September 25th, The 21st Summer course

Bayesian sample selection model with multinomial endogenous switching for non-randomly missing travel behavior outcomes

Hajime Watanabe[a)] and Takuya Maruyama[b)]

a) The University of Tokyo b) Kumamoto University

Background

Experimental study (e.g., randomized control trial (RCT))

- An experiment **randomly** assigns people to treatment or control group
- The causal effect is a difference between treatment and control groups

Quasi-experimental study (Observational study)

- It uses observational data and estimates the causal effect statistically
- It **lacks** the element of random assignment to treatment or control group
- It needs to address the potential **non-random** assignment
  ➡ **Residential Self-Selection (RSS)**

✓ Travel-related attitudes play an important role in residential choice
  ➡ causes the **non-random** assignment, namely, **RSS**

✓ Travel-related attitudes are rarely observed in travel survey

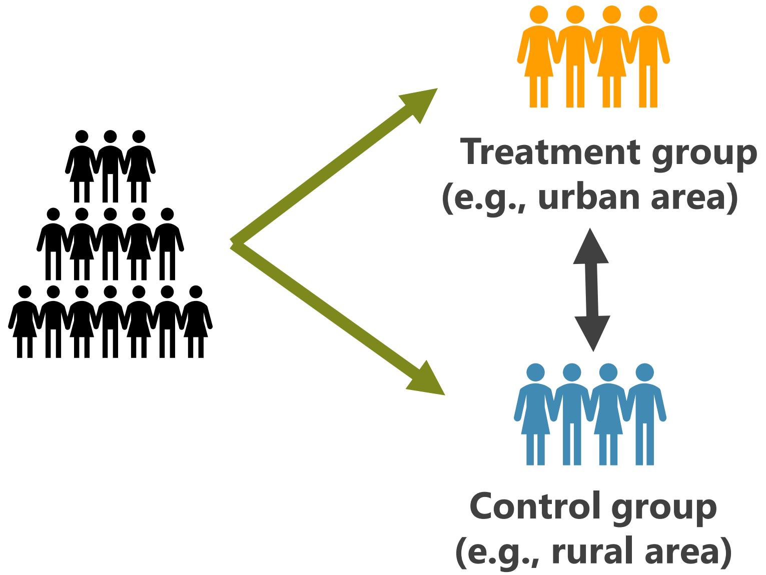


Fig1. Assignment to treatment or control group

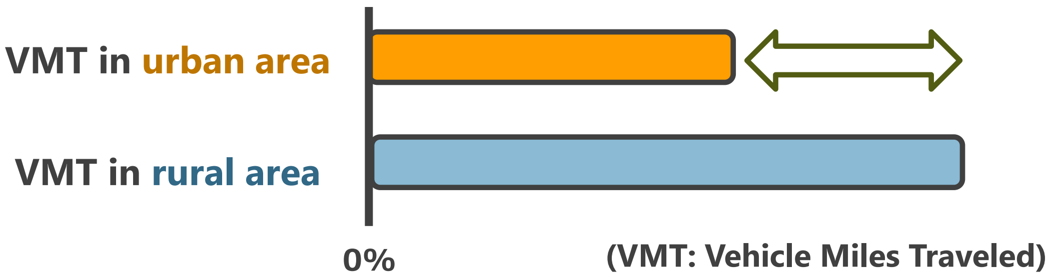


Fig2. Difference in VMT between distinct two regions

Research objective

The sample selection modeling approach is a quasi-experimental study framework and handles the non-random assignment to treatment or control group (endogeneity issue due to RSS)

Strength: This approach does not require instrumental variables (IV) and other indicators unlike IV and MIS approach

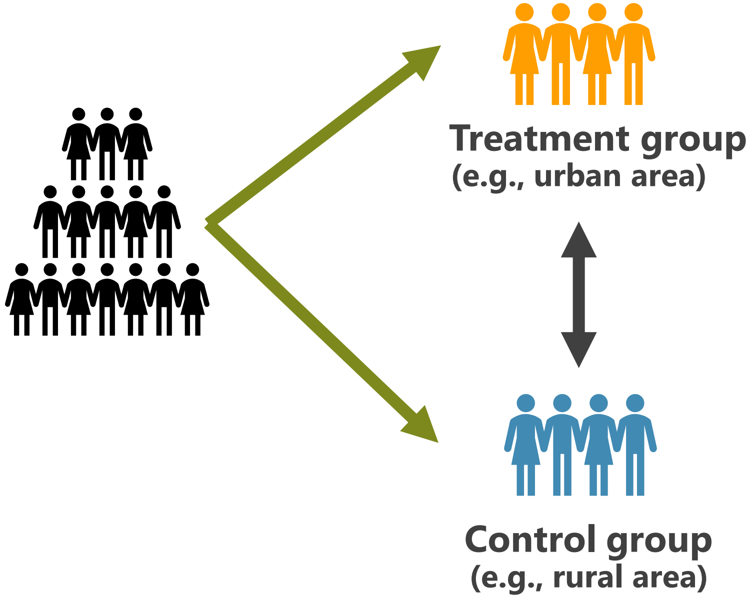
Weakness: This approach must assume treatment and control groups, which is restrictive for analysis

Challenge

Existing sample selection models are too simple for travel behavior analysis

Objective

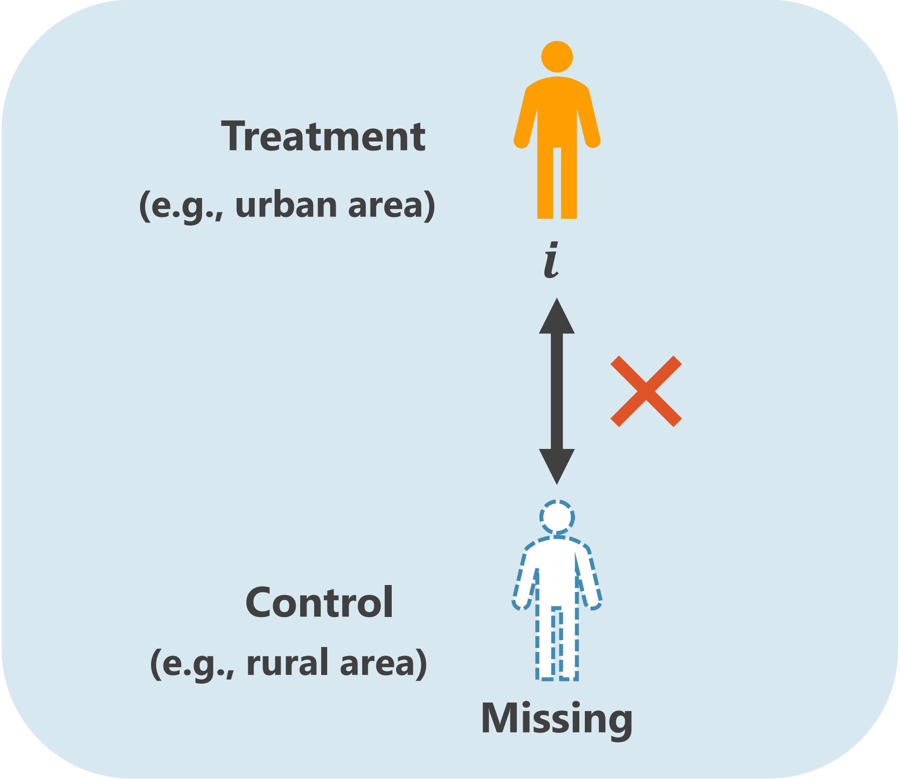
- To propose a new extended sample selection model to identify the causal and RSS effects on travel behavior



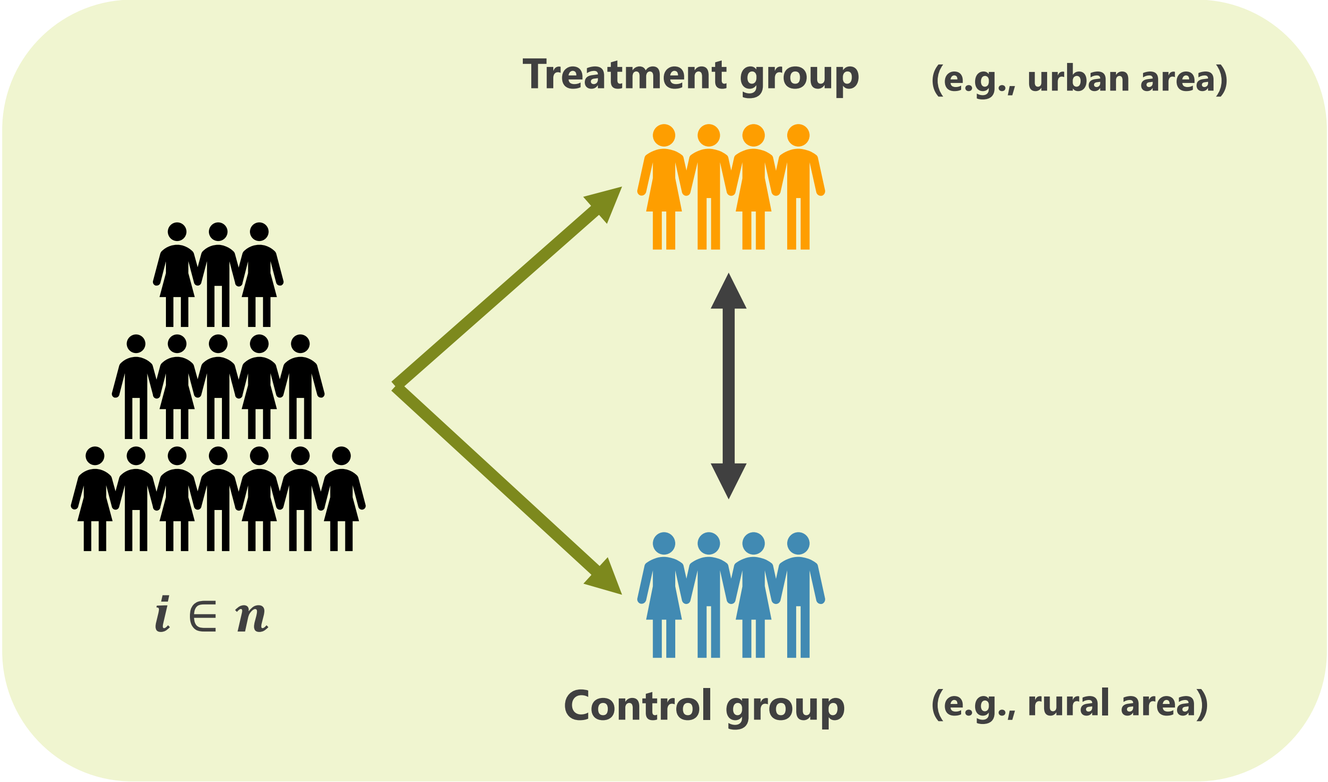
Quasi-experimental study:

Sample selection model in the Rubin Causal Model (RCM) framework

Rubin Causal Model (RCM) framework for causal inference



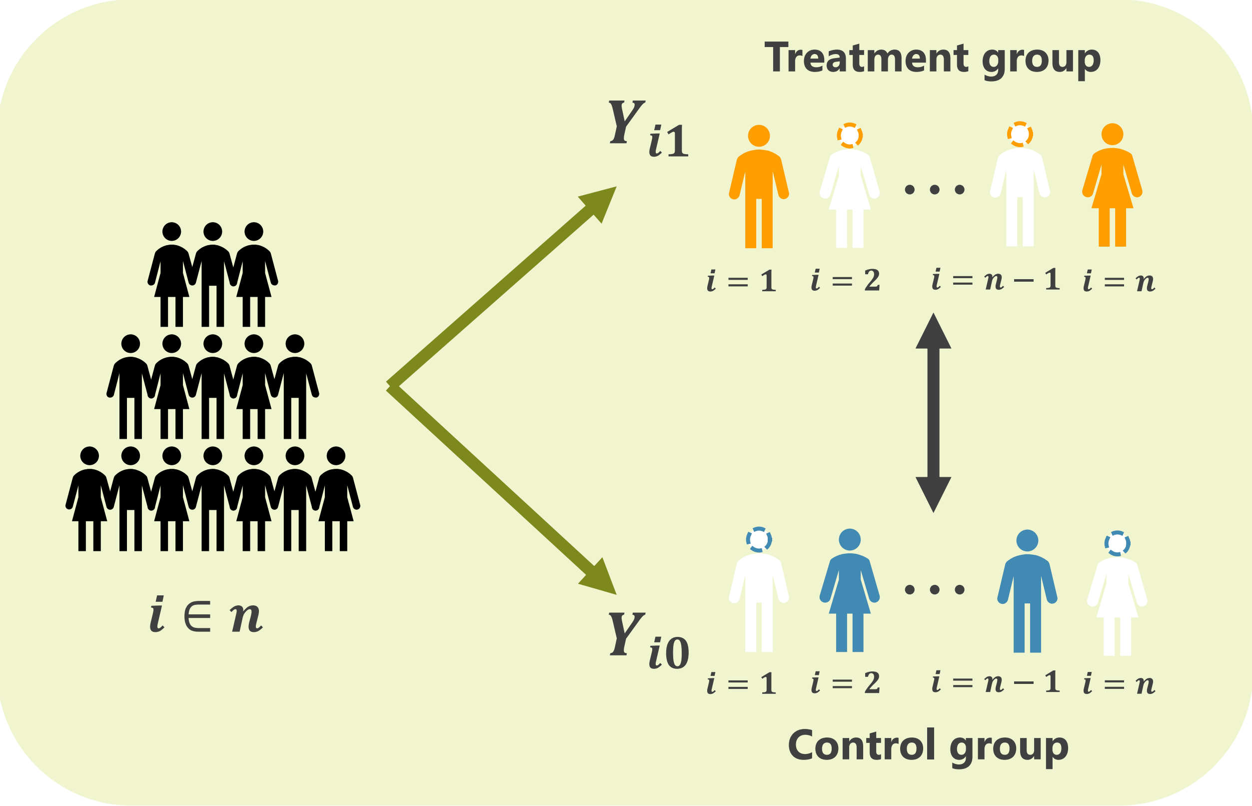
Individual-level Causal Effect



Population-level Causal Effect

- Individual-level causal effect cannot be directly observed ("fundamental problem of causal inference")
- RCM identifies a population-level causal effect in experimental and quasi-experimental studies

Rubin Causal Model (RCM) framework for causal inference



i	Group	Y_{i1}	Y_{i0}
1	Treatment	Observed	Missing
2	Control	Missing	Observed
⋮	⋮	⋮	⋮
n-1	Control	Missing	Observed
n	Treatment	Observed	Missing
Total / n		$\frac{1}{n}\sum_{i=1}^{n} Y_{i1}$	$\frac{1}{n}\sum_{i=1}^{n} Y_{i0}$

One of the potential outcomes is always missing since it is impossible to see both potential outcomes at once

Population-level Causal Effect

$$\text{ATE} = \frac{1}{n}\sum_{i=1}^{n} Y_{i1} - \frac{1}{n}\sum_{i=1}^{n} Y_{i0}$$

Y_i : e.g., Vehicle miles traveled (VMT)

ATE: Average Treatment Effect

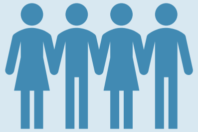
Quasi-experiment: Choice modeling in the Rubin Causal Model (RCM) framework

Treatment group (e.g., urban area)

$$Y_{i1} = x_i'\beta_1 + u_{i1}$$

Estimated $\widehat{\beta_1}$

Control group (e.g., rural area)



$$Y_{i0} = x_i'\beta_0 + u_{i0}$$

Estimated $\widehat{\beta_0}$

i	Group	$\mathrm{E}[Y_{i1}]$	$\mathrm{E}[Y_{i0}]$
1	Treatment	$x_1'\widehat{\beta_1}$	$x_1'\widehat{\beta_0}$
2	Control	$x_2'\widehat{\beta_1}$	$x_2'\widehat{\beta_0}$
⋮	⋮	⋮	⋮
n-1	Control	$x_{n-1}'\widehat{\beta_1}$	$x_{n-1}'\widehat{\beta_0}$
n	Treatment	$x_n'\widehat{\beta_1}$	$x_n'\widehat{\beta_0}$
Total / n		$\frac{1}{n}\sum_{i=1}^{n} x_i'\widehat{\beta_1}$	$\frac{1}{n}\sum_{i=1}^{n} x_i'\widehat{\beta_0}$

x_i : Explanatory variable
β : Parameter
u_i : Error term

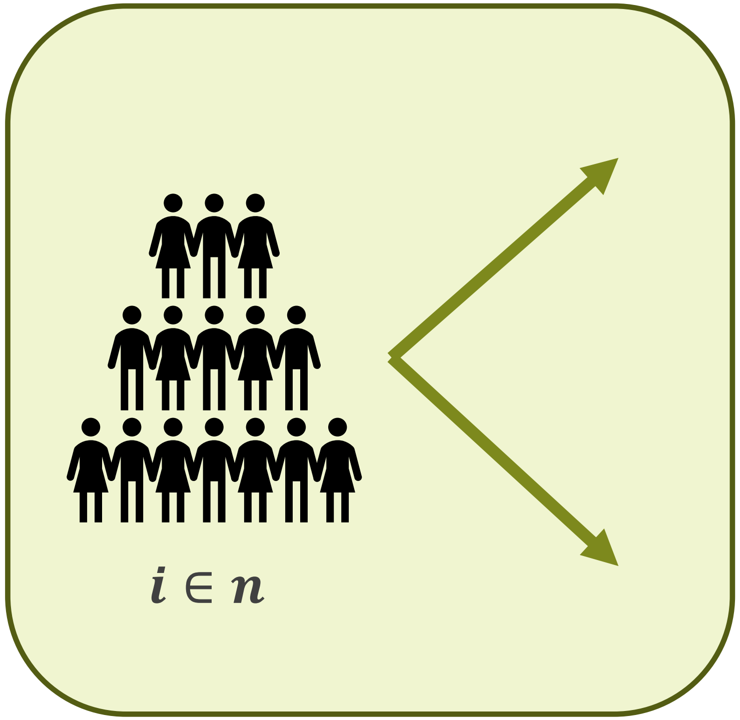
$x_i'\beta$ control the effect of socio-demographic attributes that cause the non-random missing

Population-level Causal Effect

$$\text{ATE} = \frac{1}{n}\sum_{i=1}^{n} x_i'\widehat{\beta_1} - \frac{1}{n}\sum_{i=1}^{n} x_i'\widehat{\beta_0}$$

Y_i : e.g., Vehicle miles traveled (VMT)

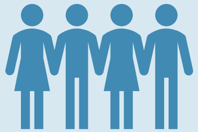
Quasi-experiment: Choice modeling in the Rubin Causal Model (RCM) framework



Treatment group (e.g., urban area)

$$Y_{i1} = x_i'\beta_1 + u_{i1}$$

Control group (e.g., rural area)



$$Y_{i0} = x_i'\beta_0 + u_{i0}$$

i	Group	Y_{i1}	Y_{i0}
1	Treatment	Observed	Missing
2	Control	Missing	Observed
⋮	⋮	⋮	⋮
n-1	Control	Missing	Observed
n	Treatment	Observed	Missing

Assumption of The Rubin Causal Model

People must be **randomly** assigned to treatment/control group conditional on x_i

otherwise →

Estimated $\hat{\beta} \neq$ True β

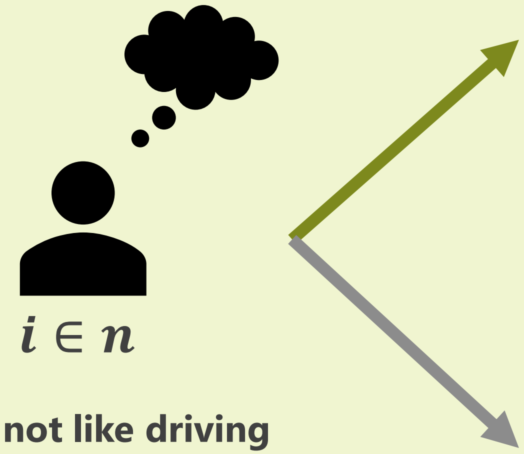
Biased estimates

Y_i : e.g., Vehicle miles traveled (VMT)

i.e., endogeneity due to residential self-selection

Residential self-selection (RSS) as missing data mechanism

subjective and attitudinal factors



$i \in n$

does not like driving

is more likely to use public transport

Treatment group (e.g., urban area)

$$Y_{i1} = x_i'\beta_1 + u_{i1}$$

Control group (e.g., rural area)

$$Y_{i0} = x_i'\beta_0 + u_{i0}$$

i	Group	Y_{i1}	Y_{i0}
1	Treatment	Observed	Missing
2	Control	Missing	Observed
⋮	⋮	⋮	⋮
n-1	Control	Missing	Observed
n	Treatment	Observed	Missing

- People choose a residential location while considering their future travel behaviors in the candidate locations
- Travel behavior outcomes are non-randomly missing due to subjective and attitudinal factors (Missing Not At Random, MNAR)
- Subjective and attitudinal variables are rarely observed → we cannot include these variables in x_i

Y_i : e.g., Vehicle miles traveled (VMT)

Sample selection model in the Rubin Causal Model (Heckman; 1976, 2003)

Residential choice model Z
(Binary endogenous switching)

$i \in n$

$Z_i = 1 \;\; \text{if } z_i^* > 0$

$z_i^* = w_i' \alpha + \varepsilon_i$

$Z_i = 0 \;\; \text{if } z_i^* \leq 0$

Travel behavior model Y

Treatment group (e.g., urban area)

$$Y_{i1} = x_i' \beta_1 + u_{i1}$$

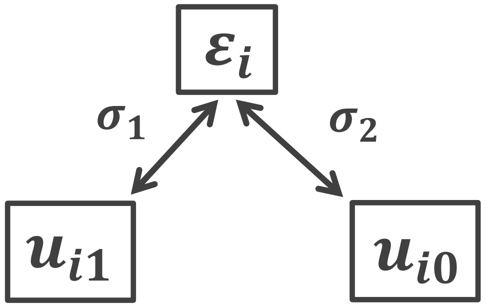
Control group (e.g., rural area)

$$Y_{i0} = x_i' \beta_0 + u_{i0}$$

The error structure

$$\begin{pmatrix} \varepsilon_i \\ u_{i1} \\ u_{i0} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_1 & \sigma_2 \\ \sigma_1 & v_1^2 & 0 \\ \sigma_2 & 0 & v_2^2 \end{pmatrix} \right]$$

Residential choice model Z



Travel behavior model Y

- Introduce the residential choice model with binary endogenous switching and estimate the error covariances σ
- The error structure describes the non-randomly assignment due to unobserved subjective and attitudinal factors

Y_i : e.g., Vehicle miles traveled (VMT)

Sample selection model with multinomial endogenous switching

Residential choice model Z (multinomial endogenous switching)

$i \in n$

$$z_{ij}^* = w_{ij}'\alpha_j + \varepsilon_{ij}$$

$$Z_i = j \quad \text{if } max(z_i^*) = z_{ij}^*$$

$$z_i^* = \{z_{i1}^*, z_{i2}^*, \dots, z_{ij}^*, \dots, z_{iJ}^*\}'$$

Travel behavior model Y

Treatment group

$$Y_{i1} = x_i'\beta_1 + u_{i1}$$

Control group 1

$$Y_{i2} = x_i'\beta_2 + u_{i2}$$

⋮

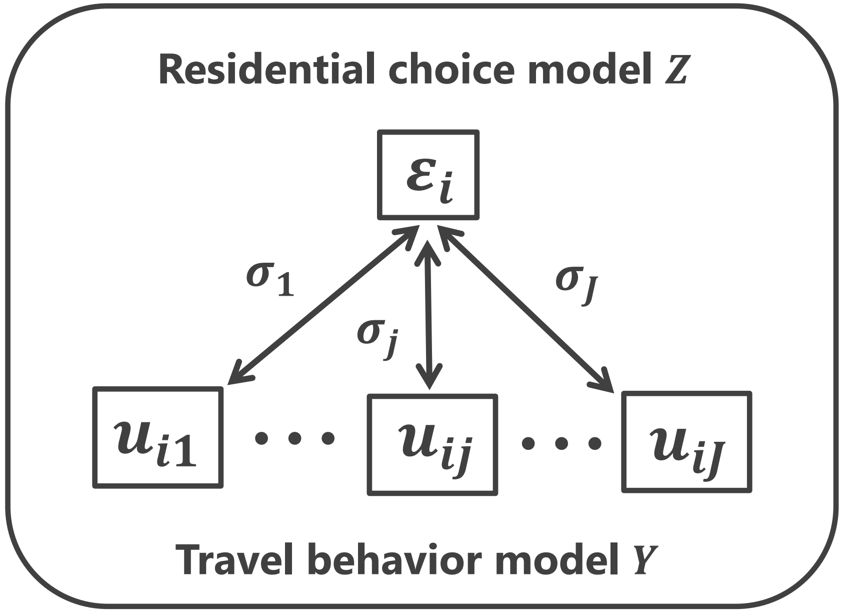
Control group J-1

$$Y_{iJ} = x_i'\beta_J + u_{iJ}$$

The error structure

$$\begin{pmatrix} \varepsilon_i \\ u_i \end{pmatrix} \sim N_{2J}\left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_Z & \Sigma_{Z,Y} \\ \Sigma_{Z,Y}^T & \Sigma_Y \end{pmatrix}\right]$$

$$\varepsilon_i = \{\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iJ}\}' \qquad u_i = \{u_{i1}, u_{i2}, \dots, u_{iJ}\}'$$



Sample selection model with multinomial endogenous switching

Travel behavior model Y

Treatment group

$$Y_{i1} = x_i'\beta_1 + u_{i1}$$

Estimated $\widehat{\beta_1}$

Control group 1

$$Y_{i2} = x_i'\beta_2 + u_{i2}$$

$\widehat{\beta_2}$

⋮

Control group J-1

$$Y_{iJ} = x_i'\beta_J + u_{iJ}$$

$\widehat{\beta_J}$

i	Group Z_i	Y_{i1}	Y_{i2}	Y_{i3}	⋯	Y_{iJ}
1	Treatment	Observed	Missing	Missing	⋯	Missing
2	Control 1	Missing	Observed	Missing	⋯	Missing
3	Control 2	Missing	Missing	Observed	⋯	Missing
⋮	⋮	⋮	⋮	⋮	⋯	⋮
n-1	Control J-1	Missing	Missing	Missing	⋯	Observed
n	Treatment	Observed	Missing	Missing	⋯	Missing

The proposed model:

Sample selection model with multinomial endogenous switching

Travel behavior model Y

Treatment group

$$Y_{i1} = x_i'\beta_1 + u_{i1}$$

Estimated $\widehat{\beta_1}$

Control group 1

$$Y_{i2} = x_i'\beta_2 + u_{i2}$$

$\widehat{\beta_2}$

⋮

Control group J-1

$$Y_{iJ} = x_i'\beta_J + u_{iJ}$$

$\widehat{\beta_J}$

i	Group Z_i	$\mathrm{E}[Y_{i1}]$	$\mathrm{E}[Y_{i2}]$	$\mathrm{E}[Y_{i3}]$	$\cdots$	$\mathrm{E}[Y_{iJ}]$
1	Treatment	$x_1'\widehat{\beta_1}$	$x_1'\widehat{\beta_2}$	$x_1'\widehat{\beta_3}$	$\cdots$	$x_1'\widehat{\beta_J}$
2	Control 1	$x_2'\widehat{\beta_1}$	$x_2'\widehat{\beta_2}$	$x_2'\widehat{\beta_3}$	$\cdots$	$x_2'\widehat{\beta_J}$
3	Control 2	$x_3'\widehat{\beta_1}$	$x_3'\widehat{\beta_2}$	$x_3'\widehat{\beta_3}$	$\cdots$	$x_3'\widehat{\beta_J}$
⋮	⋮	⋮	⋮	⋮	$\cdots$	⋮
n-1	Control J-1	$x_{n-1}'\widehat{\beta_1}$	$x_{n-1}'\widehat{\beta_2}$	$x_{n-1}'\widehat{\beta_3}$	$\cdots$	$x_{n-1}'\widehat{\beta_J}$
n	Treatment	$x_n'\widehat{\beta_1}$	$x_n'\widehat{\beta_2}$	$x_n'\widehat{\beta_3}$	$\cdots$	$x_n'\widehat{\beta_J}$
Total / n		$\frac{1}{n}\sum_{i=1}^{n} x_i'\widehat{\beta_1}$	$\frac{1}{n}\sum_{i=1}^{n} x_i'\widehat{\beta_2}$	$\frac{1}{n}\sum_{i=1}^{n} x_i'\widehat{\beta_3}$	$\cdots$	$\frac{1}{n}\sum_{i=1}^{n} x_i'\widehat{\beta_J}$

Residential choice model Z

$$i \in n$$

$$Z_i = j$$

$$z_{ij}^* = w_{ij}'\alpha_j + \varepsilon_{ij}$$

$$Z_i = j \quad \text{if } max(z_i^*) = z_{ij}^*$$

$$z_i^* = \{z_{i1}^*, z_{i2}^*, \dots, z_{ij}^*, \dots, z_{iJ}^*\}'$$

Travel behavior model Y

Treatment group

$$Y_{i1} = x_i'\beta_1 + u_{i1}$$

Control group j-1

$$Y_{ij} = x_i'\beta_j + u_{ij}$$

Control group J-1

$$Y_{iJ} = x_i'\beta_J + u_{iJ}$$

$$\begin{pmatrix} z_i^* \\ Y_i \end{pmatrix} \sim N_{2J}\left[\begin{pmatrix} W_i\alpha \\ X_i\beta \end{pmatrix}, \begin{pmatrix} \Sigma_Z & \Sigma_{Z,Y} \\ \Sigma_{Z,Y}^{\mathrm{T}} & \Sigma_Y \end{pmatrix}\right]$$

$$\begin{pmatrix} z_i^* \\ Y_{ij} \end{pmatrix} \sim N_{J+1}\left[\begin{pmatrix} W_i\alpha \\ x_{ij}'\beta_j \end{pmatrix}, \begin{pmatrix} \Sigma_Z & \Sigma_{Z,Y_j} \\ \Sigma_{Z,Y_j}^{\mathrm{T}} & \Sigma_{Y_j} \end{pmatrix}\right]$$

Missing travel behaviors Y_{-j} are marginalized

The likelihood

$$f(Z, Y|\theta) = \prod_{j\in J}\prod_{i:Z_i=j} f(Z_i = j, Y_{ij}|\theta)$$

θ: model parameter vector

The error structure and Bayesian estimation

The overall model structure

Residential choice model

$$Z^*_{ij} = w'_{ij}\alpha_j + \varepsilon_{ij}$$

Error term

$$\mathrm{Cov}(\varepsilon_{ij}, u_{ij}) \quad \sigma_j$$

$$Y_{ij} = x'_i\beta_j + u_{ij}$$

Travel behavior model Error term

$$\Sigma_Z = \begin{pmatrix} 1 & \gamma_{1,2} & \cdots & \gamma_{1,J-1} & \gamma_{1,J} \\ \gamma_{1,2} & 1 & \cdots & \gamma_{2,J-1} & \gamma_{2,J} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{1,J-1} & \gamma_{2,J-1} & \cdots & 1 & \gamma_{J-1,J} \\ \gamma_{1,J} & \gamma_{2,J} & \cdots & \gamma_{J-1,J} & 1 \end{pmatrix}$$

$$\Sigma_{Z,Y} = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_J \end{pmatrix}$$

$$\begin{pmatrix} z^*_i \\ Y_i \end{pmatrix} \sim N_{2J}\left[\begin{pmatrix} W_i\alpha \\ X_i\beta \end{pmatrix}, \begin{pmatrix} \Sigma_Z & \Sigma_{Z,Y} \\ \Sigma^{\mathrm{T}}_{Z,Y} & \Sigma_Y \end{pmatrix}\right]$$

$$\Sigma_Y = \begin{pmatrix} \Sigma_{Y_1} & 0 & \cdots & 0 \\ 0 & \Sigma_{Y_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_{Y_J} \end{pmatrix}$$

The error structure consists of:

- **multinomial probit model's variance-covariance matrix for allowing correlated alternatives**
- **diagonal matrix of covariance parameters σ for describing the non-randomness of the assignment[1)]**

1) i.e., Missing not at random (MNAR) of travel behavior outcomes

Differences from existing sample selection models

$$\begin{pmatrix} z_i^* \\ Y_i \end{pmatrix} \sim N_{2J}\left[\begin{pmatrix} W_i\alpha \\ X_i\beta \end{pmatrix}, \begin{pmatrix} \Sigma_Z & \Sigma_{Z,Y} \\ \Sigma_{Z,Y}^{\mathrm{T}} & \Sigma_Y \end{pmatrix}\right]$$

Lee (1983) and Spissu (2009)'s model

$$\Sigma_Z = \begin{pmatrix} \frac{\pi^2}{6\mu^2} & 0 & \cdots & 0 & 0 \\ 0 & \frac{\pi^2}{6\mu^2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{\pi^2}{6\mu^2} & 0 \\ 0 & 0 & \cdots & 0 & \frac{\pi^2}{6\mu^2} \end{pmatrix}$$

✓ Travel behavior outcome Y_i is only continuous

The proposed sample selection model

$$\Sigma_Z = \begin{pmatrix} 1 & \gamma_{1,2} & \cdots & \gamma_{1,J-1} & \gamma_{1,J} \\ \gamma_{1,2} & 1 & \cdots & \gamma_{2,J-1} & \gamma_{2,J} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{1,J-1} & \gamma_{2,J-1} & \cdots & 1 & \gamma_{J-1,J} \\ \gamma_{1,J} & \gamma_{2,J} & \cdots & \gamma_{J-1,J} & 1 \end{pmatrix}$$

This describes correlated alternatives for residential choice
➡ dealing with RSS more properly

✓ Travel behavior outcome Y_i is continuous / binary

Bayesian estimation framework

From Bayes' theorem, the posterior distribution is

$$f\left(\alpha, \beta, \Sigma_Z, \Sigma_{Z,Y}, \Sigma_Y \middle| Z, Y\right) \propto f(\theta) f(Z, Y|\theta),$$

θ : parameter vector (Prior: $f(\theta)$; Likelihood: $f(Z, Y|\theta)$)

where $f(Z, Y|\theta) = \prod_{j\in J} \prod_{i:Z_i=j} f\left(Z_i = j, Y_{ij}\middle|\theta\right).$

Evaluating the likelihood is computationally intensive

J+1 dimensional normal distribution

$$\begin{pmatrix} z_i^* \\ Y_{ij} \end{pmatrix} \sim N_{J+1}\left[\begin{pmatrix} W_i\alpha \\ x'_{ij}\beta_j \end{pmatrix}, \begin{pmatrix} \Sigma_Z & \Sigma_{Z,Y_j} \\ \Sigma_{Z,Y_j}^{\mathrm{T}} & \Sigma_{Y_j} \end{pmatrix}\right]$$

$$Z_i = j \quad \text{if } max(z_i^*) = z_{ij}^*$$

$$f(\alpha|Z, Y, \theta_{-\alpha}) \propto f(\alpha) f(Z|\theta),$$

$$f\left(\beta, \Sigma_{Z,Y} \middle| Z, Y, \theta_{-\beta,\Sigma_{Z,Y}}\right) \propto f(\beta, \Sigma_{Z,Y}) f(Y|Z, \theta),$$

$$f\left(\Sigma_Y \middle| Z, Y, \theta_{-\Sigma_Y}\right) \propto f(\Sigma_Y) f(Y|Z, \theta),$$

$$f\left(\Sigma_Z \middle| Z, Y, \theta_{-\Sigma_Z}\right) \propto f(\Sigma_Z) f(Z, Y|\theta),$$

where $f(Z|\theta) = \prod_{i\in n} f(z_i^*|\theta),$

J dimensional normal distribution

$$z_i^* \sim N_J[W_i\alpha, \Sigma_Z]$$

where $f(Y|Z, \theta) = \prod_{i\in n} f\left(Y_{ij}\middle|z_i^*, \theta\right),$

1 dimensional normal distribution

$$\left(Y_{ij}\middle|z_i^*\right) \sim N\left[x'_{ij}\beta_j + \Sigma_{Z,Y_j}^{\mathrm{T}} \Sigma_Z^{-1}(z_i^* - W_i\alpha), \nu_j^2\right]$$

Markov chain Monte Carlo (MCMC) algorithm

Step 1

Sample $z_i^* | [Y_{ij}, \theta]$ by data augmentation from

$$(z_i^* | Y_i, \theta) \sim N_J \left[\left(W_i\alpha + \Sigma_{Z,Y}\Sigma_Y^{-1}(Y_i - X_i\beta) \right), \left(\Sigma_Z - \Sigma_{Z,Y}\Sigma_Y^{-1}\Sigma_{Z,Y}^{T} \right) \right]$$

Sampling of latent utilities $z_i^* = (z_{i1}^*, z_{i2}^*, .., z_{iJ}^*)$

while fulfilling Z_i e.g., $max(z_i^*) = z_{ij}^*$ if $Z_i = j$

Step 2 Sample $\alpha | [z^*, Y, \theta_{-\alpha}]$ by Gibbs sampling

Step 3 Sample $\beta, \Sigma_{Z,Y} | [z^*, Y, \theta_{-\beta, \Sigma_{Z,Y}}]$ by Gibbs sampling

Step 4 Sample $\Sigma_Y | [z^*, Y, \theta_{-\Sigma_Y}]$ by Gibbs sampling

Step 5 Sample $\Sigma_Z | [z^*, Y, \theta_{-\Sigma_Z}]$ by Metropolis-Hastings

Back to Step 1 and repeat

The model structure

$$\begin{pmatrix} z_i^* \\ Y_i \end{pmatrix} \sim N_{2J} \left[\begin{pmatrix} W_i\alpha \\ X_i\beta \end{pmatrix}, \begin{pmatrix} \Sigma_Z & \Sigma_{Z,Y} \\ \Sigma_{Z,Y}^{T} & \Sigma_Y \end{pmatrix} \right]$$

➢ A tailored MCMC algorithm for efficient parameter estimation while allowing the complex error structure

Simulation study

Simulation①: Data generation

Step 1

Generating z_i^*, Y_i for residential choice and travel behavior ($i \in n = 3,000$)

z_i^*, Y_i follow the six-dimensional normal distribution

$$\begin{pmatrix} z_{i1}^* \\ z_{i2}^* \\ z_{i3}^* \\ Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{pmatrix} = N\left[\begin{pmatrix} \mathbf{w}'_{i1}\boldsymbol{\alpha}_1 \\ \mathbf{w}'_{i2}\boldsymbol{\alpha}_2 \\ \mathbf{w}'_{i3}\boldsymbol{\alpha}_3 \\ x'_i\beta_1 \\ x'_i\beta_2 \\ x'_i\beta_3 \end{pmatrix}, \begin{pmatrix} 1 & \gamma_{1,2} & \gamma_{1,3} & \sigma_1 & 0 & 0 \\ \gamma_{1,2} & 1 & \gamma_{2,3} & 0 & \sigma_2 & 0 \\ \gamma_{1,3} & \gamma_{2,3} & 1 & 0 & 0 & \sigma_3 \\ \sigma_1 & 0 & 0 & \Sigma_{Y_1} & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & \Sigma_{Y_2} & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 & \Sigma_{Y_3} \end{pmatrix}\right]$$

Settings on the error structure

correlation parameter $(\gamma_{1,2}, \gamma_{1,3}, \gamma_{2,3}) = (0, 0, 0)$

covariance parameter $(\sigma_1, \sigma_2, \sigma_3) = (0, 0.3, -0.3)$

Step 2

Discarding travel behavior outcomes based on residential choice

i	$max(z_i^*)$	Y_{i1}	Y_{i2}	Y_{i3}
1	z_{i1}^*	Observed	Missing	Missing
2	z_{i2}^*	Missing	Observed	Missing
⋮	⋮	⋮	⋮	⋮
2,999	z_{i3}^*	Missing	Missing	Observed
3,000	z_{i1}^*	Observed	Missing	Missing

✓ Y_{i1} is randomly missing and Y_{i2}, Y_{i3} are non-randomly missing because of $(\sigma_1, \sigma_2, \sigma_3) = (0, 0.3, -0.3)$

Simulation②: Estimated results

Parameters [True value]	Freely estimated σ		Fixed $\sigma = 0$	
	Estimates	t-value	Estimates	t -value
β_{10} [1.00]	0.94	16.96	0.95	16.60
β_{11} [0.25]	0.21	6.60	0.21	6.60
β_{12} [-0.50]	-0.46	-14.23	-0.45	-14.30
β_{20} [1.00]	0.89	9.87	1.17	21.33
β_{21} [-0.25]	-0.26	-7.91	-0.26	-7.88
β_{22} [-0.25]	-0.22	-6.82	-0.23	-7.13
β_{30} [1.00]	0.96	13.63	0.78	13.91
β_{31} [-0.25]	-0.25	-7.92	-0.25	-7.94
β_{32} [-0.50]	-0.49	-15.00	-0.49	-15.03
σ_1 [0.0]	0.07	1.07	Fixed to 0	
σ_2 [0.3]	0.37	3.84	Fixed to 0	
σ_3 [-0.3]	-0.35	-4.35	Fixed to 0	

i	$max(z_i^*)$	$\mathrm{E}[Y_{i1}]$	$\mathrm{E}[Y_{i2}]$	$\mathrm{E}[Y_{i3}]$
1	z_{i1}^*	$x_1'\widehat{\beta_1}$	$x_1'\widehat{\beta_2}$	$x_1'\widehat{\beta_3}$
2	z_{i2}^*	$x_2'\widehat{\beta_1}$	$x_2'\widehat{\beta_2}$	$x_2'\widehat{\beta_3}$
⋮	⋮	⋮	⋮	⋮
2,999	z_{i3}^*	$x_{2,999}'\widehat{\beta_1}$	$x_{2,999}'\widehat{\beta_2}$	$x_{2,999}'\widehat{\beta_3}$
3,000	z_{i1}^*	$x_{3,000}'\widehat{\beta_1}$	$x_{3,000}'\widehat{\beta_2}$	$x_{3,000}'\widehat{\beta_3}$
Total / n		$\frac{1}{n}\sum_{i=1}^{n} x_i'\widehat{\beta_1}$	$\frac{1}{n}\sum_{i=1}^{n} x_i'\widehat{\beta_2}$	$\frac{1}{n}\sum_{i=1}^{n} x_i'\widehat{\beta_3}$

ATE	True	Freely estimated σ	Fixed $\sigma = 0$
$\mathrm{E}[Y_1] - \mathrm{E}[Y_2]$	0.25	0.29	0.02
$\mathrm{E}[Y_1] - \mathrm{E}[Y_3]$	0.49	0.48	0.66

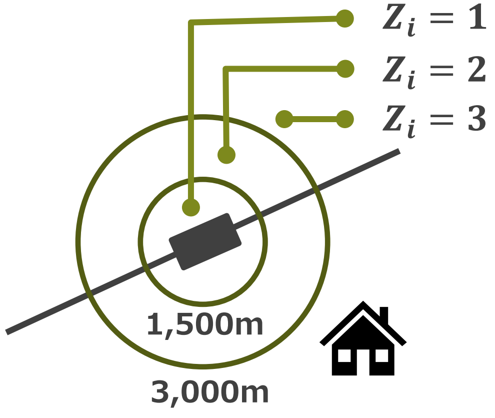
ATE: Average Treatment Effect

Case study

Case study①: Setting

Aim

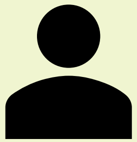
To examine the causal effect of relocation to around a train station on individual car ownership probability in Kumamoto city, Japan



Access to train station

Residential choice model Z
(multinomial endogenous switching)

$i \in n$



$z_{ij}^* = w_{ij}'\alpha_j + \varepsilon_{ij}$

$Z_i = j \quad \text{if } max(z_i^*) = z_{ij}^*$

Car ownership $Y \subset \{1, 0\}$

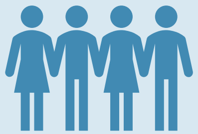
Treatment group

$y_{i1}^* = x_i'\beta_1 + u_{i1}$

$Y_{i1} = 1 \text{ if } y_{i1}^* > 0$

$Y_{i1} = 0 \text{ if } y_{i1}^* \leq 0$

Control group 1

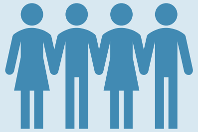


$y_{i2}^* = x_i'\beta_2 + u_{i2}$

$Y_{i2} = 1 \text{ if } y_{i2}^* > 0$

$Y_{i2} = 0 \text{ if } y_{i2}^* \leq 0$

Control group 2



$y_{i3}^* = x_i'\beta_3 + u_{i3}$

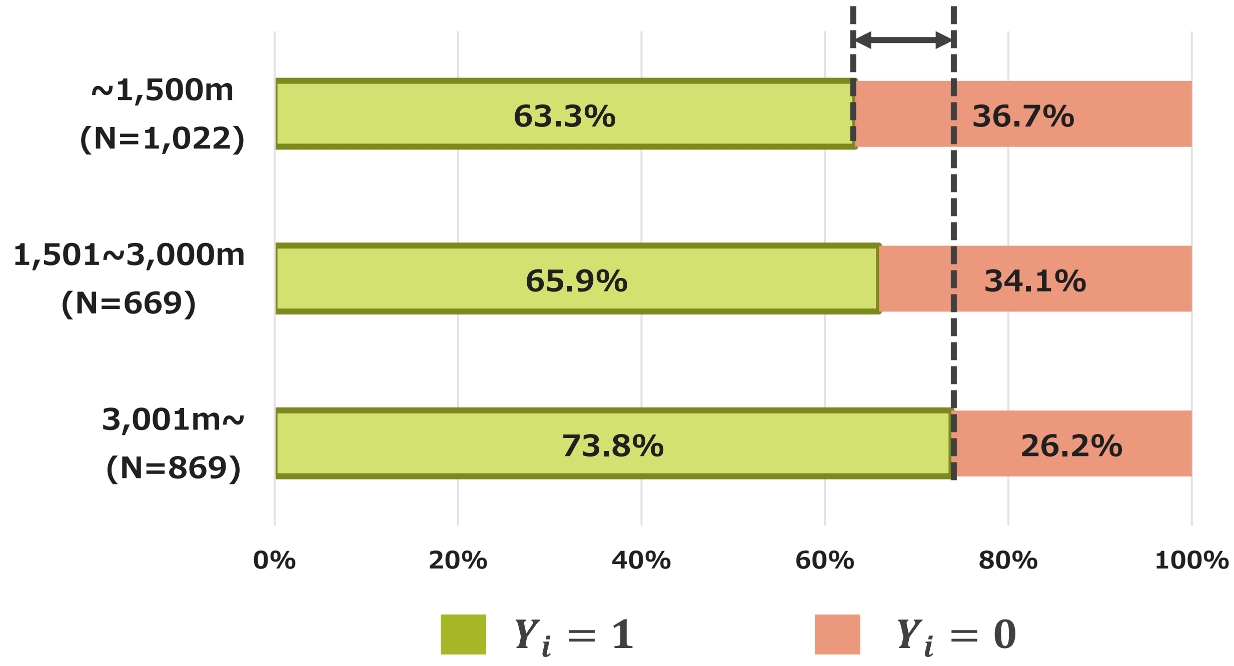
$Y_{i3} = 1 \text{ if } y_{i3}^* > 0$

$Y_{i3} = 0 \text{ if } y_{i3}^* \leq 0$

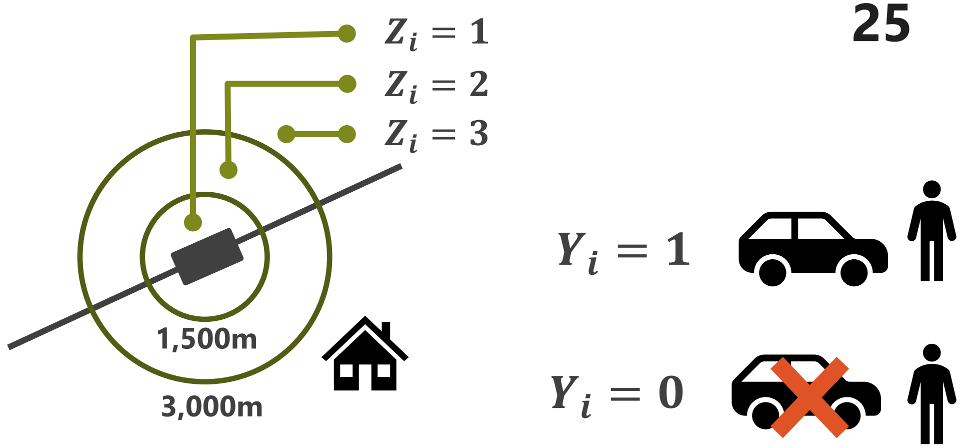
i	$max(z_i^*)$	Y_{i1}	Y_{i2}	Y_{i3}
1	z_1^*	Observed	Missing	Missing
2	z_2^*	Missing	Observed	Missing
⋮	⋮	⋮	⋮	⋮
n-1	z_3^*	Missing	Missing	Observed
n	z_1^*	Observed	Missing	Missing

Case study②: Data

- The 2012 household travel survey in Kumamoto City, Japan
- Respondents: 2,560 householders over 17 years old



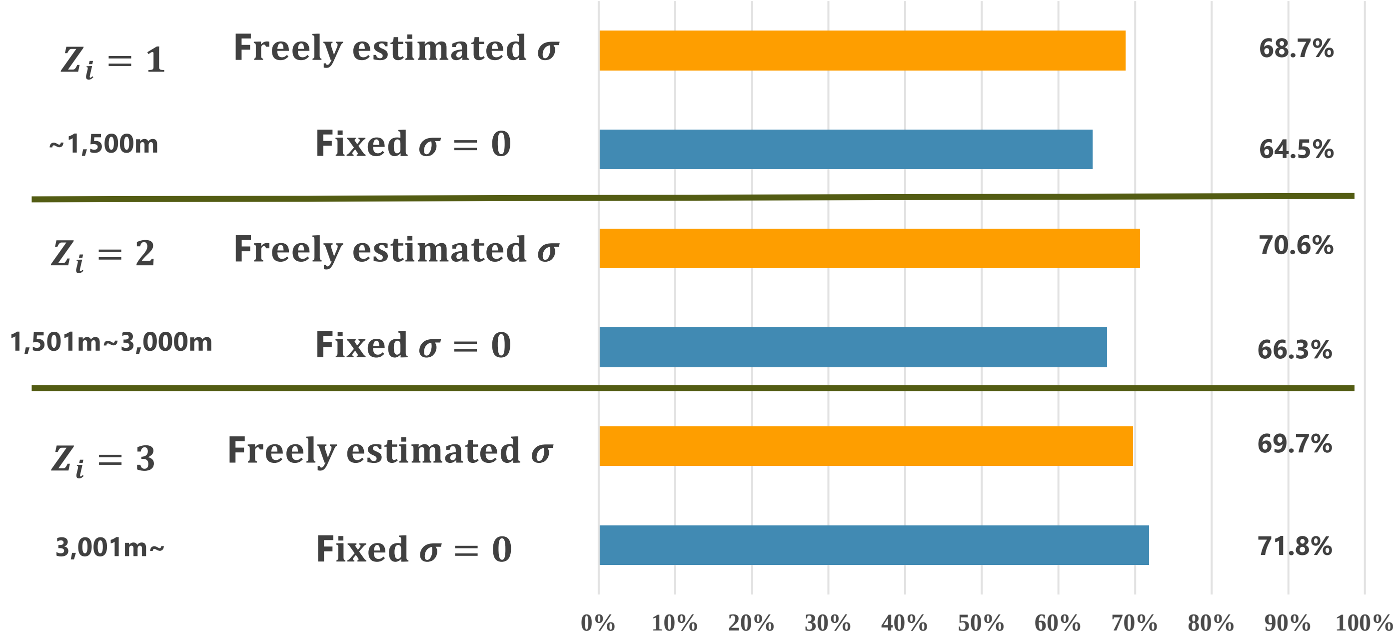
✓ 10.5% point difference between people living within 1,500m and over 3,000m from the nearest station



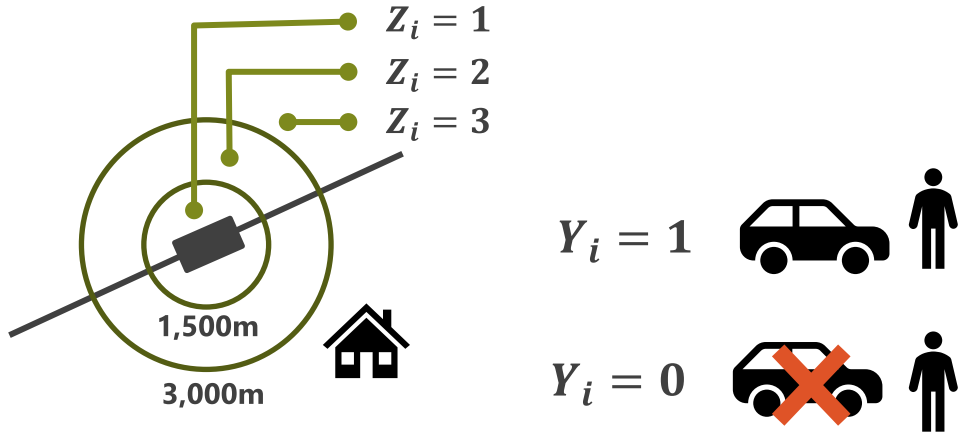
Access to train station

i	$max(z_i^*)$	Y_{i1}	Y_{i2}	Y_{i3}
1	z_1^*	Observed	Missing	Missing
2	z_2^*	Missing	Observed	Missing
⋮	⋮	⋮	⋮	
2,559	z_3^*	Missing	Missing	Observed
2,560	z_1^*	Observed	Missing	Missing

Case study③: Result



Average of the expected car ownership probability $\mathrm{E}\left[P\left(Y_j = 1\right)\right]$



Access to train station

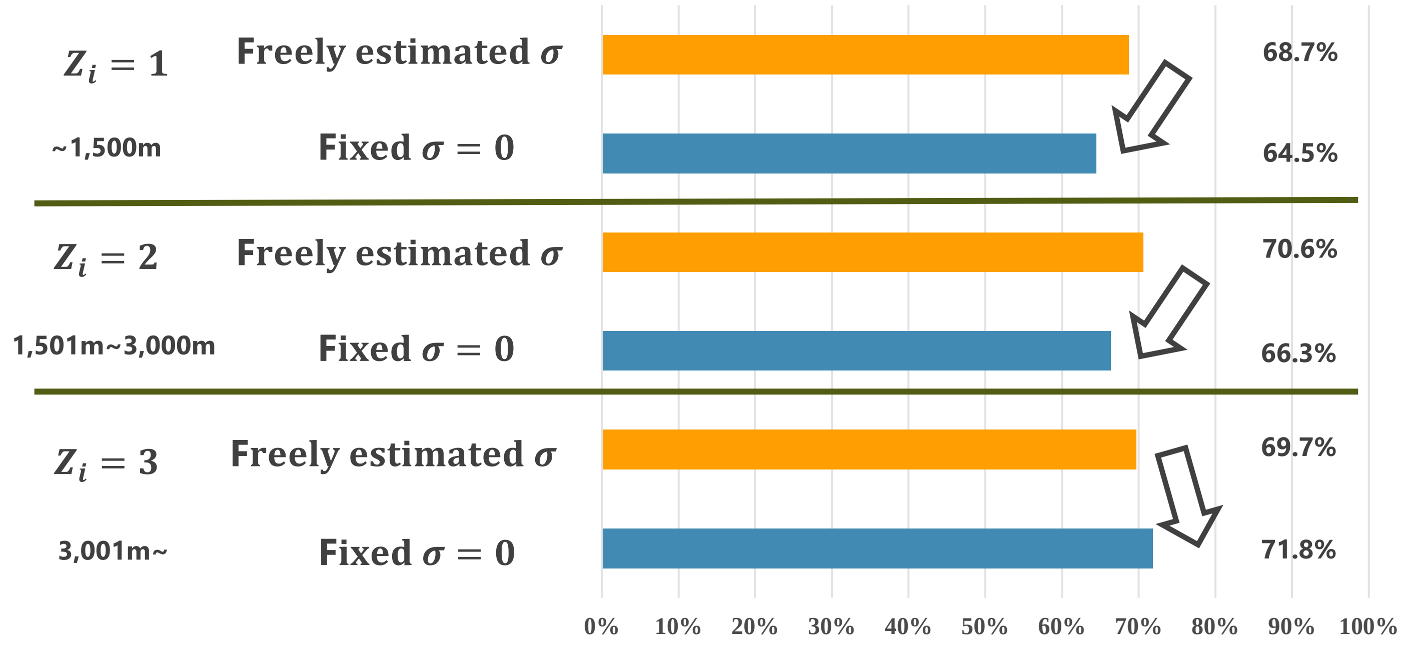
ATE	Freely estimated σ	Fixed $\sigma = 0$
$\mathrm{E}[P(Y_1 = 1)] - \mathrm{E}[P(Y_2 = 1)]$	-2.0%	-2.0%
$\mathrm{E}[P(Y_1 = 1)] - \mathrm{E}[P(Y_3 = 1)]$	-1.0%	-7.0%

Freely estimated σ : dealing with non-randomly missing car ownership outcomes (addressing endogeneity due to RSS)

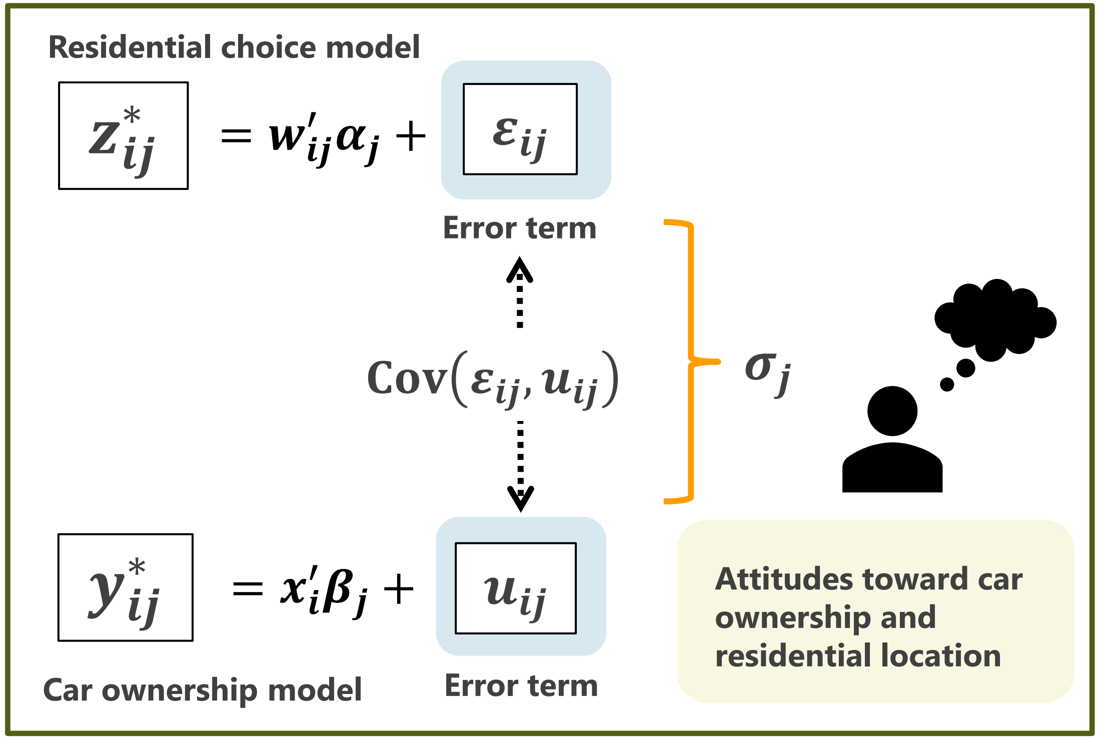
Fixed $\sigma = 0$: assuming randomly missing car ownership outcomes (ignoring endogeneity due to RSS)

✓ Assuming the random assignment can lead to the false conclusion that relocation from over 3,000m to within 1,500m from the nearest train station can reduce their car ownership levels

Case study④: Discussion



Average of the expected car ownership probability $\mathrm{E}[P(Y_j = 1)]$



- The RSS effect could occur due to following unobserved travel-related attitudes
  - Attitudes toward using public transportation and living near a train station
  - Attitudes toward owning a car and living in suburban areas far from a train station

The degree of RSS can be of interest to researchers and practitioners in urban planning

Conclusion

- We proposed an extended sample selection model to identify a causal effect (ATE) of residential neighborhoods on travel behavior in the Rubin Causal Model framework
- The proposed sample selection model describes the non-randomly missing data mechanism of travel behavior outcomes, namely, residential self-selection (RSS)
- The analysis in Kumamoto city revealed that relocation around a station could not reduce their car ownership levels
- Unobserved subjective and attitudinal factors can cause the non-random assignment (i.e., endogeneity due to residential self-selection), leading to a false conclusion
- The degree of RSS can be of interest to researchers and practitioners in urban planning

Thank you!

for your attention

Hajime Watanabe, The University of Tokyo

hwatanabe@bin.t.u-tokyo.ac.jp

The error structure of the proposed model

Residential choice model

$$z_{ij}^{*} = w_{ij}'\alpha_j + \varepsilon_{ij}$$

Error term

$$\mathrm{Cov}(\varepsilon_{ij}, u_{ij}) \quad \sigma_j$$

$$Y_{ij} = x_i'\beta_j + u_{ij}$$

Travel behavior model

Error term

$$\Sigma_Z = \begin{pmatrix} 1 & \gamma_{1,2} & \cdots & \gamma_{1,J-1} & \gamma_{1,J} \\ \gamma_{1,2} & 1 & \cdots & \gamma_{2,J-1} & \gamma_{2,J} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{1,J-1} & \gamma_{2,J-1} & \cdots & 1 & \gamma_{J-1,J} \\ \gamma_{1,J} & \gamma_{2,J} & \cdots & \gamma_{J-1,J} & 1 \end{pmatrix}$$

$$\Sigma_{Z,Y} = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_J \end{pmatrix}$$

$$\begin{pmatrix} z_i^* \\ Y_i \end{pmatrix} \sim N_{2J}\left[\begin{pmatrix} W_i\alpha \\ X_i\beta \end{pmatrix}, \begin{pmatrix} \Sigma_Z & \Sigma_{Z,Y} \\ \Sigma_{Z,Y}^{T} & \Sigma_Y \end{pmatrix}\right]$$

$$\Sigma_Y = \begin{pmatrix} \Sigma_{Y_1} & 0 & \cdots & 0 \\ 0 & \Sigma_{Y_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_{Y_J} \end{pmatrix}$$

$$(z_i^*|Y_i,\theta) \sim N_J\left[\left(W_i\alpha + \Sigma_{Z,Y}\,\Sigma_Y^{-1}(Y_i - X_i\beta)\right), \left(\Sigma_Z - \Sigma_{Z,Y}\,\Sigma_Y^{-1}\Sigma_{Z,Y}^{T}\right)\right],$$

if $Y_i \neq Y_i'$, $\quad P\left(max(z_i^*) = z_{ij}^*\middle|Y_i\right) \neq P\left(max(z_i^*) = z_{ij}^*\middle|Y_i'\right).$

Travel behavior outcomes are missing not at random (MNAR)

Error structure and the full conditional distribution of σ

$$\begin{pmatrix} z_i^* \\ y_i^* \end{pmatrix} \sim N\left[\begin{pmatrix} w_i'\alpha \\ x_i'\beta \end{pmatrix}, \begin{pmatrix} 1 & \sigma \\ \sigma & 1+\sigma^2 \end{pmatrix}\right]$$

$$(y_i^*|z_i^*) \sim N[\underline{x_i'\beta + \sigma(z_i^* - w_i'\alpha)}, 1]$$

$$\underbrace{\{x_i', (z_i^* - w_i'\alpha)\}}_{\bar{x}_i'} \cdot \begin{pmatrix} \beta \\ \sigma \end{pmatrix}$$

Conditional distribution of parameter β, σ is

$$\beta, \sigma|[y^*, z^*, \alpha] \sim N[g, G]$$

$$\beta, \sigma|[y^*, z^*, \alpha] \sim N\left[G\left(G_0^{-1} g_0 + \sum_{i \in n} \bar{x}_i y_i^*\right), \left(G_0^{-1} + \sum_{i \in n} \bar{x}_i \bar{x}_i'\right)^{-1}\right]$$

g_0 : Prior mean of $f(\beta, \sigma)$
G_0 : Prior variance of $f(\beta, \sigma)$
$\bar{x}_i' = \{x_i', (z_i^* - w_i'\alpha)\}$

The conditional distribution does not include parameter β, σ (Full conditional distribution)

We can use **Gibbs sampling** to approximate the posterior distribution of β, σ

Error structure and the full conditional distribution of σ

$$\begin{pmatrix} z_i^* \\ y_i^* \end{pmatrix} \sim N\left[\begin{pmatrix} w_i'\alpha \\ x_i'\beta \end{pmatrix}, \begin{pmatrix} 1 & \sigma \\ \sigma & 1 \end{pmatrix}\right]$$

$$(y_i^* | z_i^*) \sim N\left[x_i'\beta + \sigma(z_i^* - w_i'\alpha), \underline{1 - \sigma^2}\right]$$

Conditional distribution of parameter β, σ is

$$\beta, \sigma | [y^*, z^*, \alpha] \sim N[g, G]$$

$$\beta, \sigma | [y^*, z^*, \alpha] \sim N\left[G\left(G_0^{-1} g_0 + [1 - \sigma^2]^{-1} \sum_{i \in n} \bar{x}_i y_i^*\right), \left(G_0^{-1} + [1 - \sigma^2]^{-1} \sum_{i \in n} \bar{x}_i \bar{x}_i'\right)^{-1}\right]$$

g_0 : Prior mean of $f(\beta, \sigma)$
G_0 : Prior variance of $f(\beta, \sigma)$
$\bar{x}_i' = \{x_i', (z_i^* - w_i'\alpha)\}$

The conditional distribution includes parameter σ

The full conditional distribution cannot be derived and need to accept−reject sampling (e.g., Metropolis − Hastings)

→ inefficient

The error structure of the proposed sample selection model

$$\begin{pmatrix} z_i^* \\ Y_{ij} \end{pmatrix} \sim N_{J+1}\left[\begin{pmatrix} W_i\alpha \\ x'_{ij}\beta_j \end{pmatrix}, \begin{pmatrix} \Sigma_Z & \Sigma_{Z,Y_j} \\ \Sigma_{Z,Y_j}^{\mathrm{T}} & \Sigma_{Y_j} \end{pmatrix}\right]$$

$$\Sigma_{Y_j} = \nu_j^2 + \Sigma_{Z,Y_j}^{\mathrm{T}} \Sigma_Z^{-1} \Sigma_{Z,Y_j}$$

$$\left(Y_{ij} | z_i^*\right) \sim N\left[x'_{ij}\beta_j + \Sigma_{Z,Y_j}^{\mathrm{T}} \Sigma_Z^{-1}(z_i^* - W_i\alpha), \nu_j^2\right]$$

g_{0j} : Prior mean of $f\left(\beta, \Sigma_{Z,Y_j}\right)$

G_{0j} : Prior variance of $f\left(\beta, \Sigma_{Z,Y_j}\right)$

$V'_j = \left\{x'_{ij}, (z_i^* - W_i\alpha)\right\}$

$$\beta, \Sigma_{Z,Y_j} | \left[Y_j, z^*, \alpha, \Sigma_Z, \Sigma_{Y_j}\right] \sim N[g_j, G_j]$$

$$\text{where } G_j = \left(G_{oj}^{-1} + \nu_j^{-2} V'_j V_j\right)^{-1}, g_j = G_j\left(G_{0j}^{-1} g_{0j} + \nu_j^{-2} V'_j Y_j\right),$$

The conditional distribution does not include parameter β, Σ_{Z,Y_j} (Full conditional distribution)