

2023. 9. 20

第22回行動モデル夏の学校

# Causal-logit:

A brand-new unbiased logit model

with Behavioral changes due to disruptions in rail operations

---

東京大学交通研Aチーム

倉澤龍平, 須賀理帆, 須藤旺大, 林由翔

# Background

## 災害の常襲化

Disasters become a regular occurrence

## 鉄道施設の老朽化

Aging railroad facilities

## 鉄道の運転見合わせ・遅延の発生

Suspension or delay of railroad services

[1]

変圧器内ショートか 21年変電所火災、JR東日本原因発表

社会・調査  
2022年4月12日 19:24

保存



火災で煙を上げる埼玉県蕨市のJR東日本の変電所（2021年10月10日、近隣住民提供）=共同

[2]

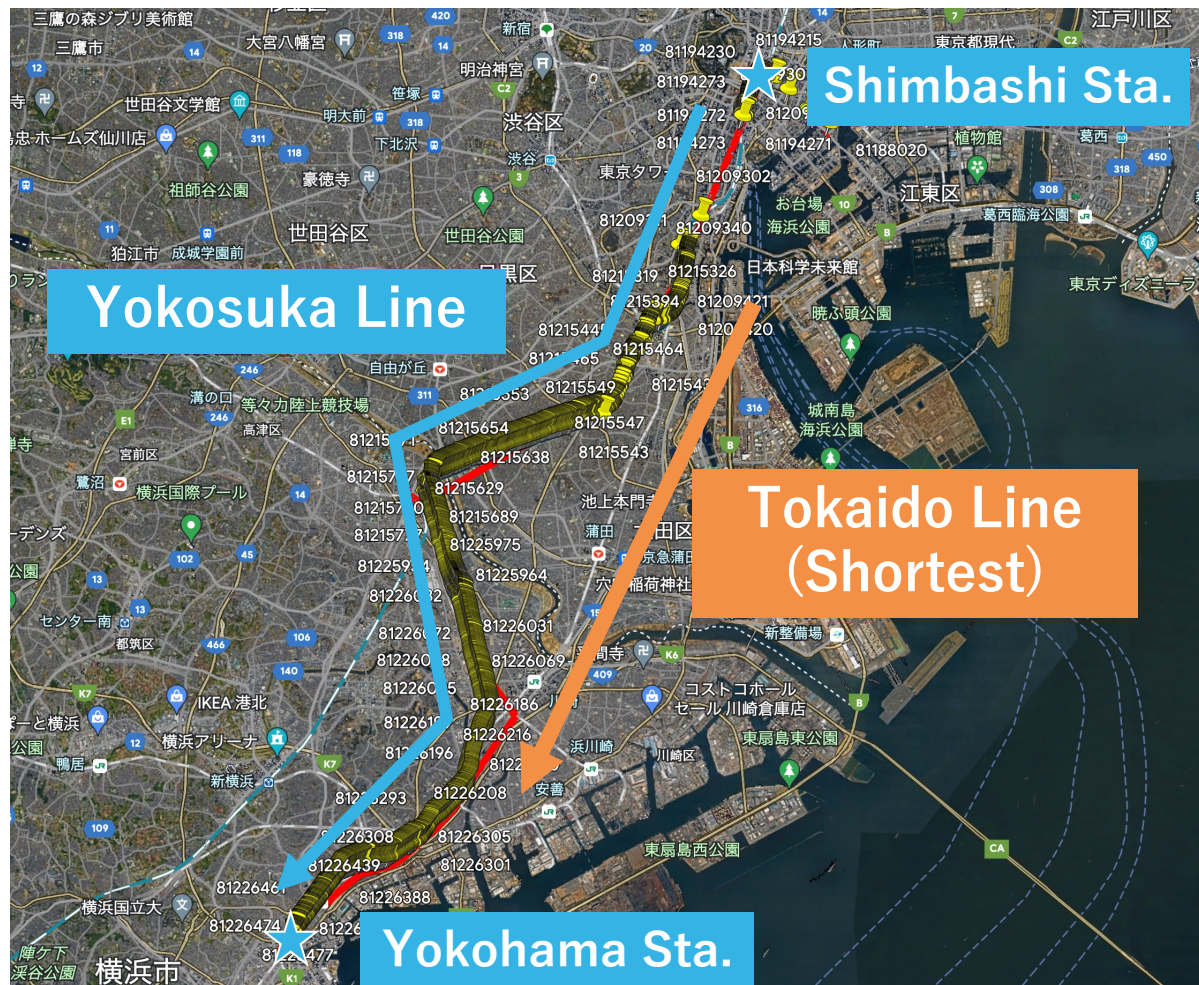
【まとめ】台風19号情報 各交通機関が計画運休  
2019年10月11日 15時01分



# Background

## 運転見合わせ時に行動変更する可能性

Possibility to change one's behavior when operation is suspended



Ex) 2021/11/08の新橋→横浜の移動

Travel from Shimbashi to Yokohama on 2021/11/08

[3]

JR東海道線横浜駅で信号トラブル、原因は線路上の鉄片

JR東海道線 信号 運転見合わせ

事件事故 | 神奈川新聞 | 2021年11月8日(月) 16:29



[3]<https://www.kanaloco.jp/news/social/case/article-738590.html>

東海道線運転見合わせのため、  
横須賀線を選択

Select Yokosuka Line due to Tokaido Line suspension

# Purpose

## 鉄道トラブル発生時の経路選択行動の平常時との違いを説明

Understanding the differences in route selection behavior during railroad trouble compared to normal conditions.

### 課題：遅延や運転見合わせへの遭遇頻度の非ランダム性

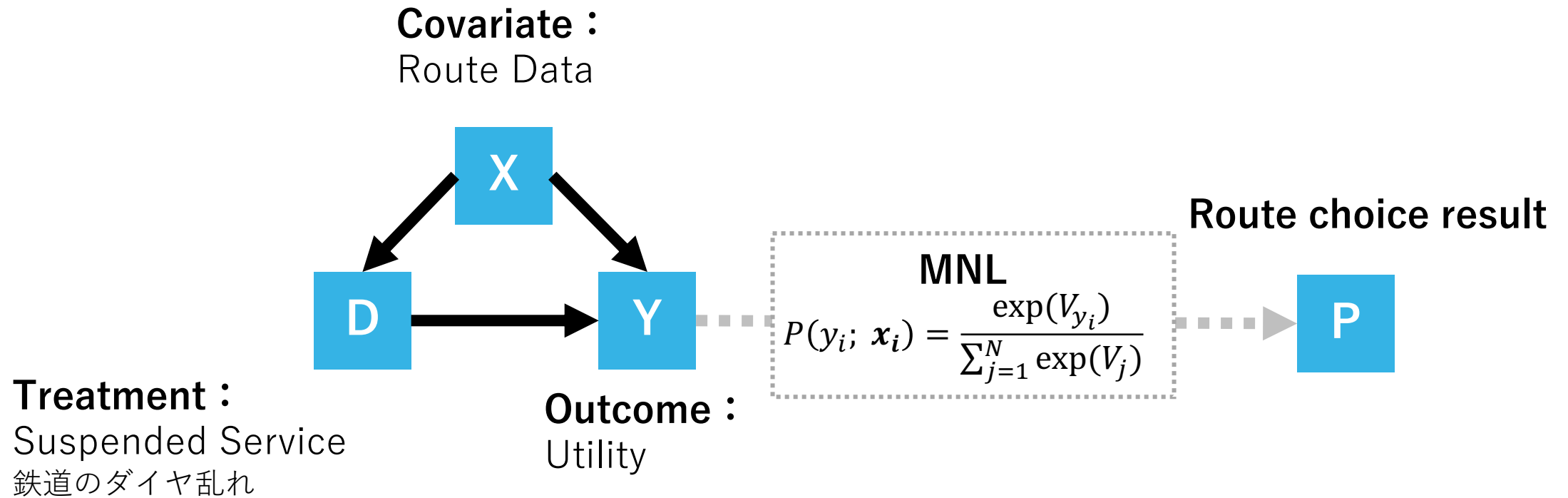
**Problem: Non-randomness in the frequency of encounters with delays and operational disruptions**

遅延や運転見合わせの発生頻度は時間帯や経路の乗り換え回数などに依存  
Frequency of delays and suspensions depends on time of day, number of route transfers, etc.

**通常のMNLでは事故データの偏りからバイアスのかかった推定になってしまう！**

The usual MNL results in biased estimates due to the bias of the accident data!

# Framework



# Framework

## How to collect data

Information from X



とれいんふぉ 首都圏エリア 非公式運行情報など  
@Trainfo

【東武東上線 運転再開見込(変更) 21:20】

東上線は、19:08頃、北坂戸～高坂での人身事故の影響で、川越市～森林公園の上下線で運転を見合わせています。現場状況により、再開見込はさらに前後する場合があります。運転区間の一部列車に遅れと運休・運転変更がでています。

★振替輸送利用可能★

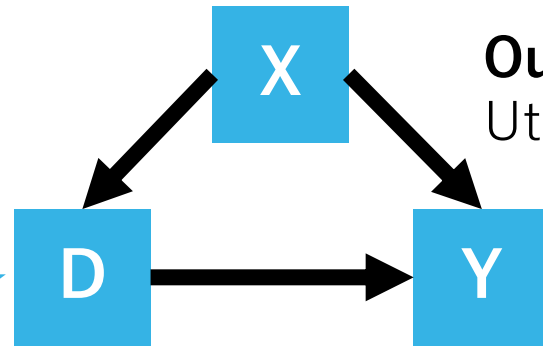


Accident data csv

2021/08/09 14:21, 東日本旅客鉄道, 京葉線, 新木場, 強風による運転規制値超過, 東京の上下線の一部列車で運転見合わせ  
2021/08/09 13:40, 東京急行電鉄, 大井町線, 大井町～旗の台, 架線断線に伴う停電・強風による運転規制値超過, 大井町～溝の口の上下線で運転見合わせ  
2021/08/08 08:52, 京王電鉄, 京王線, 平山城址公園～長沼, 踏切内自動車立往生, 高幡北野の上下線で運転見合わせ  
2021/08/06 20:30, 小田急電鉄, 小田原線, 祖師ヶ谷大蔵駅付近, 走行中の列車内においてを使用した無差別傷害事件発生(犯人確保済), 新宿～向ヶ丘遊園の上下線で運転見合わせ  
2021/08/06 15:50, 東京都交通局, 新宿線, 東大島～船堀, 架線に飛来物の影響, 大島～の上下線で運転見合わせています。

Covariate :  
Route Data

Outcome :  
Utility



Treatment :  
Suspended Service  
鉄道のダイヤ乱れ

$$P(y_i; x_i) = \frac{\exp(V_{y_i})}{\sum_{j=1}^N \exp(V_j)}$$

Route choice  
result



# Framework

## How to collect data

Information from X



とれいんふお 首都圏エリア 非公式運行情報など  
@Trainfo

【東武東上線 運転再開見込(変更) 21:20】

東上線は、19:08頃、北坂戸～高坂での人身事故の影響で、川越市～森林公園の上下線で運転を見合わせています。現場状況により、再開見込はさらに前後する場合があります。運転区間の一部列車に遅れと運休・運転変更がでています。

★振替輸送利用可能★



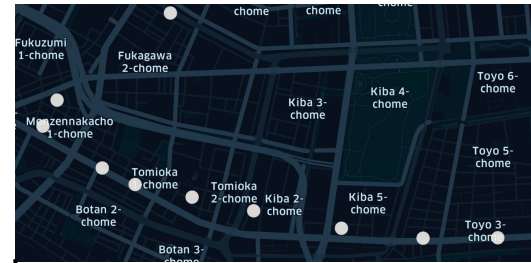
Accident data csv

2021/08/09 14:21, 東日本旅客鉄道, 京葉線, 新木場, 強風による運転規制値超過, 東京の上下線の一部列車で運転見合わせ  
2021/08/09 13:40, 東京急行電鉄, 大井町線, 大井町～旗の台, 架線断線に伴う停電・強風による運転規制値超過, 大井町～溝の口の上下線で運転見合わせ  
2021/08/08 08:52, 京王電鉄, 京王線, 平山城址公園～長沼, 踏切内自動車立往生, 高幡北野の上下線で運転見合わせ  
2021/08/06 20:30, 小田急電鉄, 小田原線, 祖師ヶ谷大蔵駅付近, 走行中の列車内においてを使用した無差別傷害事件発生(犯人確保済), 新宿～向ヶ丘遊園の上下線で運転見合わせ  
2021/08/06 15:50, 東京都交通局, 新宿線, 東大島～船堀, 架線に飛来物の影響, 大島～の上下線で運転見合わせています。

Covariate :  
Route Data

## Data collection

PP data

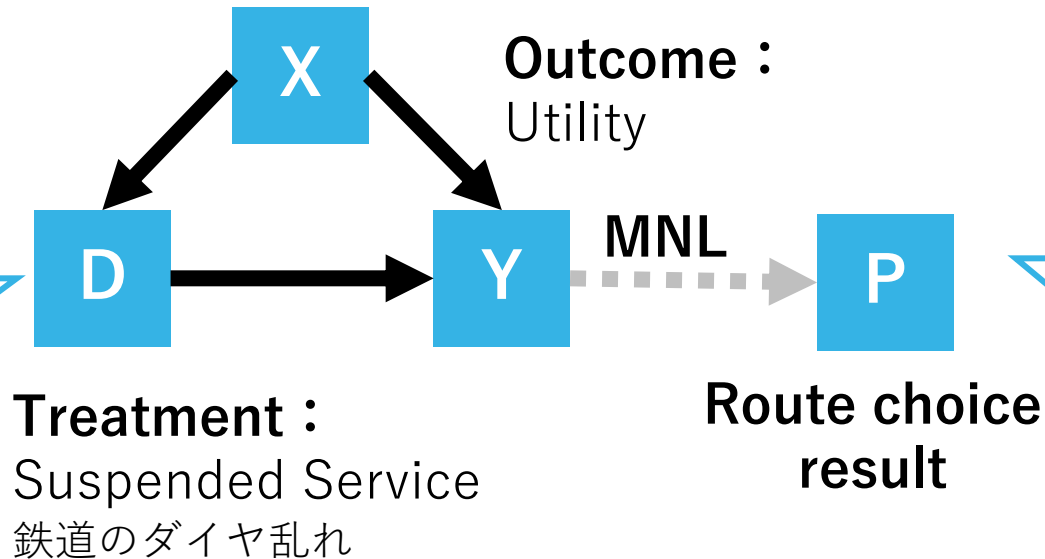
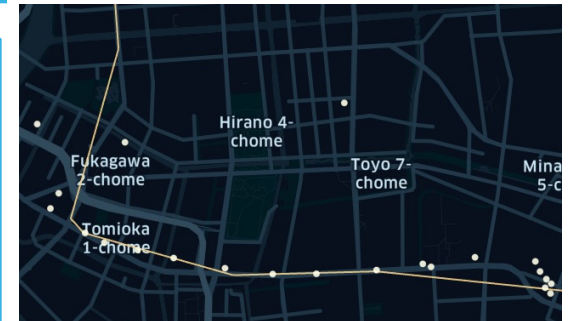


Train Network



Map matching

Route data



# Formulation

## Ordinary MNL

### Utility Function

$$V_i = (\alpha_1 + \alpha_2 \cdot D_i) \cdot \text{Time}_i + (\beta_1 + \beta_2 \cdot D_i) \cdot \text{Transition}_i$$

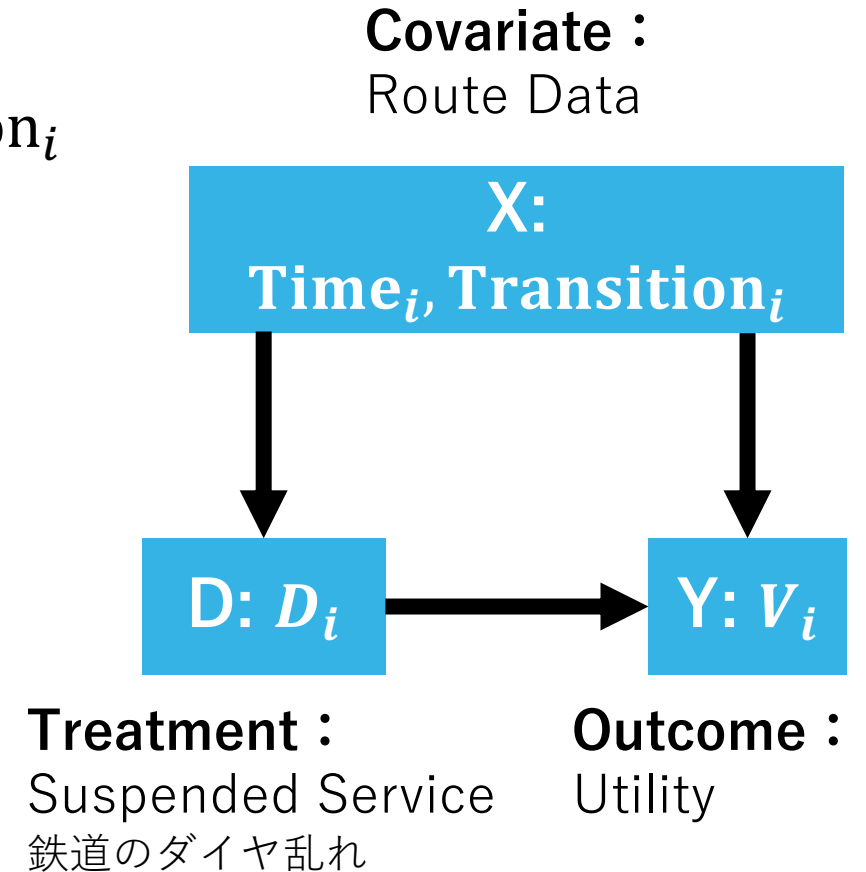
$i$ : Route

$D_i$ : Accident dummy

$\alpha, \beta$ : Parameters

**XとDに因果関係があると、正しく効果が推定できない**  
( $D_i = 1$ となるデータの性質が偏っているため)

**If there is a causal relationship between X and D, the effect cannot be estimated correctly**  
(due to the biased nature of the data where  $D_i = 1$ )





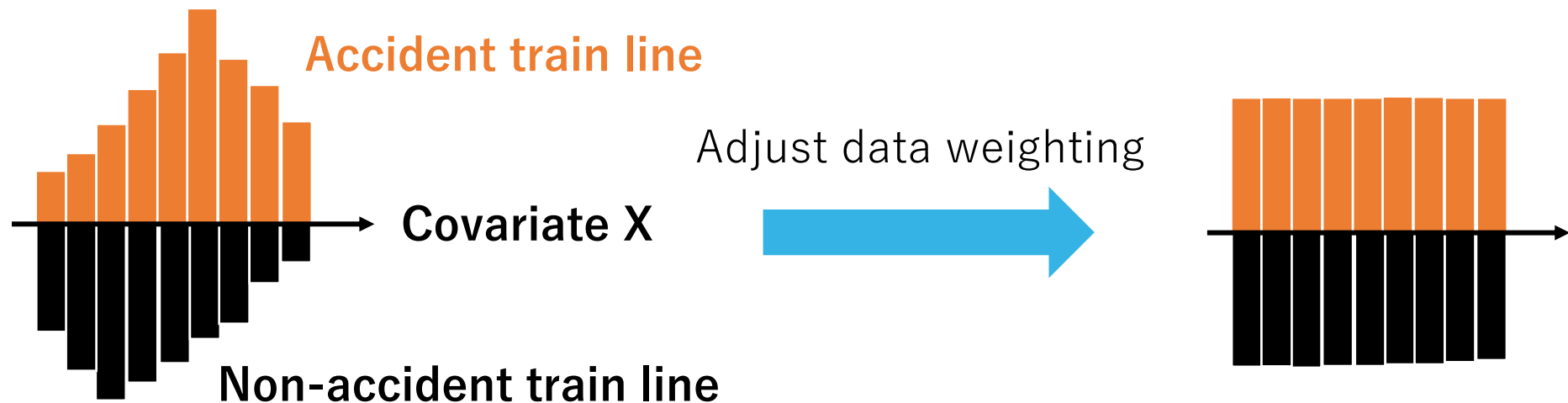
# Formulation

## Causal logit (Our model)

「データに適切な重み付けをかける&反実仮想を推定する」ことで、効果をバイアスなく推定する

Estimating effects without bias by "applying appropriate weights to the data & estimating counterfactuals"

Image



Covariate bias

Elimination of bias

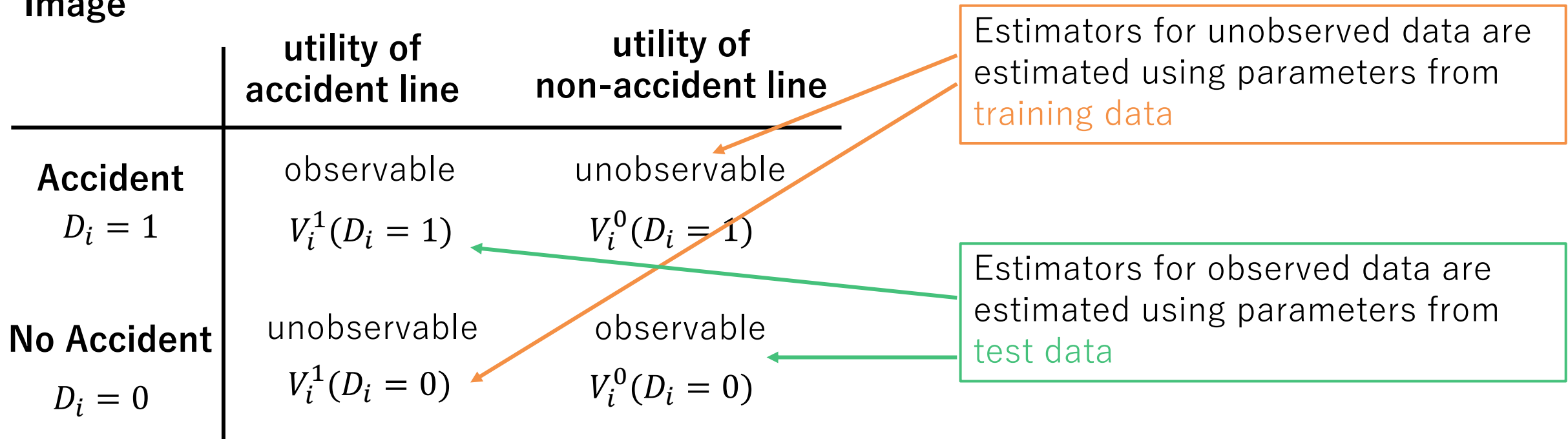
# Formulation

## Causal logit (Our model)

「データに適切な重み付けをかける&反実仮想を推定する」ことで、効果をバイアスなく推定する

Estimating effects without bias by "applying appropriate weights to the data & **estimating counterfactuals**"

Image



# Formulation

## Causal logit (Our model)

### Step1: Split data



### Step2: Make prediction models of utility and prosperity score

(by using training data)

#### Utility

$$V_i = 1(D_i = 1) \cdot (\alpha_1^1 \cdot \text{Time}_i + \alpha_2^1 \cdot \text{Transition}_i) + 1(D_i = 0) \cdot (\alpha_1^0 \cdot \text{Time}_i + \alpha_2^0 \cdot \text{Transition}_i)$$

$\alpha_k^j$ :  $k$ th parameter when  $D = j$

#### Prosperity Score (= Probability to be treated)

$$e(X_i) = \text{Prob}(D_i = 1|X_i)$$

Create model by using **Random Forest**

# Formulation

## Causal logit (Our model)



Utility estimation from **training data**

$$\hat{V}_i^1 = \hat{\alpha}_1^1 \cdot \text{Time}_i + \hat{\alpha}_2^1 \cdot \text{Transition}_i$$

$$\hat{V}_i^0 = \hat{\alpha}_1^0 \cdot \text{Time}_i + \hat{\alpha}_2^0 \cdot \text{Transition}_i$$

$\hat{V}_i^j$ : Route  $i$ 's predicted utility  
(from training data) when  $D_i = j$

### Prosperity Score

$$\hat{e}(X_i) = \text{Prob}(D_i = 1|X_i)$$

## Step3: AIPW estimation

Utility estimation from **test data**

$$V_i^1 = \alpha_1^1 \cdot \text{Time}_i + \alpha_2^1 \cdot \text{Transition}_i$$

$$V_i^0 = \alpha_1^0 \cdot \text{Time}_i + \alpha_2^0 \cdot \text{Transition}_i$$

AIPW estimator (Robins et al., 1995)

$$\Delta V_{i,\text{AIPW}} \quad \begin{array}{l} \text{use predicted utility in case that } V_i^1 \text{ is} \\ \text{unobserved} \end{array}$$

$$= \left( \frac{D_i}{\hat{e}(X_i)} V_i^1 + \left( 1 - \frac{D_i}{\hat{e}(X_i)} \right) \hat{V}_i^1 \right)$$

↓

$$- \left( \frac{1 - D_i}{1 - \hat{e}(X_i)} V_i^0 + \left( 1 - \frac{1 - D_i}{1 - \hat{e}(X_i)} \right) \hat{V}_i^0 \right)$$

↑ weighting by prosperity score to balance covariates

# Formulation

## Causal logit (Our model)

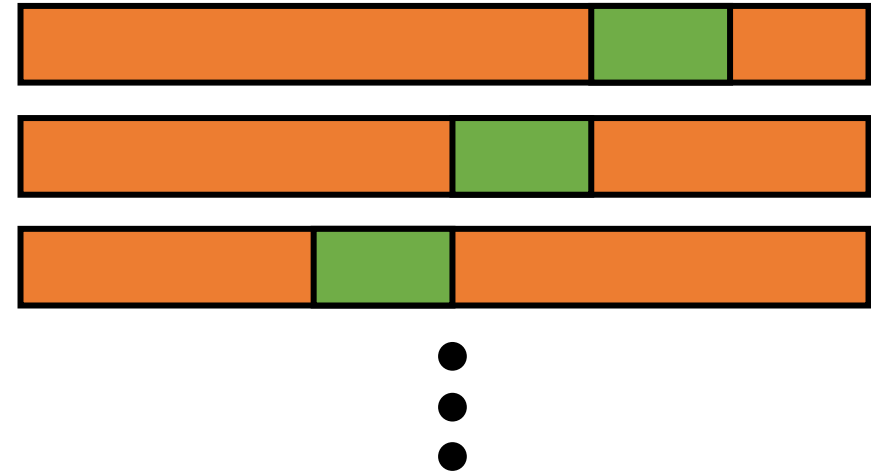


AIPW estimator (Robins et al., 1995)

$$\begin{aligned} \Delta V_{i,\text{AIPW}} &= \left( \frac{D_i}{\hat{e}(X_i)} V_i^1 + \left( 1 - \frac{D_i}{\hat{e}(X_i)} \right) \hat{V}_i^1 \right) \\ &\quad - \left( \frac{1 - D_i}{1 - \hat{e}(X_i)} V_i^0 + \left( 1 - \frac{1 - D_i}{1 - \hat{e}(X_i)} \right) \hat{V}_i^0 \right) \end{aligned}$$

## Step4: Cross Estimation

Training Data    Test Data



Obtain  $\Delta V_{i,\text{AIPW}}$  with different combinations of data

## Step5: Best Linear Projection

(Semenova et al., 2021)

Linear regression of  $\Delta V_{i,\text{AIPW}}$

$$\Delta V_{i,\text{IPW}} = C + \Delta\alpha_{1,\text{IPW}} \cdot \text{Time}_i + \Delta\alpha_{2,\text{IPW}} \cdot \text{Transition}_i$$

# Ordinary MNL VS Causal-logit

$$\text{Real utility function: } V = \begin{cases} 5x_1 + 20x_2 + \varepsilon & (D = 1) \\ 3x_1 + 40x_2 + \varepsilon & (D = 0) \end{cases}$$

$$Pr(D = 1) = (1 + x_3) \times 0.01$$

$x_3$  not exist

→ Causal inference (ratio) should be  $[x_1, x_2, x_3] = [2, -20, 0]$

Ordinary MNL

Causal Logit

param	x	t-value	
$x_1(D = 1)$	$-4.92 \times 10^{-2}$	-27.73 ***	✗ Causal inference? → $\tau_1 = -3.59 \times 10^{-2}$
$x_1(D = 0)$	$-1.33 \times 10^{-2}$	-13.39 ***	
$x_2(D = 1)$	$-5.83 \times 10^{-2}$	-17.12 ***	→ $\tau_2 = -3.65 \times 10^{-1}$
$x_2(D = 0)$	$-4.13 \times 10^{-1}$	-52.26 ***	
$x_3(D = 1)$	$8.05 \times 10^{-2}$	3.62 ***	→ $\tau_3 = -0.91 \times 10^{-2}$
$x_3(D = 0)$	$1.71 \times 10^{-1}$	12.02 ***	

param	x	t-value
$\tau_1$	$7.96 \times 10^{-2}$	5.09 ***
$\tau_2$	$-2.90 \times 10^{-2}$	-70.10 ***
$\tau_3$	$3.19 \times 10^{-3}$	124.70 ***
const	$-1.58 \times 10^{-1}$	-25.54 ***
n	6000	
$R^2$	0.42	

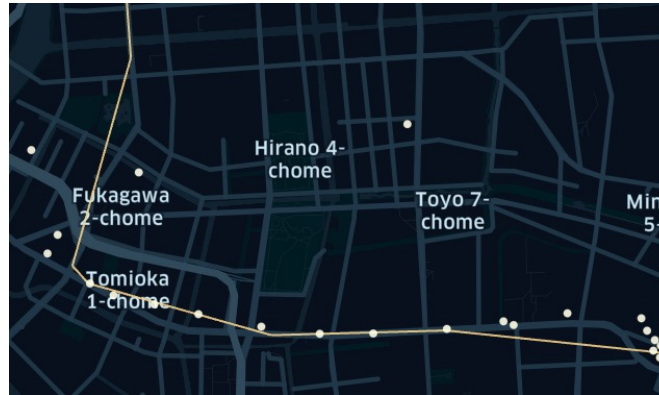
Could estimate that  $x_3$  is not causally related to utility  $V$

n	6000
Initial log likelihood	92103.40
Final log likelihood	89761.50
Corrected $\rho^2$	0.025

# Estimation with route choice data

## Ordinary MNL

param	x	t-value
time ( $D = 1$ )	$-8.39 \times 10^{-2}$	-1.03
time ( $D = 0$ )	$-1.79 \times 10^{-1}$	-6.48 ***
transfer ( $D = 1$ )	$-9.31 \times 10^{-2}$	-1.00
transfer ( $D = 0$ )	$-1.87 \times 10^{-1}$	-15.11 ***
commute * time ( $D = 1$ )	$-2.70 \times 10^{-2}$	-2.48 *
commute * time ( $D = 0$ )	$1.99 \times 10^{-4}$	0.04
n	6913 trips	
Initial log likelihood	6386.60	
Final log likelihood	6048.79	
Corrected $\rho^2$	0.058	



## Causal Logit

param	x	t-value
time $\tau$	$2.66 \times 10^{-1}$	2.52 *
transfer $\tau$	$-2.24 \times 10^{-3}$	-0.63
[commute * time] $\tau$	$4.63 \times 10^{-2}$	-1.12
const	$-3.02 \times 10^{-3}$	-0.81
n	6913 trips	
$R^2$	$3.33 \times 10^{-3}$	

**Tendency to accept delays in emergencies**

# References

---

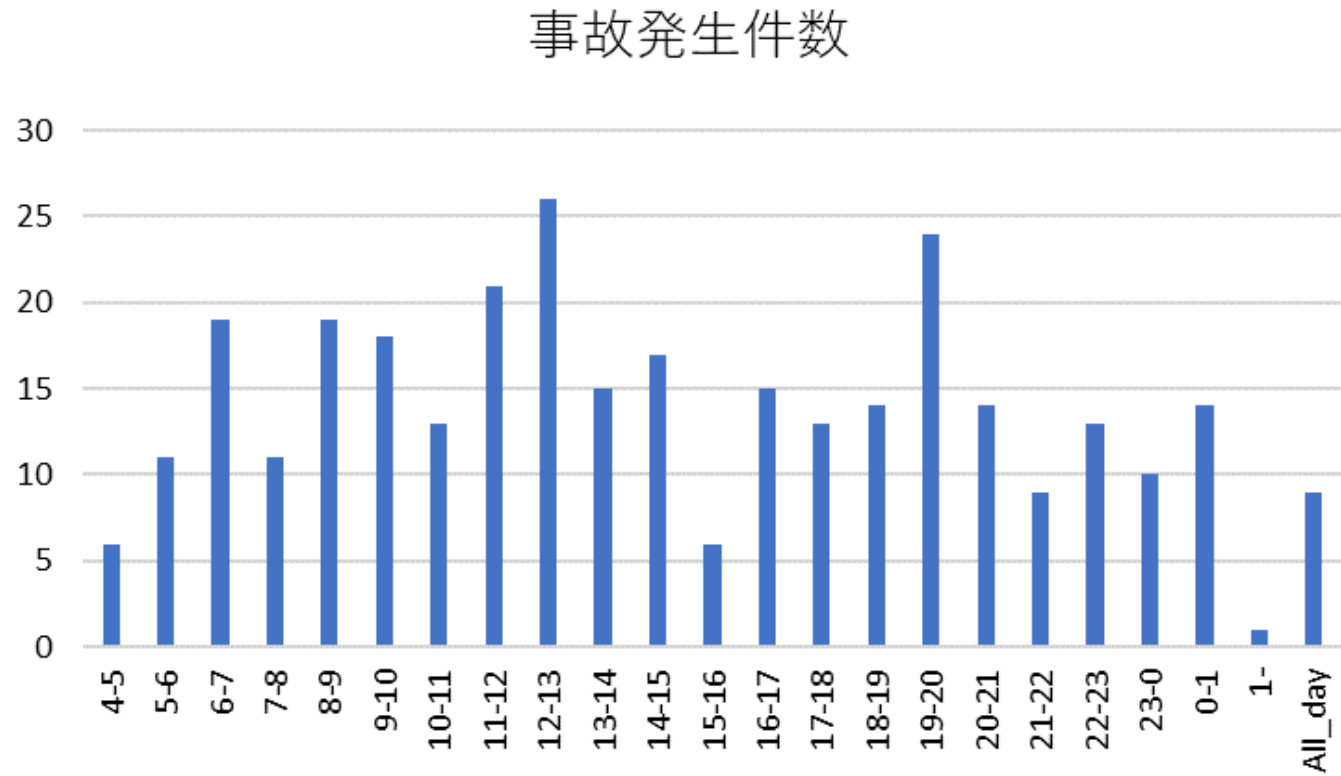
- Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the american statistical association*, 90(429), 106-121.
- Semenova, Vira, and Victor Chernozhukov. "Debiased Machine Learning of Conditional Average Treatment Effects and Other Causal Functions". *The Econometrics Journal* 24.2 (2021)
-



# Appendix

---

# 基礎集計 時間帯ごとの鉄道事故発生件数



# 基礎集計 路線別の事故発生件数

