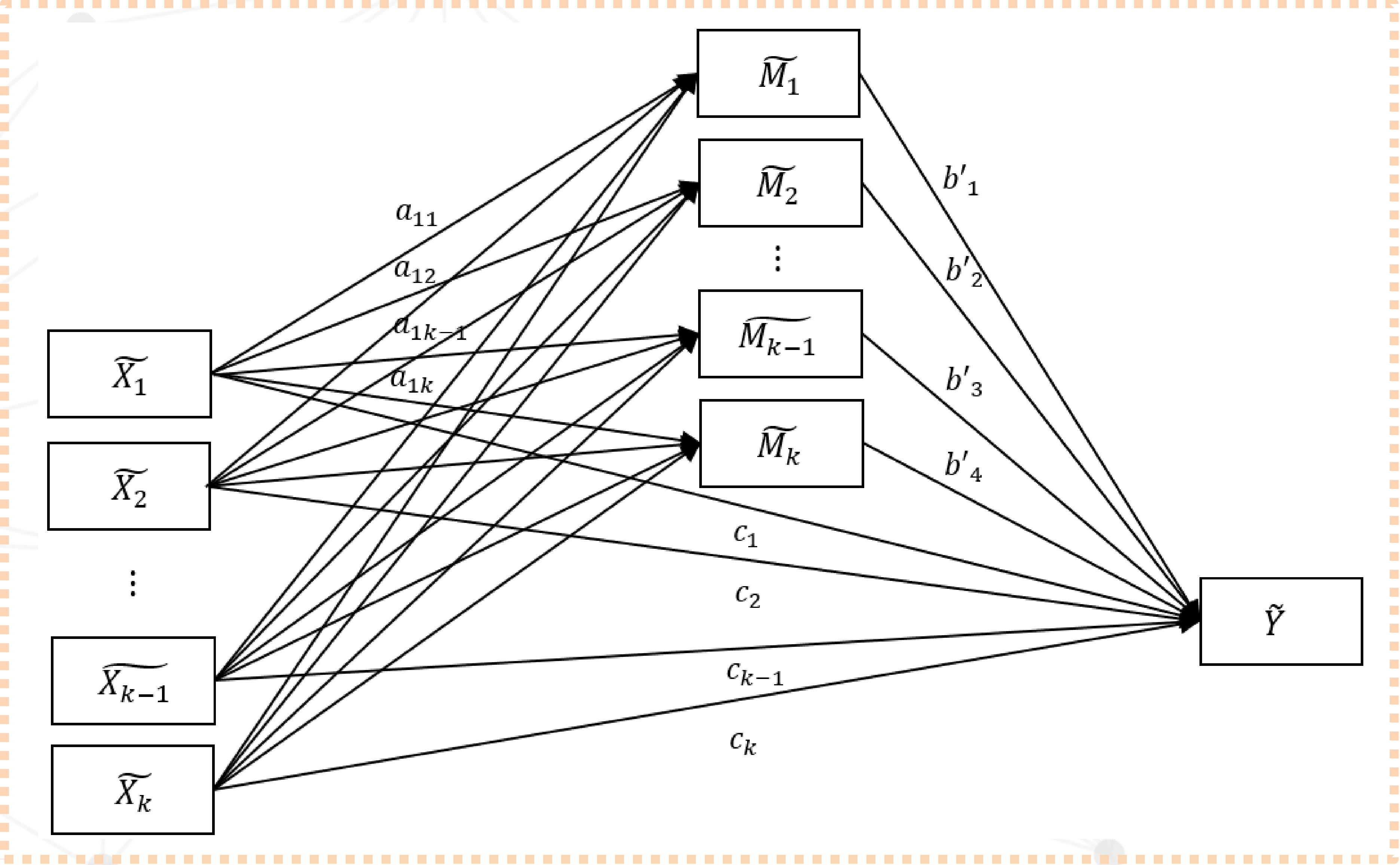
Step 03 $\tilde{Y} = \beta_{30} \oplus \sum_{j=1}^{p} \beta_{31}^{j}\tilde{X}_j \oplus \beta_{32}\tilde{M} \oplus \tilde{E}_3$

Total effect = Direct effect + Indirect effect

$$\boldsymbol{\beta_{1j} = \beta_{31}^{j} + \beta_{2j}\beta_{32}}$$

Fuzzy Mediation Analysis

Fuzzy Mediation Analysis for multiple covariates with multiple mediators



Step 01 $\tilde{Y} = \beta_{10} \oplus \beta_{11}\tilde{X}_1 \oplus \cdots \oplus \beta_{1p}\tilde{X}_p \oplus \tilde{E}_1$

Step 02 $\tilde{M}_h = \beta_{20} \oplus \beta_{21}^h\tilde{X}_1 \oplus \cdots \oplus \beta_{2p}^h\tilde{X}_p \oplus \tilde{E}_2$

Step 03 $\tilde{Y} = \beta_{30} \oplus \sum_{j=1}^{p} \beta_{31}^j\tilde{X}_i \oplus \sum_{h=1}^{k} \beta_{32}^h\tilde{M}_h \oplus \tilde{E}_3$

Total effect = Direct effect + Indirect effect

$$\boldsymbol{\beta_{1j}} = \beta_{31}^j + \beta_{2j}^h\beta_{32}^h$$

Estimation in Fuzzy Mediation Analysis

LSE Method

$$\tilde{Y}_i = \beta_0 \oplus \beta_1 \tilde{X}_{1i} \oplus \beta_2 \tilde{X}_{2i} \oplus \cdots \oplus \beta_p \tilde{X}_{pi} \oplus \tilde{E}_i$$

$$Q(\beta_{k0}, \beta_{k1}, \ldots, \beta_{kp}) = \sum_{i=1}^{n} d^2\left(\tilde{Y}_i, \sum_{j=0}^{p} \beta_{kj} \tilde{X}_{ij}\right) \rightarrow \frac{\partial Q}{\partial \beta_{kl}} = 0$$

where

$$d^2\left(\tilde{Y}_i, \sum_{j=0}^{p} \beta_{kj} \tilde{X}_{ij}\right) = \left(l_{y_i} - \sum_{j=0}^{p} \beta_{kj} l_{x_{ij}}\right)^2 + \left(y_i - \sum_{j=0}^{p} \beta_{kj} x_{ij}\right)^2 + \left(r_{y_i} - \sum_{j=0}^{p} \beta_{kj} r_{x_{ij}}\right)^2$$

$$d^2(X, Y) = D_2^2(Supp\, X, Supp\, Y) + [m_l(X) - m_l(Y)]^2 + [m_r(X) - m_r(Y)]^2$$

$$\widehat{\boldsymbol{\beta}_k} = \left(\tilde{\boldsymbol{X}}^t \diamond \tilde{\boldsymbol{X}}\right)^{-1} \tilde{\boldsymbol{X}}^t \diamond \tilde{\boldsymbol{y}}$$

where

$$\tilde{X}^t \diamond \tilde{X} = \left[\sum_{i=1}^{n} (l_{x_{il}} l_{x_{ij}} + x_{il} x_{ij} + r_{x_{il}} r_{x_{ij}})\right]_{(p+1)\times(p+1)}$$

$$\tilde{X}^t \diamond \tilde{y} = \left[\textstyle\sum_{i=1}^{n} (l_{x_{il}} l_{y_i} + x_{il} y_i + r_{x_{il}} r_{y_i})\right]_{(p+1)\times 1}$$

$$\tilde{X} = \begin{bmatrix} (1,1,1) & (l_{x_{11}}, x_{11}, r_{11}) & \cdots & (l_{x_{1p}}, x_{1p}, r_{1p}) \\ & \vdots & \ddots & \vdots \\ (1,1,1) & (l_{x_{n1}}, x_{n1}, r_{n1}) & \cdots & (l_{x_{np}}, x_{np}, r_{np}) \end{bmatrix}$$

$$\tilde{\boldsymbol{y}} = \left[(l_{y_1}, y_1, r_{y_1}), \cdots, (l_{y_n}, y_n, r_{y_n})\right]^t$$

Inference of the effects

Step 01 Inference of Total effect and Direct effect

$$(1-\alpha)100\%\ CI\ for\ the\ total\ effect\ c_T:\ c \pm z_{\frac{\alpha}{2}} \cdot se(c)$$

$$se(c) = se(c') = \frac{SD}{\sqrt{n}}$$

$$Z = \frac{c}{se(c)} \sim N(0,1)$$

$$H_0:\ c_T = 0\ v.s. H_1:\ c_T \neq 0$$

$$CSD = \sqrt{\frac{1}{n-1}\sum_{h=1}^{n}(X_{ih}-\bar{X})^2}\ , FSD = \sqrt{\frac{1}{n-1}\sum_{h=1}^{n} d^2(\tilde{X}_{ih}, \bar{\tilde{X}})}.$$

CSD : Crisp Standard Deviation
FSD : Fuzzfied Standard Deviation

Step 02 Inference of Indirect effect

$$(1-\alpha)100\%\ CI\ for\ the\ indirect\ effect\ a_T b_T:\ ab \pm z_{\frac{\alpha}{2}} \cdot se(ab) a_T$$

$$se(ab) = \sqrt{a^2 se_b^2 + b^2 se_a^2 + se_a^2 se_b^2}\ \text{(del}$$

$$Z = \frac{ab}{se(ab)} \sim N(0,1)$$

$$H_0:\ a_T b_T = 0\ v.s. H_1:\ a_T b_T \neq 0$$

1. Data Preprocessing

Climate condition data in Dangjin, Korea

Every hour from 1 a.m., Jan 1, 2015, to 11 p.m., Dec 31, 2017

Step 01 Linear interpolation for missing values

$$f(x_k) = f(x_{k-1}) + \frac{f(x_{k+1}) - f(x_{k-1})}{x_{k+1} - x_{k-1}}(x_k - x_{k-1})$$

Step 02 Significance Test for independent variables

temp, rain, wind_speed, humidity, solar radiation, sun hour, snow, cloud, solar power

☞ all significant

Step 03 Normalization

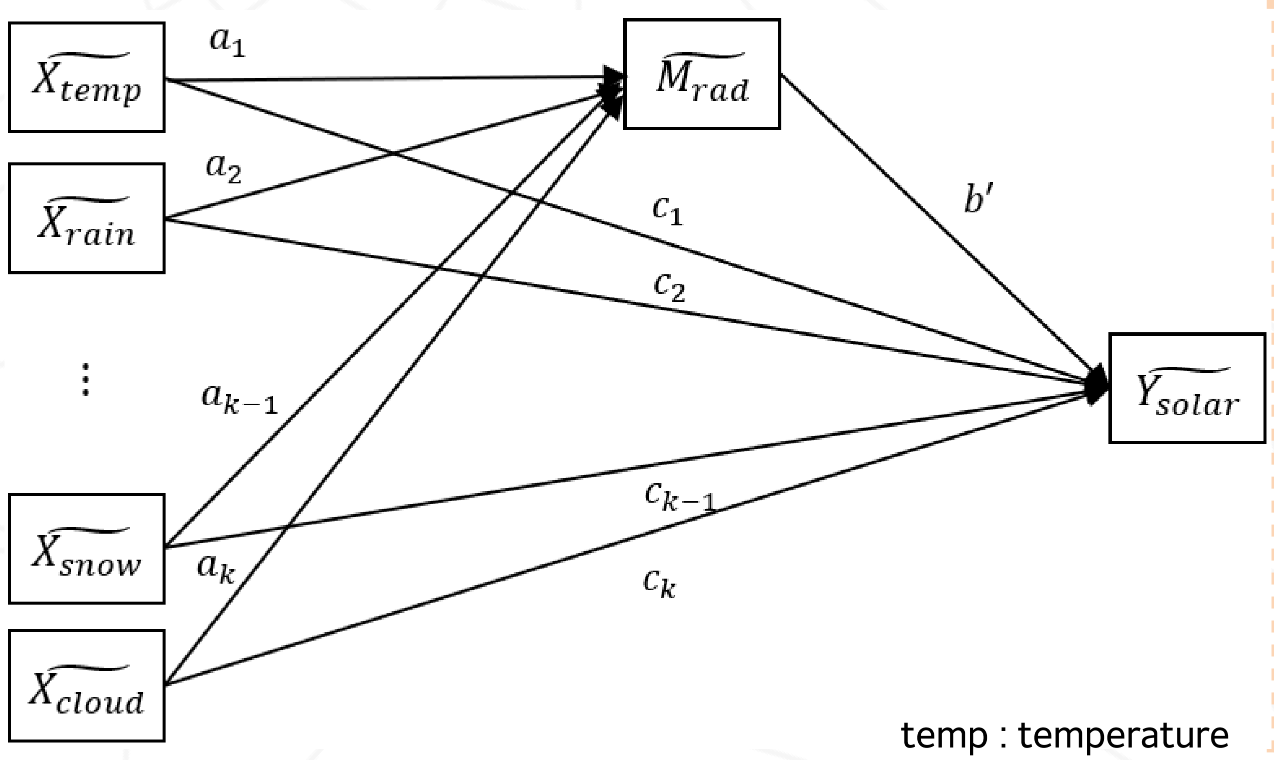
$$x = \frac{x_0 - x_{min}}{x_{max} - x_{min}}$$

Step 04 Fuzzification

Date	Crisp	m, l, r	Fuzzified
01-01-2015 1:00AM	-4.4	m : -4.4, l, r : |-4.6-(-4.4)|/2 = 0.1	(-4.5,-4.4,-4.3)
01-02-2015 2:00AM	-4.6	m : -4.6, l, r : |-4.7-(-4.6)|/2 = 0.05	(-4.65, -4.6, -4.55)
01-03-2015 3:00AM	-4.7	m : -4.7, l, r : |-5-(-4.7)|/2 = 0.15	(-4.85, -4.7, -4.55)

2. Modeling with solar power data

Fuzzy Mediation Analysis for multiple covariates with one mediator



temp : temperature
rad : solar radiation

Step 01 $\tilde{Y} = \beta_{10} \oplus \beta_{11}\tilde{X}_1 \oplus \beta_{12}\tilde{X}_2 \oplus \beta_{13}\tilde{X}_3 \oplus \beta_{14}\tilde{X}_4 \oplus \beta_{15}\tilde{X}_5 \oplus \beta_{16}\tilde{X}_6 \oplus \beta_{17}\tilde{X}_7 + \varepsilon_1$

Step 02 $\tilde{M} = \beta_{20} \oplus \beta_{21}\tilde{X}_1 \oplus \beta_{22}\tilde{X}_2 \oplus \beta_{23}\tilde{X}_3 \oplus \beta_{24}\tilde{X}_4 \oplus \beta_{25}\tilde{X}_5 \oplus \beta_{26}\tilde{X}_6 \oplus \beta_{27}\tilde{X}_7 + \varepsilon_2$

Step 03 $\tilde{Y} = \beta_{30} \oplus \beta_{31}\tilde{X}_1 \oplus \beta_{32}\tilde{X}_2 \oplus \beta_{33}\tilde{X}_3 \oplus \beta_{34}\tilde{X}_4 \oplus \beta_{35}\tilde{X}_5 \oplus \beta_{36}\tilde{M}_6 \oplus \beta_{37}\tilde{X}_7 \oplus \beta_{38}\tilde{X}_8 + \varepsilon_3$

Total effect = Direct effect + Indirect effect

$$\sum_{i=1}^{7} \beta_{1i} = \sum_{i=1, i\neq 6}^{8} \beta_{3i} + \sum_{i=1}^{7} \beta_{2i} \cdot \beta_{36}$$

3. Calculation of coefficients with CMA & FMA and Comparison of their effects

Step 01

Method	Parameter estimates							
	β_{10} const	β_{11} temp	β_{12} rain	β_{13} windspeed	β_{14} humidity	β_{15} sun hour	β_{16} snow	β_{17} cloud
CMA	0.114	0.240	- 0.718	0.153	- 0.291	0.327	0.038	0.064
FMA	0.112	0.248	- 0.585	0.170	- 0.291	0.316	0.043	0.056

Step 02

Method	Parameter estimates							
	β_{20} const	β_{21} temp	β_{22} rain	β_{23} windspeed	β_{24} humidity	β_{25} sun hour	β_{26} snow	β_{27} cloud
CMA	0.107	0.201	-0.601	0.113	-0.255	0.360	0.063	0.054
FMA	0.104	0.208	- 0.487	0.130	- 0.254	0.350	0.067	0.047

Step 03

Method	Parameter estimates								
	β_{30} const	β_{31} temp	β_{32} rain	β_{33} windspeed	β_{34} humidity	β_{35} sun hour	β_{36} solar radiation	β_{37} snow	β_{38} cloud
CMA	0.005	0.033	- 0.101	0.037	- 0.029	- 0.042	1.027	- 0.027	0.009
FMA	0.006	0.035	- 0.088	0.038	- 0.031	- 0.041	1.027	- 0.026	0.008

CMA : Crisp Mediation Analysis
FMA : Fuzzy Mediation Analysis

Method	Effect		
	Total effect	Direct effect	Indirect effect
CMA	-0.187	-0.120	-0.067^{a}
FMA	-0.043	-0.105	0.062^{b}

Fuzzy Mediation Analysis
prevents the effects from overestimating

4. Significance Test of effects with CMA & FMA

temperature

Effect	Method	95% CI (lower bound)	95% CI (upper bound)	$z(t)$	p-value
Total	CMA	-0.191	-0.183	-82.318	<0.001
	FMA	-0.048	-0.038	-18.344	<0.001
Direct	CMA	-0.128	-0.112	-27.853	<0.001
	FMA	-0.110	-0.100	-44.795	<0.001
Indirect	CMA	-0.074	-0.060	-17.607	<0.001
	FMA	0.057	0.067	25.849	<0.001

wind speed

Effect	Method	95% CI (lower bound)	95% CI (upper bound)	$z(t)$	p-value
Total	CMA	-0.194	-0.180	-54.155	<0.001
	FMA	-0.046	-0.040	-28.122	<0.001
Direct	CMA	-0.133	-0.107	-17.465	<0.001
	FMA	-0.108	-0.102	-68.671	<0.001
Indirect	CMA	-0.079	-0.055	-11.225	<0.001
	FMA	0.059	0.065	39.632	<0.001

rain

Effect	Method	95% CI (lower bound)	95% CI (upper bound)	$z(t)$	p-value
Total	CMA	-0.271	-0.103	-4.348	<0.001
	FMA	-0.043	-0.043	-366.673	<0.001
Direct	CMA	-0.289	0.049	-1.394	0.1633
	FMA	-0.105	-0.105	-895.365	<0.001
Indirect	CMA	-0.213	0.079	-0.898	0.369
	FMA	0.061	0.063	184.145	<0.001

humidity

Effect	Method	95% CI (lower bound)	95% CI (upper bound)	$z(t)$	p-value
Total	CMA	-0.193	-0.181	-60.349	<0.001
	FMA	-0.047	-0.039	-21.180	<0.001
Direct	CMA	-0.132	-0.108	-20.259	<0.001
	FMA	-0.109	-0.101	-51.718	<0.001
Indirect	CMA	-0.077	-0.057	-12.843	<0.001
	FMA	0.058	0.066	29.798	<0.001

4. Significance Test of effects with CMA & FMA

sun hour

Effect	Method	95% CI (lower bound)	95% CI (upper bound)	$z(t)$	p-value
Total	CMA	-0.191	-0.183	-99.014	<0.001
	FMA	-0.052	-0.034	-9.500	<0.001
Direct	CMA	-0.125	-0.115	-44.084	<0.001
	FMA	-0.114	-0.096	-23.198	<0.001
Indirect	CMA	-0.072	-0.062	-24.819	<0.001
	FMA	0.053	0.071	13.391	<0.001

snow

Effect	Method	95% CI (lower bound)	95% CI (upper bound)	$z(t)$	p-value
Total	CMA	-0.201	-0.173	-26.059	<0.001
	FMA	-0.044	-0.042	-66.492	<0.001
Direct	CMA	-0.148	-0.092	-8.348	<0.001
	FMA	-0.106	-0.104	-162.365	<0.001
Indirect	CMA	-0.091	-0.043	-5.377	<0.001
	FMA	0.061	0.063	93.640	<0.001

cloud

Effect	Method	95% CI (lower bound)	95% CI (upper bound)	$z(t)$	p-value
Total	CMA	-0.189	-0.185	-161.048	<0.001
	FMA	-0.052	-0.034	-9.672	<0.001
Direct	CMA	-0.124	-0.116	-52.312	<0.001
	FMA	-0.114	-0.096	-23.618	<0.001
Indirect	CMA	-0.071	-0.063	-33.541	<0.001
	FMA	0.053	0.070	13.650	<0.001

Reject H_0 in all independent variables in FMA

VS

Cannot reject H_0 in ‘rain’ variable in CMA

☞ Using without considering ambiguous information can lead to biased results

Summary

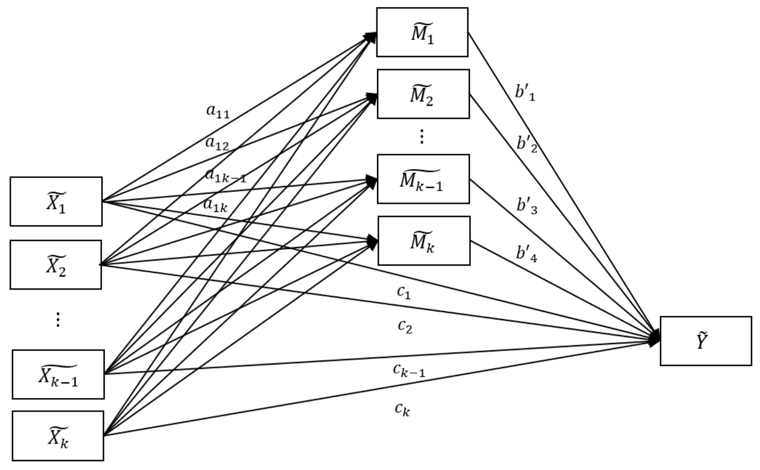
Need for study → Data preprocessing & Modeling → Result → Follow-up study

Need for study

- Lack of research on the **relationships between variables**

- Lack of reflection of **characteristics of climate information data**

Data preprocessing & Modeling



Result

Fuzzy Mediation Analysis

- prevents the effects from overestimating
- prevents the effects from biased results

Follow-up study

To be extended to Moderation Analysis, Mediation-Moderation Analysis, etc.

THANK YOU