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

Hajime Watanabe\textsuperscript{a)} and Takuya Maruyama\textsuperscript{b)}

\textsuperscript{a)} The University of Tokyo \quad \textsuperscript{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 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

\[ \sigma_i = \frac{1}{n} \sum_{i=1}^{n} Y_{i1} - \frac{1}{n} \sum_{i=1}^{n} Y_{i0} \]

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

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

\( \text{ATE}: \text{Average Treatment Effect} \)
Quasi-experiment: Choice modeling in the Rubin Causal Model (RCM) framework

<table>
<thead>
<tr>
<th>i</th>
<th>Group</th>
<th>( E[Y_{i1}] )</th>
<th>( E[Y_{i0}] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Treatment</td>
<td>( x_{i1} \beta_1 )</td>
<td>( x_{i1} \beta_0 )</td>
</tr>
<tr>
<td>2</td>
<td>Control</td>
<td>( x_{i2} \beta_1 )</td>
<td>( x_{i2} \beta_0 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>n-1</td>
<td>Control</td>
<td>( x_{n-1} \beta_1 )</td>
<td>( x_{n-1} \beta_0 )</td>
</tr>
<tr>
<td>n</td>
<td>Treatment</td>
<td>( x_n \beta_1 )</td>
<td>( x_n \beta_0 )</td>
</tr>
</tbody>
</table>

Total / n

\[
\frac{1}{n} \sum_{i=1}^{n} x_i \beta_1 \quad \frac{1}{n} \sum_{i=1}^{n} x_i \beta_0
\]

Population-level Causal Effect

\[
ATE = \frac{1}{n} \sum_{i=1}^{n} x_i \beta_1 - \frac{1}{n} \sum_{i=1}^{n} x_i \beta_0
\]
Quasi-experiment: Choice modeling in the Rubin Causal Model (RCM) framework

\[ Y_{i1} = x_i' \beta_1 + u_{i1} \]

\[ Y_{i0} = x_i' \beta_0 + u_{i0} \]

Assumption of The Rubin Causal Model

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

<table>
<thead>
<tr>
<th>( i )</th>
<th>Group</th>
<th>( Y_{i1} )</th>
<th>( Y_{i0} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Treatment</td>
<td>Observed</td>
<td>Missing</td>
</tr>
<tr>
<td>2</td>
<td>Control</td>
<td>Missing</td>
<td>Observed</td>
</tr>
<tr>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>n-1</td>
<td>Control</td>
<td>Missing</td>
<td>Observed</td>
</tr>
<tr>
<td>n</td>
<td>Treatment</td>
<td>Observed</td>
<td>Missing</td>
</tr>
</tbody>
</table>

**otherwise**

\[ \hat{\beta} \neq \beta \]  
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

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)

\[
Y_{i1} = x'_i \beta_1 + u_{i1} \\
Y_{i0} = x'_i \beta_0 + u_{i0}
\]

**Table:**

<table>
<thead>
<tr>
<th>$i$</th>
<th>Group</th>
<th>$Y_{i1}$</th>
<th>$Y_{i0}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Treatment</td>
<td>Observed</td>
<td>Missing</td>
</tr>
<tr>
<td>2</td>
<td>Control</td>
<td>Missing</td>
<td>Observed</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>n-1</td>
<td>Control</td>
<td>Missing</td>
<td>Observed</td>
</tr>
<tr>
<td>n</td>
<td>Treatment</td>
<td>Observed</td>
<td>Missing</td>
</tr>
</tbody>
</table>

**Notes:**
- $i \in n$ does not like driving.
- Is more likely to use public transport.
- Residential self-selection (RSS) is used as a missing data mechanism.
Sample selection model in the Rubin Causal Model (Heckman; 1976, 2003)

Residential choice model $Z$

(Binary endogenous switching)

- $i \in n$
  - $Z_i = 1$ if $z_i^* > 0$
  - $z_i^* = w_i' \alpha + \varepsilon_i$
  - $Z_i = 0$ 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

- $Y_i : e.g.,$ Vehicle miles traveled (VMT)
- Introduce the residential choice model with binary endogenous switching and estimate the error covariances $\sigma$
- The error structure describes the non-randomly assignment due to unobserved subjective and attitudinal factors
The proposed model:

Sample selection model with multinomial endogenous switching

### Residential choice model $Z$
(multinomial endogenous switching)

- $i \in n$
- $i \in n$
- $Z_{ij}^* = w_{ij}'\alpha_j + \varepsilon_{ij}$
- $Z_i = j \text{ if } \max(z_i^*) = z_{ij}^*$
- $z_i^* = \{z_{i1}^*, z_{i2}^*, ..., z_{ij}^*, ..., z_{ij}^*\}'$

### 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

$$
\begin{pmatrix}
\varepsilon_i \\
u_i
\end{pmatrix} \sim \mathcal{N}_{2J}
\begin{pmatrix}
0 \\
\Sigma_Z \\
\Sigma_{Z,Y} \\
\Sigma_Y
\end{pmatrix}
$$

$\varepsilon_i = \{\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{ij}\}'$

$u_i = \{u_{i1}, u_{i2}, ..., u_{ij}\}'$

### Residential choice model $Z$

- $\varepsilon_i$
- $\sigma_i$
- $\sigma_j$

### Travel behavior model $Y$

- $u_{i1}$
- $\cdots$
- $u_{ij}$
- $\cdots$
- $u_{ij}$
The proposed model:

### Sample selection model with multinomial endogenous switching

#### 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 $J-1$**

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

<table>
<thead>
<tr>
<th>$i$</th>
<th>Group $Z_i$</th>
<th>$Y_{i1}$</th>
<th>$Y_{i2}$</th>
<th>$Y_{i3}$</th>
<th>$\cdots$</th>
<th>$Y_{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Treatment</td>
<td>Observed</td>
<td>Missing</td>
<td>Missing</td>
<td>$\cdots$</td>
<td>Missing</td>
</tr>
<tr>
<td>2</td>
<td>Control 1</td>
<td>Missing</td>
<td>Observed</td>
<td>Missing</td>
<td>$\cdots$</td>
<td>Missing</td>
</tr>
<tr>
<td>3</td>
<td>Control 2</td>
<td>Missing</td>
<td>Missing</td>
<td>Observed</td>
<td>$\cdots$</td>
<td>Missing</td>
</tr>
<tr>
<td>$\cdots$</td>
<td>$\cdots$</td>
<td>$\cdots$</td>
<td>$\cdots$</td>
<td>$\cdots$</td>
<td>$\cdots$</td>
<td>$\cdots$</td>
</tr>
<tr>
<td>$n-1$</td>
<td>Control $J-1$</td>
<td>Missing</td>
<td>Missing</td>
<td>Missing</td>
<td>$\cdots$</td>
<td>Observed</td>
</tr>
<tr>
<td>$n$</td>
<td>Treatment</td>
<td>Observed</td>
<td>Missing</td>
<td>Missing</td>
<td>$\cdots$</td>
<td>Missing</td>
</tr>
</tbody>
</table>

- Estimated $\hat{\beta}_1$
- Estimated $\hat{\beta}_2$
- Estimated $\hat{\beta}_j$
The proposed model:

Sample selection model with multinomial endogenous switching

<table>
<thead>
<tr>
<th>$i$</th>
<th>Group $Z_i$</th>
<th>$E[Y_{i1}]$</th>
<th>$E[Y_{i2}]$</th>
<th>$E[Y_{i3}]$</th>
<th>$\cdots$</th>
<th>$E[Y_{ij}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Treatment</td>
<td>$x'_1\beta_1$</td>
<td>$x'_2\beta_2$</td>
<td>$x'_3\beta_3$</td>
<td>$\cdots$</td>
<td>$x'_1\beta_J$</td>
</tr>
<tr>
<td>2</td>
<td>Control 1</td>
<td>$x'_2\beta_1$</td>
<td>$x'_2\beta_2$</td>
<td>$x'_2\beta_3$</td>
<td>$\cdots$</td>
<td>$x'_2\beta_J$</td>
</tr>
<tr>
<td>3</td>
<td>Control 2</td>
<td>$x'_3\beta_1$</td>
<td>$x'_3\beta_2$</td>
<td>$x'_3\beta_3$</td>
<td>$\cdots$</td>
<td>$x'_3\beta_J$</td>
</tr>
<tr>
<td></td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\cdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>n-1</td>
<td>Control J-1</td>
<td>$x'_{n-1}\beta_1$</td>
<td>$x'_{n-1}\beta_2$</td>
<td>$x'_{n-1}\beta_3$</td>
<td>$\cdots$</td>
<td>$x'_{n-1}\beta_J$</td>
</tr>
<tr>
<td>n</td>
<td>Treatment</td>
<td>$x'_n\beta_1$</td>
<td>$x'_n\beta_2$</td>
<td>$x'_n\beta_3$</td>
<td>$\cdots$</td>
<td>$x'_n\beta_J$</td>
</tr>
<tr>
<td>Total / n</td>
<td>$\frac{1}{n}\sum_{i=1}^{n} x'_1\beta_1$</td>
<td>$\frac{1}{n}\sum_{i=1}^{n} x'_2\beta_2$</td>
<td>$\frac{1}{n}\sum_{i=1}^{n} x'_3\beta_3$</td>
<td>$\cdots$</td>
<td>$\frac{1}{n}\sum_{i=1}^{n} x'_J\beta_J$</td>
<td></td>
</tr>
</tbody>
</table>

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**
  \[ Y_{ij} = x'_i\beta_J + u_{ij} \]

- **Control group J**
  \[ Y_{iJ} = x'_i\beta_J + u_{iJ} \]
Residential choice model $Z$

- $i \in n$
- $Z_i = j$
- $z_{ij}^* = w^j_i \alpha_j + \varepsilon_{ij}$
- $Z_i = j$ if max($z_{ij}^*$) = $z_{ij}^*$

$z_i = \{z_{i1}, z_{i2}, ..., z_{ij}, ..., z_{ij}\}'$

Travel behavior model $Y$

- Treatment group
  - $Y_{i1} = x_i' \beta_1 + u_{i1}$
  - $(Z_i^* Y_i) \sim \mathcal{N}_{2J} \left( \begin{bmatrix} W_i \alpha \\ X_i \beta \end{bmatrix}, \begin{bmatrix} \Sigma_Z & \Sigma_{Z,Y} \\ \Sigma_{Z,Y}^T & \Sigma_Y \end{bmatrix} \right)$
- Control group $j-1$
  - $Y_{ij} = x_i' \beta_j + u_{ij}$
- Control group $J-1$
  - $Y_{ij} = x_i' \beta_j + u_{ij}$

- Missing travel behaviors $Y_{-j}$ are marginalized

The likelihood

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

$\theta$: model parameter vector
The error structure and Bayesian estimation
The overall model structure

The error structure consists of:
- multinomial probit model’s variance-covariance matrix for allowing correlated alternatives
- diagonal matrix of covariance parameters \( \sigma \) for describing the non-randomness of the assignment

1) i.e., Missing not at random (MNAR) of travel behavior outcomes
Differences from existing sample selection models

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

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

\[
f(\alpha, \beta, \Sigma_Z, \Sigma_{Z,Y}, \Sigma_Y | Z, Y) \propto f(\theta)f(Z, Y | \theta),
\]

where

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

Evaluating the likelihood is computationally intensive.

\[
\begin{align*}
& f(\alpha | Z, Y, \theta_{-\alpha}) \propto f(\alpha)f(Z | \theta), \\
& f(\beta, \Sigma_{Z,Y} | Z, Y, \theta_{-\beta,\Sigma_{Z,Y}}) \propto f(\beta, \Sigma_{Z,Y})f(Y | Z, \theta), \\
& f(\Sigma_Y | Z, Y, \theta_{-\Sigma_Y}) \propto f(\Sigma_Y)f(Y | Z, \theta), \\
& f(\Sigma_Z | Z, Y, \theta_{-\Sigma_Z}) \propto f(\Sigma_Z)f(Z, Y | \theta),
\end{align*}
\]

where

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

\[
J \text{ dimensional normal distribution}
\]

\[
z_i^* \sim N_j [W_i \alpha, \Sigma_Z]
\]

\[
(\Sigma_Y)^{-1} X_{ij} \beta_j \sim [x_{ij}' \beta_j, \Sigma_{Y,i}]
\]

where

\[
f(Y | Z, \theta) = \prod_{i \in n} f(Y_{ij} | z_i^*, \theta),
\]

\[
1 \text{ dimensional normal distribution}
\]

\[
(Y_{ij} | z_i^*) \sim N [x_{ij}' \beta_j + \Sigma_{Z,Y_j}^{-1} (z_i^* - W_i \alpha), \nu_j^2]
\]
Markov chain Monte Carlo (MCMC) algorithm

Step 1
Sample $z_i^* | [Y_{ij}, \theta]$ by data augmentation from
$$(z_i^* | Y_{ij}, \theta) \sim N \left[ \left( W_i \alpha + \Sigma_{Z,Y} \Sigma_Y^{-1} (Y_i - X_i \beta) \right), \left( \Sigma_Z - \Sigma_{Z,Y} \Sigma_Y^{-1} \Sigma_Y^T \right) \right]$$

Sampling of latent utilities

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

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

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

Back to Step 1 and repeat
Simulation study
Simulation①: Data generation

Step 1

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

$$
\begin{bmatrix}
    z_{i1}^* \\
    z_{i2}^* \\
    z_{i3}^* \\
    Y_{i1} \\
    Y_{i2} \\
    Y_{i3}
\end{bmatrix} = N

\begin{bmatrix}
    w_{i1}^* \alpha_1 \\
    w_{i2}^* \alpha_2 \\
    w_{i3}^* \alpha_3 \\
    x_i^* \beta_1 \\
    x_i^* \beta_2 \\
    x_i^* \beta_3
\end{bmatrix}

\begin{pmatrix}
    1 & y_{1,2} & y_{1,3} & \sigma_1 & 0 & 0 \\
    y_{1,2} & 1 & y_{2,3} & 0 & \sigma_2 & 0 \\
    y_{1,3} & y_{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}

$$

Settings on the error structure

- correlation parameter
  $$
  (y_{1,2}, y_{1,3}, y_{2,3}) = (0, 0, 0)
  $$

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

Step 2

Discarding travel behavior outcomes based on residential choice

<table>
<thead>
<tr>
<th>$i$</th>
<th>$\text{max}(z_i^*)$</th>
<th>$Y_{i1}$</th>
<th>$Y_{i2}$</th>
<th>$Y_{i3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$z_{i1}^*$</td>
<td>Observed</td>
<td>Missing</td>
<td>Missing</td>
</tr>
<tr>
<td>2</td>
<td>$z_{i2}^*$</td>
<td>Missing</td>
<td>Observed</td>
<td>Missing</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>2,999</td>
<td>$z_{i3}^*$</td>
<td>Missing</td>
<td>Missing</td>
<td>Observed</td>
</tr>
<tr>
<td>3,000</td>
<td>$z_{i1}^*$</td>
<td>Observed</td>
<td>Missing</td>
<td>Missing</td>
</tr>
</tbody>
</table>

✓ $Y_{i1}$ 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

<table>
<thead>
<tr>
<th>Parameters [True value]</th>
<th>Freely estimated $\sigma$</th>
<th>Fixed $\sigma = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimates</td>
<td>t-value</td>
</tr>
<tr>
<td>$\beta_{10}$ [1.00]</td>
<td>0.94</td>
<td>16.96</td>
</tr>
<tr>
<td>$\beta_{11}$ [0.25]</td>
<td>0.21</td>
<td>6.60</td>
</tr>
<tr>
<td>$\beta_{12}$ [-0.50]</td>
<td>-0.46</td>
<td>-14.23</td>
</tr>
<tr>
<td>$\beta_{20}$ [1.00]</td>
<td>0.89</td>
<td>9.87</td>
</tr>
<tr>
<td>$\beta_{21}$ [-0.25]</td>
<td>-0.26</td>
<td>-7.91</td>
</tr>
<tr>
<td>$\beta_{22}$ [-0.25]</td>
<td>-0.22</td>
<td>-6.82</td>
</tr>
<tr>
<td>$\beta_{30}$ [1.00]</td>
<td>0.96</td>
<td>13.63</td>
</tr>
<tr>
<td>$\beta_{31}$ [-0.25]</td>
<td>-0.25</td>
<td>-7.92</td>
</tr>
<tr>
<td>$\beta_{32}$ [-0.50]</td>
<td>-0.49</td>
<td>-15.00</td>
</tr>
<tr>
<td>$\sigma_1$ [0.0]</td>
<td>0.07</td>
<td>1.07</td>
</tr>
<tr>
<td>$\sigma_2$ [0.3]</td>
<td>0.37</td>
<td>3.84</td>
</tr>
<tr>
<td>$\sigma_3$ [-0.3]</td>
<td>-0.35</td>
<td>-4.35</td>
</tr>
</tbody>
</table>

### ATE: Average Treatment Effect

<table>
<thead>
<tr>
<th>$i$</th>
<th>$\max(z_i^*)$</th>
<th>$E[Y_{i1}]$</th>
<th>$E[Y_{i2}]$</th>
<th>$E[Y_{i3}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$z_{i1}^*$</td>
<td>$x'_1\beta_1$</td>
<td>$x'_1\beta_2$</td>
<td>$x'_1\beta_3$</td>
</tr>
<tr>
<td>2</td>
<td>$z_{i2}^*$</td>
<td>$x'_2\beta_1$</td>
<td>$x'_2\beta_2$</td>
<td>$x'_2\beta_3$</td>
</tr>
<tr>
<td>3,999</td>
<td>$z_{i3}^*$</td>
<td>$x'_{2,999}\beta_1$</td>
<td>$x'_{2,999}\beta_2$</td>
<td>$x'_{2,999}\beta_3$</td>
</tr>
<tr>
<td>3,000</td>
<td>$z_{i1}^*$</td>
<td>$x'_{3,000}\beta_1$</td>
<td>$x'_{3,000}\beta_2$</td>
<td>$x'_{3,000}\beta_3$</td>
</tr>
</tbody>
</table>

| Total / $n$ | $\frac{1}{n}\sum_{i=1}^{n}x'_i\beta_1$ | $\frac{1}{n}\sum_{i=1}^{n}x'_i\beta_2$ | $\frac{1}{n}\sum_{i=1}^{n}x'_i\beta_3$ |

<table>
<thead>
<tr>
<th>ATE</th>
<th>True $F_{\text{really estimated}}$</th>
<th>Fixed $\sigma = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[Y_1] - E[Y_2]$</td>
<td>0.25</td>
<td>0.29</td>
</tr>
<tr>
<td>$E[Y_1] - E[Y_3]$</td>
<td>0.49</td>
<td>0.48</td>
</tr>
</tbody>
</table>
Case study
To examine the causal effect of relocation to around a train station on individual car ownership probability in Kumamoto city, Japan.

Case study①: Setting

Residential choice model $Z$
(multinomial endogenous switching)

Access to train station

<table>
<thead>
<tr>
<th>i</th>
<th>max($z_i^*$)</th>
<th>$Y_{i1}$</th>
<th>$Y_{i2}$</th>
<th>$Y_{i3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$z_1^*$</td>
<td>Observed</td>
<td>Missing</td>
<td>Missing</td>
</tr>
<tr>
<td>2</td>
<td>$z_2^*$</td>
<td>Missing</td>
<td>Observed</td>
<td>Missing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n-1</td>
<td>$z_{n-1}^*$</td>
<td>Missing</td>
<td>Missing</td>
<td>Observed</td>
</tr>
<tr>
<td>n</td>
<td>$z_n^*$</td>
<td>Observed</td>
<td>Missing</td>
<td>Missing</td>
</tr>
</tbody>
</table>
Case study 2: Data

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

Access to train station

![Access to train station diagram](image)

<table>
<thead>
<tr>
<th>$i$</th>
<th>$\max(z_i^*)$</th>
<th>$Y_{i1}$</th>
<th>$Y_{i2}$</th>
<th>$Y_{i3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$z_1^*$</td>
<td>Observed</td>
<td>Missing</td>
<td>Missing</td>
</tr>
<tr>
<td>2</td>
<td>$z_2^*$</td>
<td>Missing</td>
<td>Observed</td>
<td>Missing</td>
</tr>
<tr>
<td>2,559</td>
<td>$z_3^*$</td>
<td>Missing</td>
<td>Missing</td>
<td>Observed</td>
</tr>
<tr>
<td>2,560</td>
<td>$z_1^*$</td>
<td>Observed</td>
<td>Missing</td>
<td>Missing</td>
</tr>
</tbody>
</table>

- 10.5% point difference between people living within 1,500m and over 3,000m from the nearest station
Case study ③: Result

<table>
<thead>
<tr>
<th>Z_i = 1</th>
<th>Freely estimated σ</th>
<th>Fixed σ = 0</th>
<th>68.7%</th>
<th>64.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>~1,500m</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z_i = 2</td>
<td>Freely estimated σ</td>
<td>Fixed σ = 0</td>
<td>70.6%</td>
<td>66.3%</td>
</tr>
<tr>
<td>1,501m~3,000m</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z_i = 3</td>
<td>Freely estimated σ</td>
<td>Fixed σ = 0</td>
<td>69.7%</td>
<td>71.8%</td>
</tr>
<tr>
<td>3,001m~</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

$\text{ATE}$

<table>
<thead>
<tr>
<th>Freely estimated σ</th>
<th>Fixed σ = 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[P(Y_1 = 1)] - E[P(Y_2 = 1)]$</td>
<td>-2.0%</td>
</tr>
<tr>
<td>$E[P(Y_1 = 1)] - E[P(Y_3 = 1)]$</td>
<td>-1.0%</td>
</tr>
</tbody>
</table>

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

Fixed σ = 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 study: Discussion

<table>
<thead>
<tr>
<th>$Z_i$</th>
<th>Freely estimated $\sigma$</th>
<th>Fixed $\sigma = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>~1,500m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>68.7%</td>
<td>64.5%</td>
</tr>
<tr>
<td>2</td>
<td>1,501m~3,000m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>70.6%</td>
<td>66.3%</td>
</tr>
<tr>
<td>3</td>
<td>3,001m~</td>
<td></td>
</tr>
<tr>
<td></td>
<td>69.7%</td>
<td>71.8%</td>
</tr>
</tbody>
</table>

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

- 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

Residential choice model

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

Error term

Car ownership model

$$y_{ij}^* = x'_{ij} \beta_j + u_{ij}$$

Error term

Attitudes toward car ownership and residential location
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 \( \text{Cov}(\varepsilon_{ij}, u_{ij}) \)

\[ Y_{ij} = x_i' \beta_j + u_{ij} \]

Travel behavior model

\[ (z^*_i \mid Y_i, \theta) \sim \text{N}_J \left( W_i \alpha + \Sigma_{Z,Y} \Sigma_Y^{-1} (Y_i - X_i \beta), \Sigma_Z - \Sigma_{Z,Y} \Sigma_Y^{-1} \Sigma_{Z,Y}^T \right), \]

if \( Y_i \neq Y'_i \), \( P(\max(z^*_i) = z^*_{ij} \mid Y_i) \neq P(\max(z^*_i) = z^*_{ij} \mid Y'_i) \).

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

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

\[ \Sigma_Y = \begin{pmatrix} \Sigma_{Y_1} & 0 & \cdots & 0 \\ 0 & \Sigma_{Y_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_{Y_J} \end{pmatrix} \]
Error structure and the full conditional distribution of $\sigma$

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

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

Conditional distribution of parameter $\beta, \sigma$ is

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

The conditional distribution does not include parameter $\beta, \sigma$ (Full conditional distribution)

We can use Gibbs sampling to approximate the posterior distribution of $\beta, \sigma$
Error structure and the full conditional distribution of $\sigma$

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

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

Conditional distribution of parameter $\beta, \sigma$ is

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

$g_0$: 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 $\sigma$

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

\[
\begin{pmatrix} z_i^* \\ Y_{ij} \end{pmatrix} \sim N_{J+1} \left[ \begin{pmatrix} W_i \alpha \\ x_i^j \beta_j \end{pmatrix}, \begin{pmatrix} \Sigma_Z & \Sigma_{Z,Y_j} \\ \Sigma_{Z,Y_j}^T & \Sigma_{Y_j} \end{pmatrix} \right]
\]

\[
Y_{ij} | z_i^* \sim N \left[ x_i^j \beta_j + \Sigma_{Z,Y_j} \Sigma_Z^{-1} (z_i^* - W_i \alpha), \nu_j^2 \right]
\]

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

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

where \( G_j = (G_{0j}^{-1} + \nu_j^{-2} V_j V_j)^{-1} \), \( g_j = G_j (G_{0j}^{-1} g_{0j} + \nu_j^{-2} V_j Y_j) \),

The conditional distribution does not include parameter \( b, \Sigma_{Z,Y_j} \) (Full conditional distribution)