Data-Fusion Approach for Trip-Purpose Estimation of Inter-Regional Passengers

Oct, 20th, 2017

ARATA SUZUKI (TOKYO INSTITUTE OF TECHNOLOGY)

HIROMICHI YAMAGUCHI (KANAZAWA UNIVERSITY)

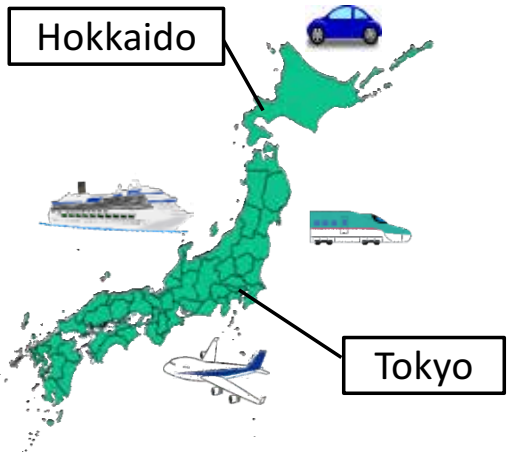
DAISUKE FUKUDA (TOKYO INSTITUTE OF TECHNOLOGY)

1. Introduction - Background

"Inter-Regional Travel Survey in Japan (IRTS)"

Survey that collects the movement of inter-regional passengers in Japan

- Target: all inter-prefectural(Pref.) travel
- Many qualitative information about trip (residence, destination, trip purpose etc…)
- Conducted **only in Autumn every 5 years**



→No data for spring, summer and winter passengers!

e.g.) Hokkaido Shinkansen

Open!

Tokyo ⟷ Hokkaido

How many tourists are there in winter in Hokkaido?

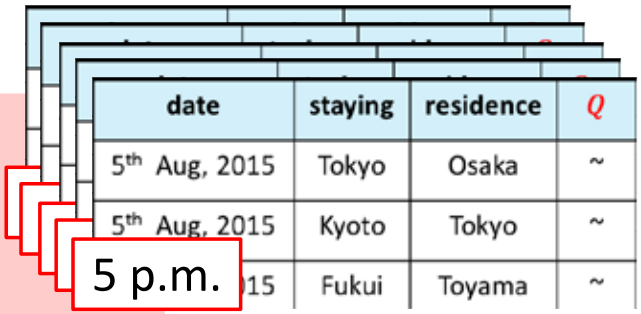
But IRTS has only Autumn data…

1. Introduction - Background

"Mobile Spatial Statistics (MSS)"

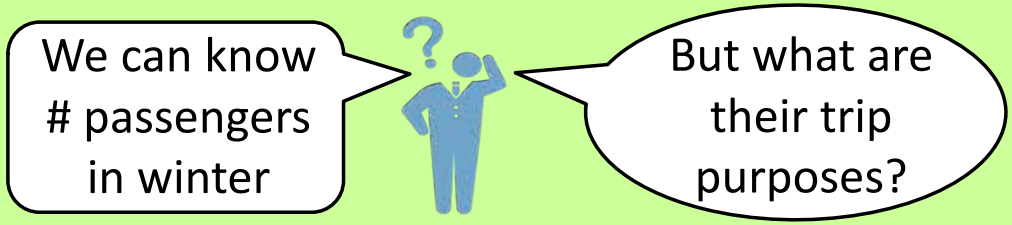
Aggregate mobile phone data
by NTT DOCOMO*

- Records the number of mobile phone users by each residential and staying place per one hour
- **Estimated from mobile phone location data**



→ No information about passengers' trip purposes!

e.g.) Hokkaido Shinkansen



*One of the biggest mobile phone operator (company) in Japan

1. Introduction - datasets

	IRTS	MSS	JNTS*
Survey interval	Every 5 years (one weekday and holiday in Autumn)	Every 1 hour every day	Every 3 months
Gender and age	Yes	Yes	Yes
Residential Pref.	Yes	Yes	Yes
Staying night Pref.	Yes	Yes	Yes
Trip-purposes	Yes	No	Yes
Travel mode	Yes	No	No
Flow or Staying information	flow	staying	flow
Survey Method	Questionnaire (sampling survey data)	mobile phone data from NTT docomo users (scaled up to the population in Japan)	Questionnaire (sampling survey data)

* Japan National Tourism Survey

1. Introduction - Objective

Data fusion

MSS
(data for all the year)

×

IRTS
(data with trip purpose)

To estimate inter-regional passenger flow by each trip purpose for various days and seasons

Previous Studies:

Muroi et.al.(2015)

Used only MSS to understand estimate inter-regional passenger flow, but this method does not consider trip purpose.

Janzen et.al.(2016)

Estimated the trip purpose by data-fusion approach of non-aggregated mobile phone location data × questionnaire survey

2. Model Formulation (regression model)

Trip purposes include (1)Business, (2)Sightseeing and (3)Private

Business | Sightseeing | Private

$$Q_{h,s,d} = V_{h,B,d} R_{h,s,B} + V_{h,S,d} R_{h,s,S} + V_{h,Pr,d} R_{h,s,Pr}$$

(h: residential Pref., s: staying night Pref., d: date, B : Business purpose, S : Sightseeing purpose, Pr : Private purpose) ※only for accommodation passengers

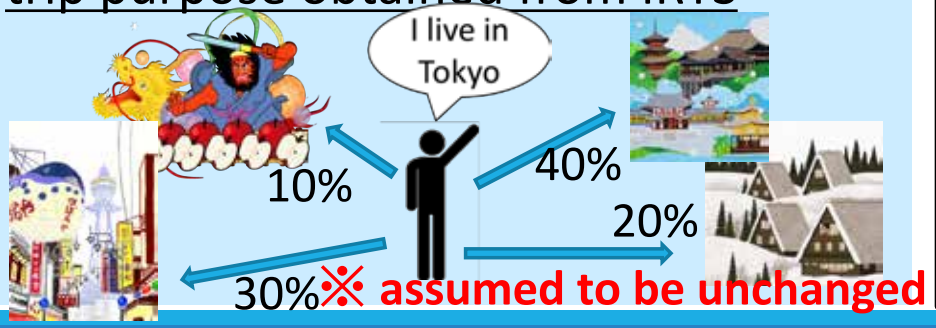
Response Variable Q: # staying people from MSS at 4 AM.

(Assumption: the staying place at 4 AM is his/her accommodation place)

Explanatory Variable from IRTS

$\boldsymbol{R}$ $(0 \leq R \leq 1)$

Share of selecting staying Pref. by each trip purpose obtained from IRTS



※ assumed to be unchanged

Parameters to be estimated

(By constrained least squares method)

$\boldsymbol{V}(\geq 0)$

passenger for each trip purpose from each residential Pref.

e.g.)
for sightseeing
live in Tokyo

2. Model Formulation (regression model): Example

$$Q_{h,s,d} = V_{h,B,d}R_{h,s,B} + V_{h,S,d}R_{h,s,S} + V_{h,Pr,d}R_{h,s,Pr}$$

(h: residential Pref., s: staying night Pref., d: date, B : Business purpose, S : Sightseeing purpose, Pr : Private purpose) ※only for accommodation passengers

In case of residence(h)=Tokyo and date(d)=Oct. 21, 2015

MSS Q

share of selecting staying Pref. by each trip purpose obtained from IRTS R

	h=Tokyo	h=Tokyo, Business	h=Tokyo, Sightseeing	h=Tokyo, Private
s=Hokkaido	15081	0.0518	0.0260	0.0355
s=Aomori	2993	0.0188	0.0062	0.0185
⋮	⋮	⋮	⋮	⋮
s=Kagoshima	2886	0.0103	0.0059	0.0193
s=Okinawa	5794	0.0184	0.0231	0.0238

$= V_{h,B,d}$ $+ V_{h,S,d}$ $+ V_{h,Pr,d}$

parameter : # passenger for each trip purpose from each residential Pref.

"#date × #residential Pref." parameters are estimated

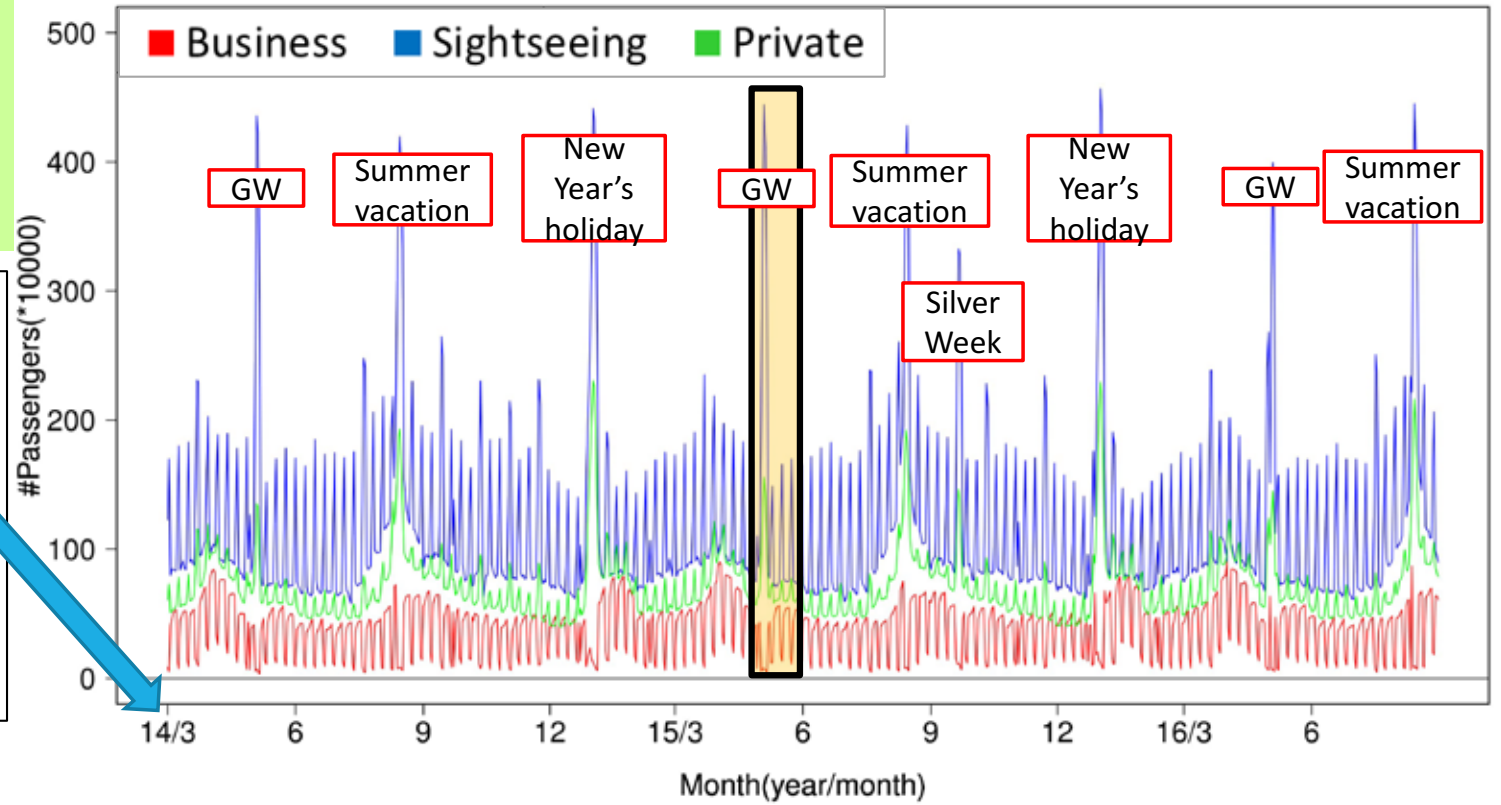
3. Estimated "Total trip generations" $\widehat{Vs}$ across different days

Total trip generations by purposes

$$\sum_h \hat{V}$$

Data
IRTS:2010
MSS : Mar, 2014
~ Aug, 2016
(915 days)

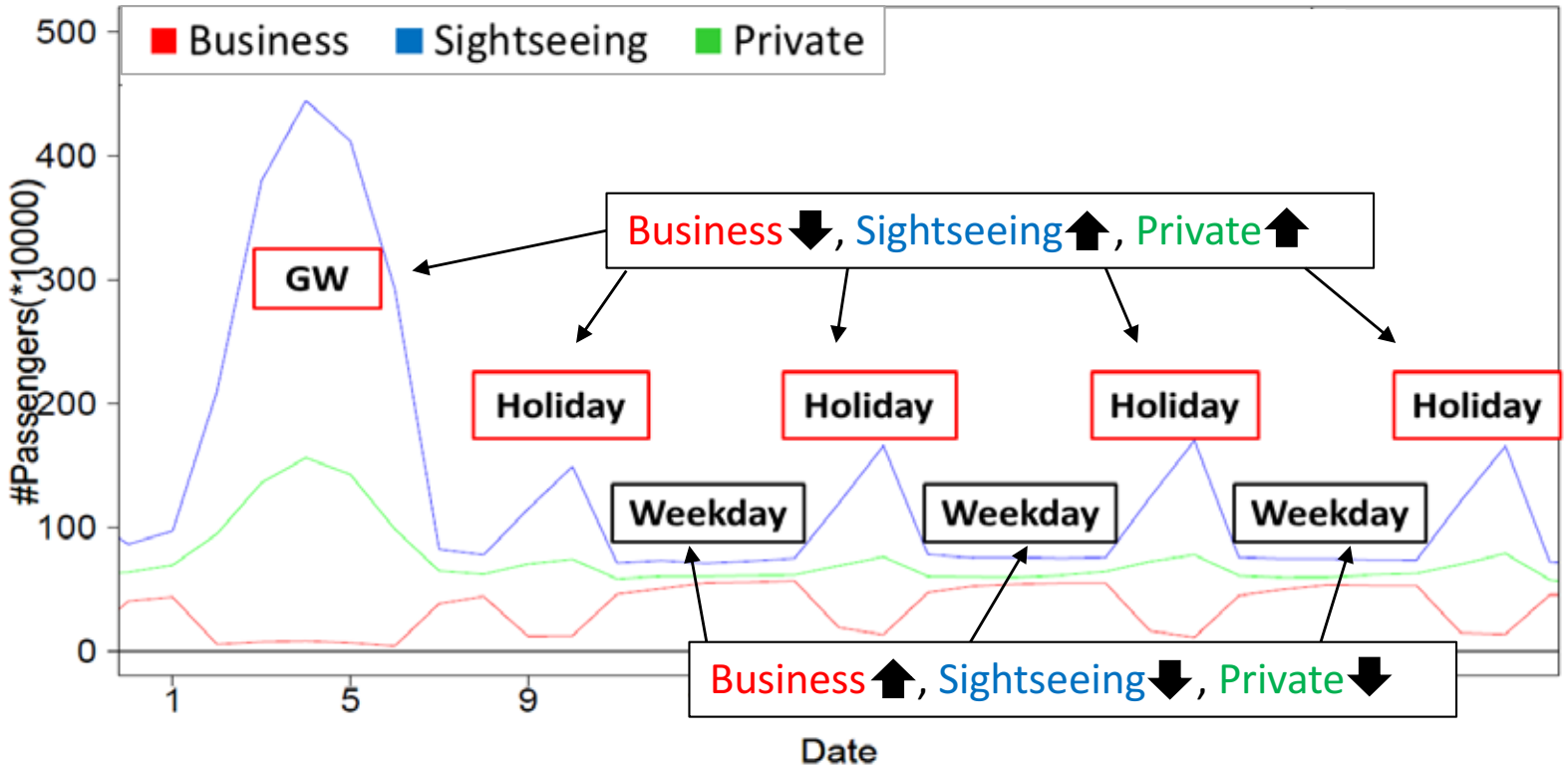
Spatial Unit
Prefecture



The coefficient of determination is 0.52(Summer vacation) ～0.94(around survey date of IRTS)

There are many passengers for sightseeing and private in long vacation such as summer vacation, New year's holiday

3. Estimated "Total trip generations" $\widehat{Vs}$ in May, 2015

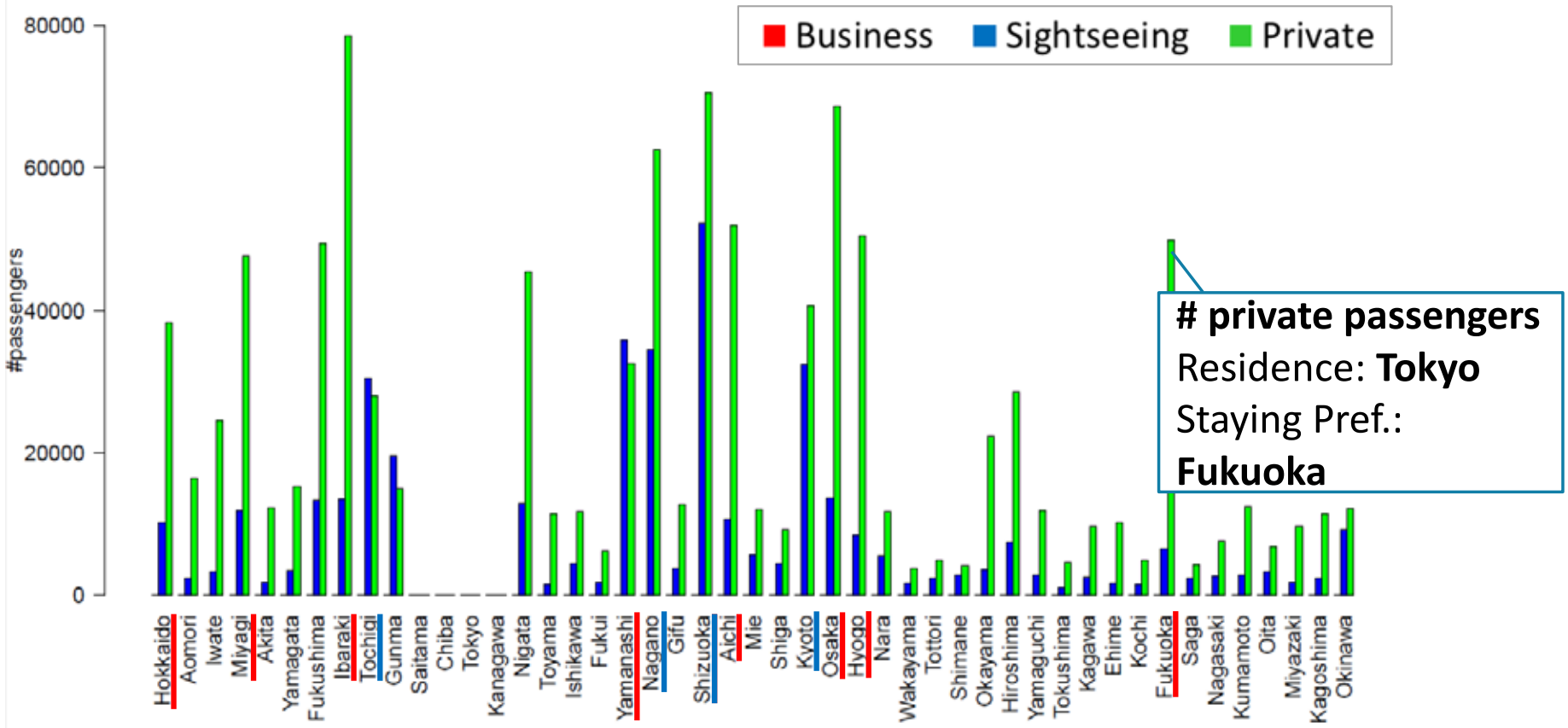


The model seems to reproduce seasonal / daily variation of passenger flow well.

3. Estimated inter regional passenger flow

~ Residence : Tokyo(metropolitan), date : New Year's day ~

Result of Jan. 1, 2015

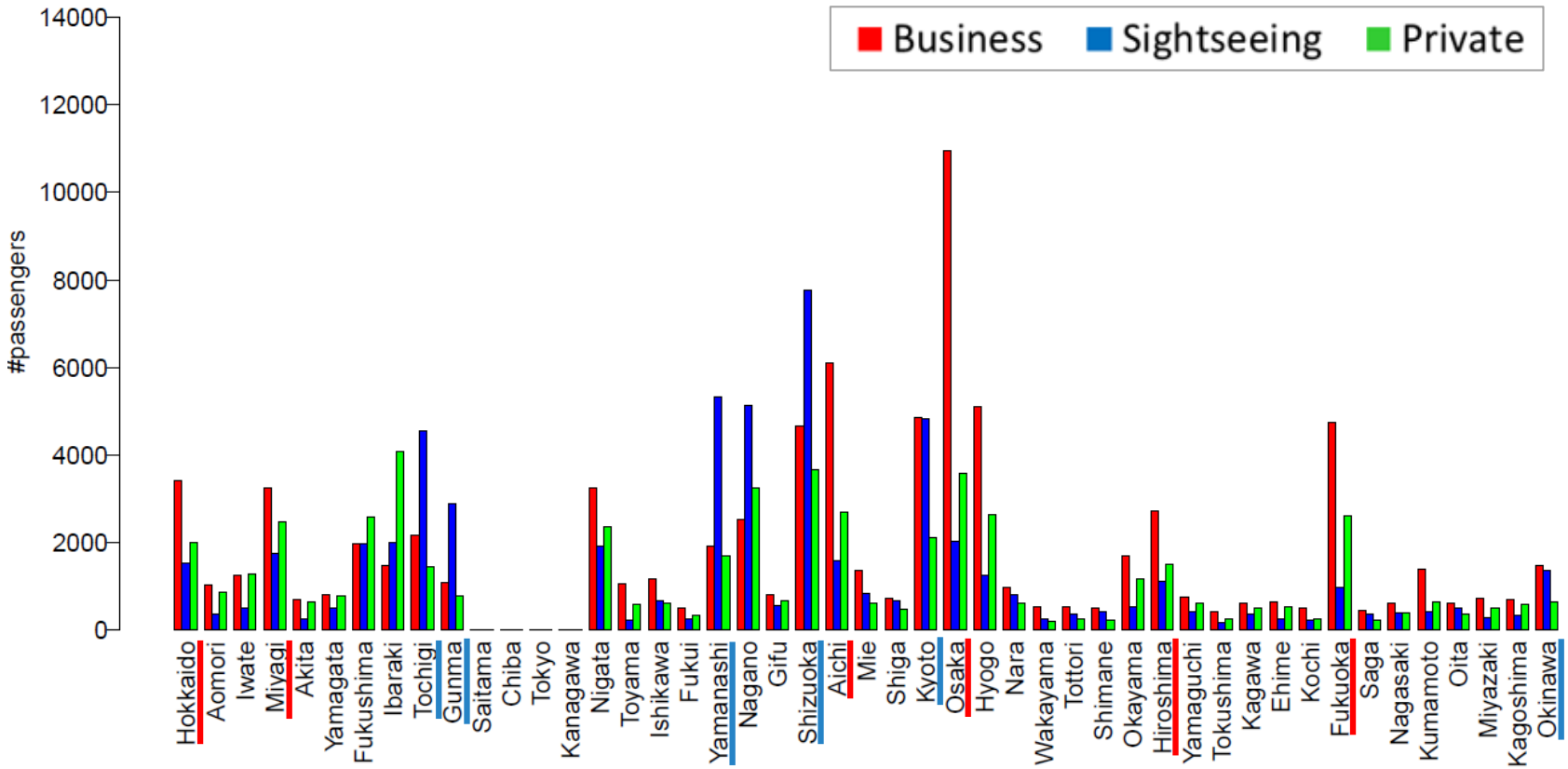


- Many private purpose passengers ⇒ Pref. with metropolitan areas (ex.Osaka)
- Many sightseeing purpose passengers ⇒ Pref. with famous sightseeing spot

3. Estimated inter regional passenger flow

~ Residence : Tokyo(metropolitan), date : weekday in Oct ~

Result of Oct. 21, 2015 (Wed)

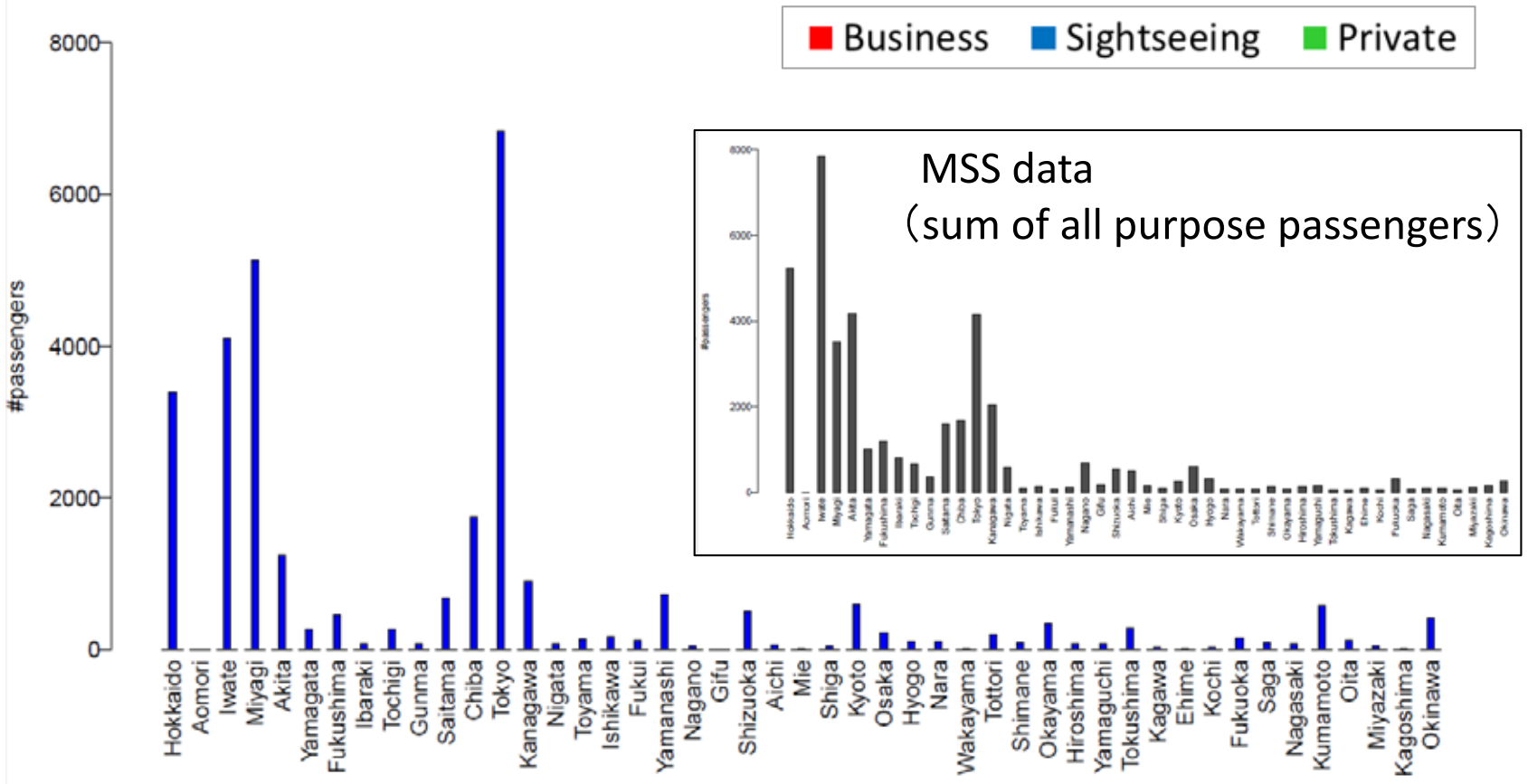


- Many business passengers ⇒ Pref. with metropolitan areas (e.g. Osaka)
- Many sightseeing purpose passengers ⇒ Pref. with famous sightseeing spot

3. Estimated inter regional passenger flow

~ Residence : Aomori(rural Pref.), date : New Year's day ~

Result of Jan. 1, 2015

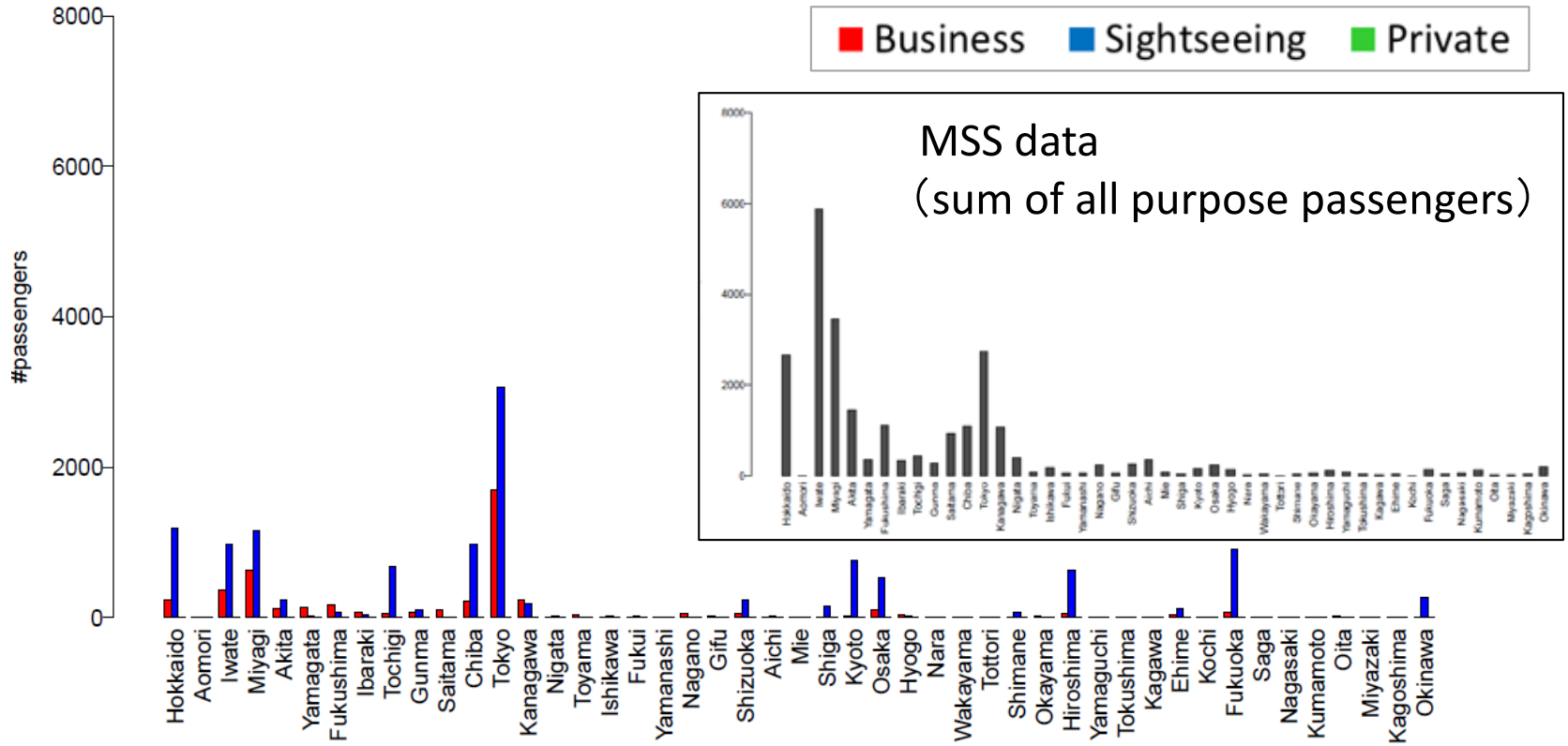


- Only sightseeing passengers
- Overestimation to Tokyo ⇔ Underestimation near Aomori

3. Estimated inter regional passenger flow

~ Residence : Aomori(rural Pref.), date : weekday in Oct ~

Result of Oct. 21, 2015 (Wed)

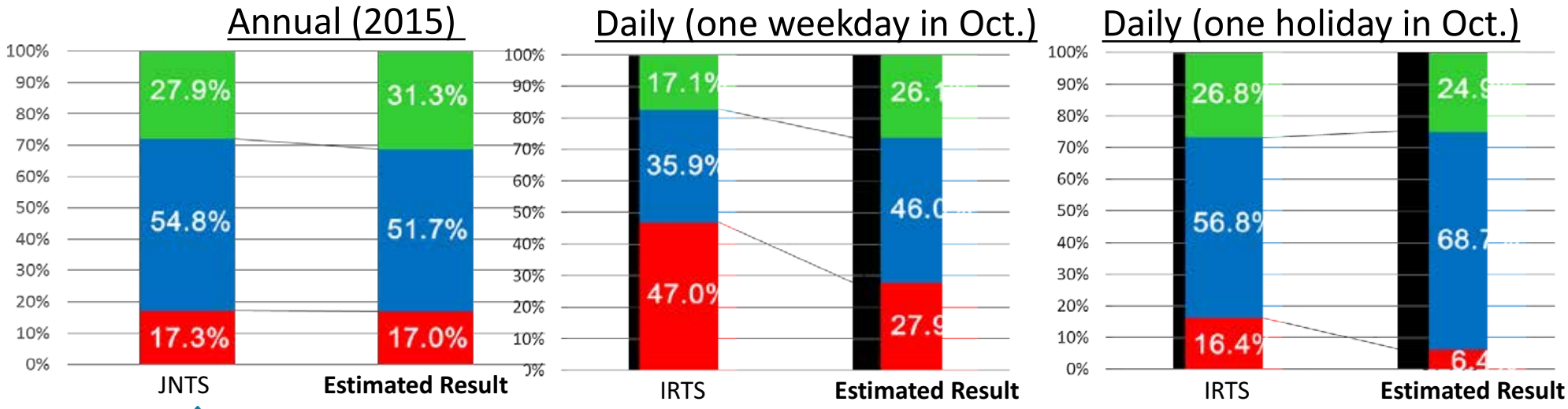


- In spite of weekday, # Business passengers ＜ # Sightseeing passengers
- Overestimation to Tokyo ⇔ Underestimation near Aomori

4. Consistency Check

Trip purposes share of estimated results

⇔ Trip purposes share of other inter regional data



Japan National Tourism Survey data (JNTS)

Data of inter regional questionnaire survey every 3 month
(trip-purposes, residential Pref., # travel, travel place)

The 3rd data sources

The estimated share of trip purposes fit well for annual aggregation, but are not close to the share of IRTS for weekday or holiday.

5. Clustering date by residuals $\hat{\varepsilon}$

Grouping 915 days' residuals $\hat{\varepsilon}$

$$\hat{\varepsilon}_{h,s,d} = Q_{h,s,d} - (\hat{V}_{h,B,d}R_{h,s,B} + \hat{V}_{h,S,d}R_{h,s,S} + \hat{V}_{h,Pr,d}R_{h,s,Pr})$$

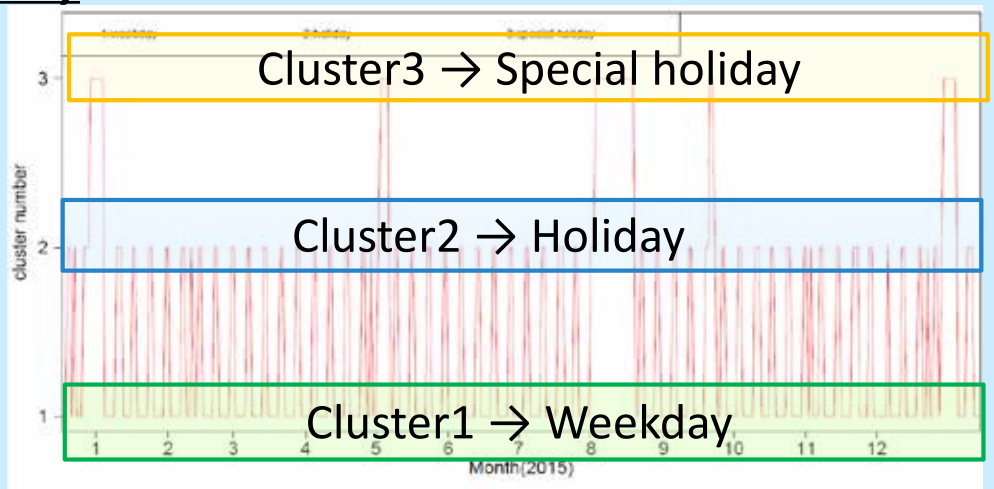
Cluster analysis by k-means (k=3)

cluster 1
−25664～+34861
cluster 2
−24408～+26093
cluster 3
−45781～+49121

Residual range (person/day)

915 days (2014/3~2016/8)

Result in 2015



Cluster1 → Weekday

Cluster2 → Holiday

Cluster3 → Special holiday

Goodness-of-fit for special holidays (Cluster3) are worse than that of normal days (Cluster1 and Cluster2)

6. Conclusions and Future works

Conclusions

- Propose the data fusion methodology for estimating the trip purpose
  - Trip generations by purposes：
    ≫ substantially appropriate result
    （According to the coefficient of determination and Consistency Check ）
  - Inter regional passenger flow by each trip purpose ：
    ≫Unreasonable results in some local area departure(e.g. Aomori)
    ≫Problem of goodness-of-fit for special holidays

Future works

- Improving the model, constructing highly accurate model
  - Integrating other inter-regional data（e.g. Japan National Tourism Survey data ）

Appendices

Appendix 1:
Map of Japan

Appendix 2:
Re-estimation of excluding Tokyo residence and staying
（46 Pref. × 46 Pref.）

Appendix 3:
Trip concentration estimation
~using share of selecting residential Pref. by each trip purpose obtained from IRTS~

Appendix 4:
Integrating 3 data sources (MSS, IRTS, JNTS)
~Maximum likelihood estimation~

Thank you for your attention!

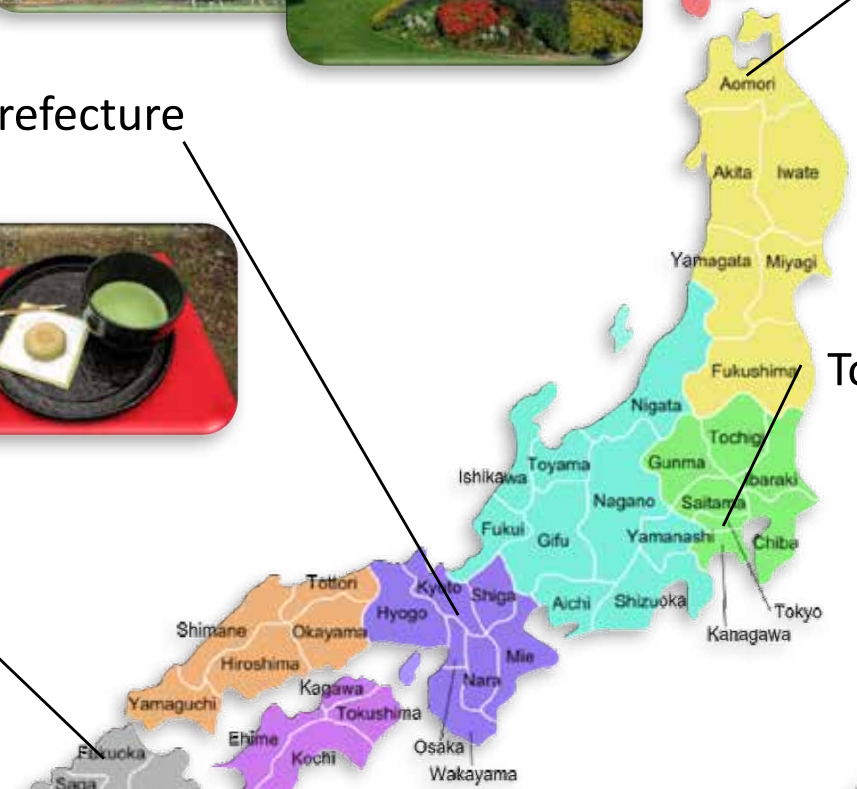
a.suzuki@plan.cv.titech.ac.jp

Appendix 1: Map of Japan

Hokkaido Prefecture

Aomori Prefecture

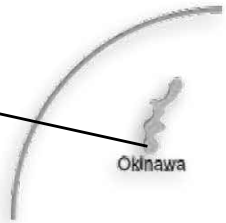
Kyoto Prefecture



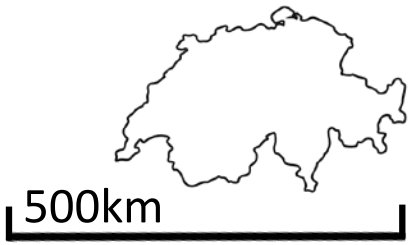
Tokyo Prefecture

Fukuoka Prefecture

Okinawa Prefecture



Size of Switzerland



Appendix 2

RE-ESTIMATION OF EXCLUDING TOKYO RESIDENCE AND STAYING（46 PREF.×46 PREF.）

Issues in the previous estimated results

Unreasonable result

- Only sightseeing passengers in “Aomori/New Year’s day” case
- In weekday, # Business passengers ＜ # Sightseeing passengers in “Aomori/weekday in Oct” case

Causes

◆In relation to Explanatory Variable (share of selecting staying Pref. by each trip purpose obtained from IRTS $\boldsymbol{R}$)

- Possibility of “Multicollinearity” among $\boldsymbol{R}$
- Possibility of $\boldsymbol{R}$ deviation to one Pref.
- Assumption”$\boldsymbol{R}$ is unchanged” is very strong

◆In relation to data

- Problem of IRTS accuracy（few samples between one OD）

Re-estimating inter regional passenger flow excluding Tokyo arrival and departure

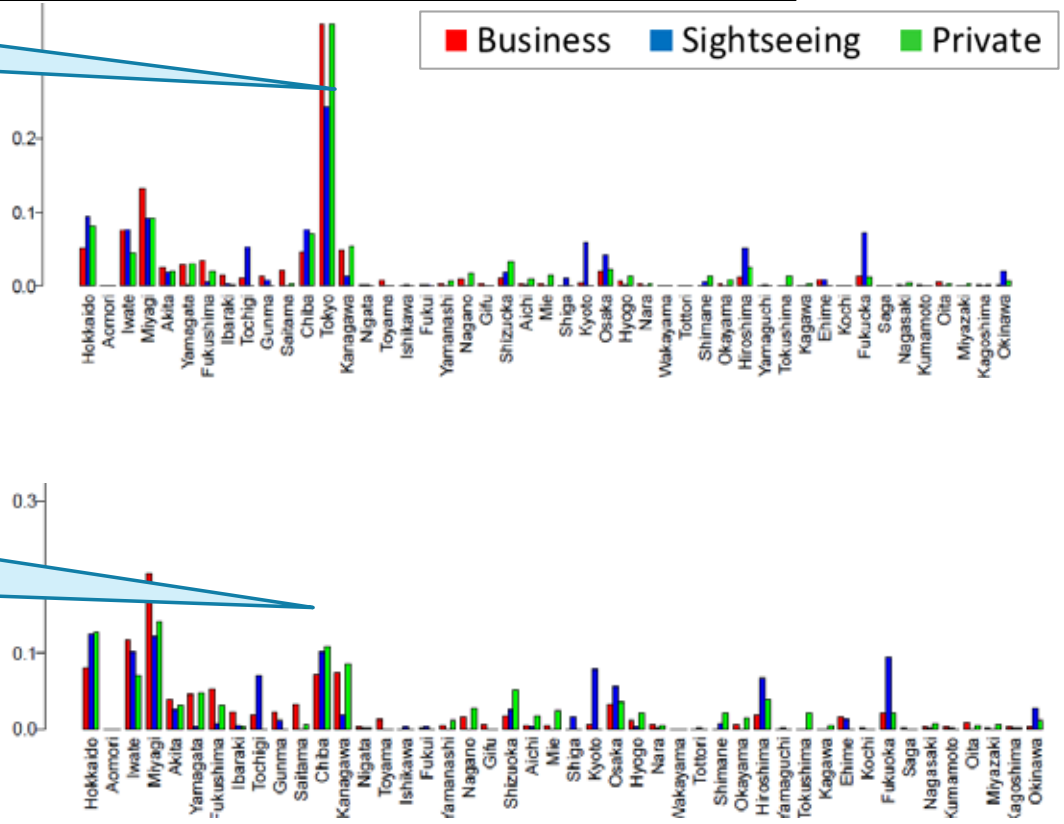
High share of selecting Tokyo in all purposes

Estimation by conforming with high share (Tokyo) (error in other Pref. is big)

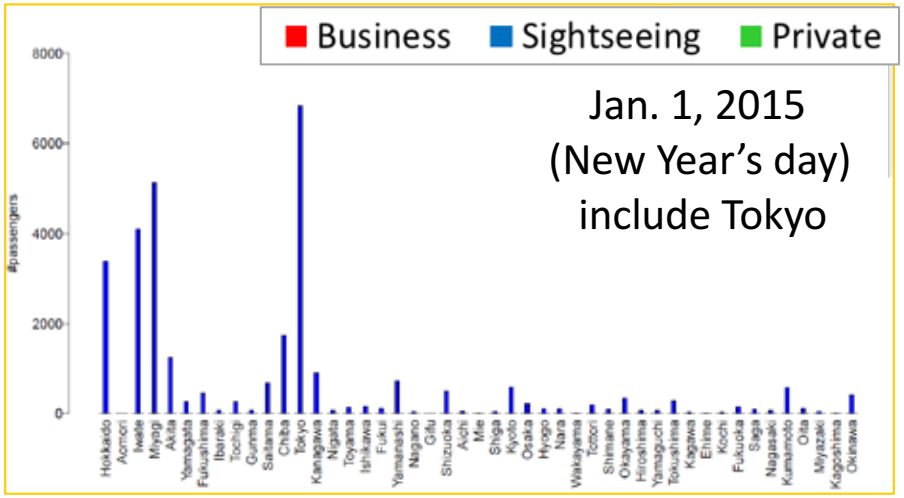
Re-estimate excluding Tokyo residence and staying (46 Pref. × 46 Pref.)

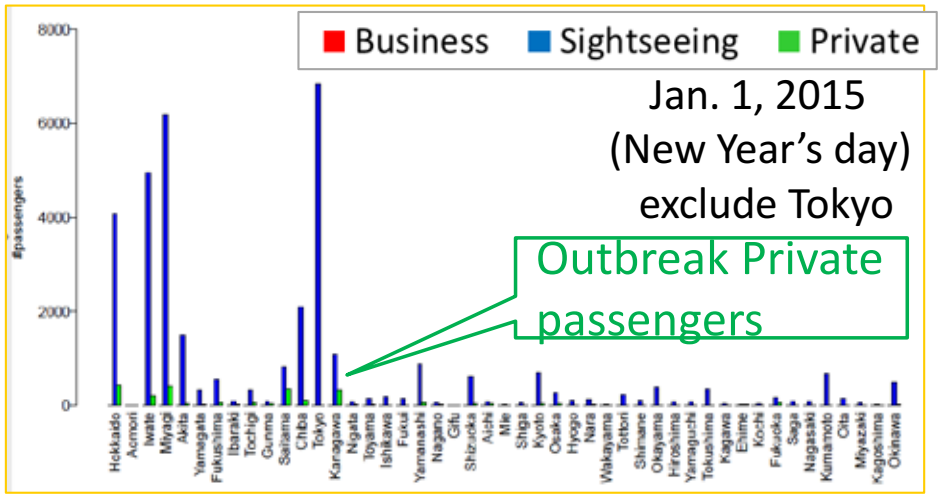
Add Tokyo residence and staying (first estimation result) to Re-estimation result

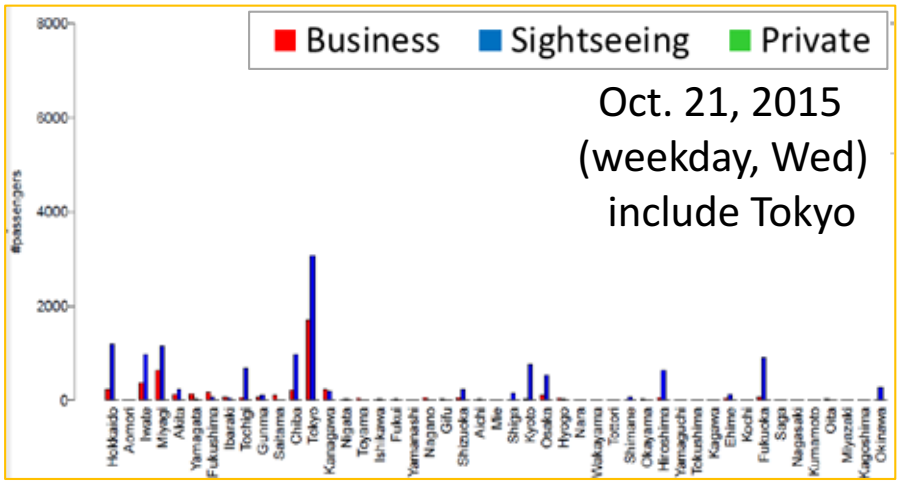
R(residence: Aomori, weekday in Oct)

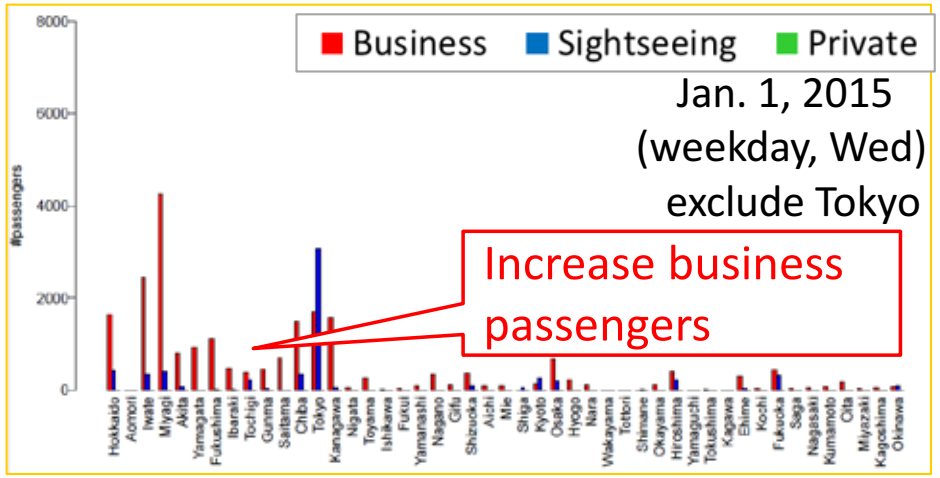


Re-estimation results





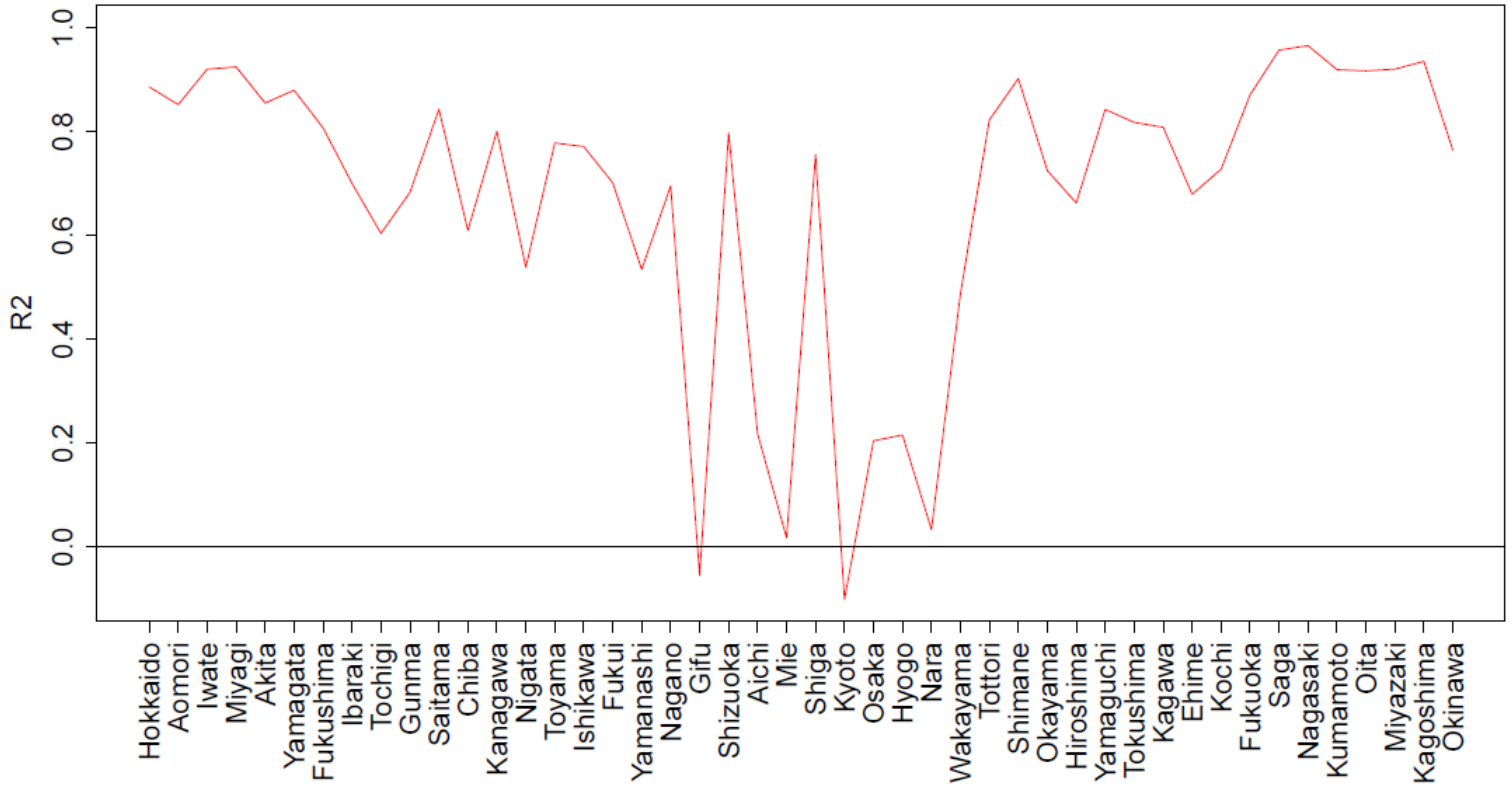




Some changes are occurred in "excluding Tokyo calculation"

Goodness of fit of the model
excluding Tokyo arrival and departure

The coefficient of determination of Re-estimation excluding Tokyo residence and staying by 46 residence Pref.（average in 915 days)



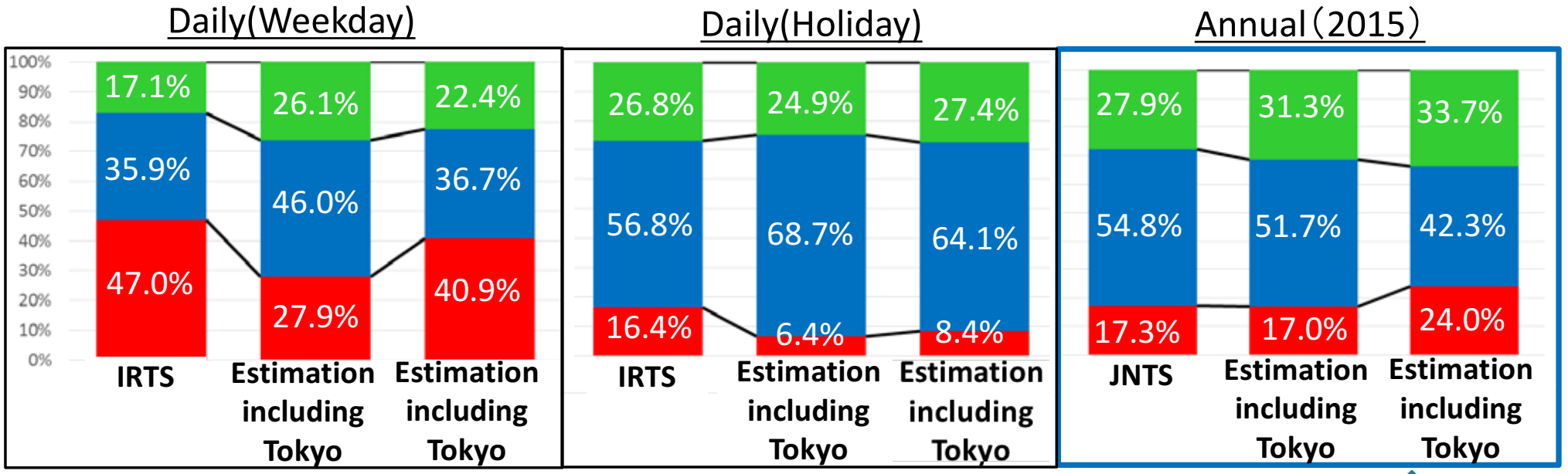
The coefficient of determination of Kyoto, Gifu and so on is small

→High share of selecting another Pref.?

Consistency Check

Trip purposes share of estimated results

⇔ Trip purposes share of other inter regional data



Japan National Tourism Survey data (JNTS) 旅行・観光消費動向調査

Data of inter regional questionnaire survey every 3 month

(the number of travel, travel date, travel place)

The share of estimated results excluding Tokyo fit well for one day aggregation, but are not close to annual aggregation

Appendix 3

TRIP CONCENTRATION ESTIMATION

(USING SHARE OF SELECTING RESIDENTIAL PREF. BY EACH TRIP PURPOSE OBTAINED FROM IRTS)

Model Formulation (regression model)

Trip purposes include (1)Business, (2)Sightseeing and (3)Private

Business | Sightseeing | Private

$$Q_{h,s,d} = V_{s,B,d} R_{h,s,B} + V_{s,S,d} R_{h,s,S} + V_{s,Pr,d} R_{h,s,Pr}$$

(h: residential Pref., s: staying night Pref., d: date, B : Business purpose, S : Sightseeing purpose, Pr : Private purpose) ※only for accommodation passengers

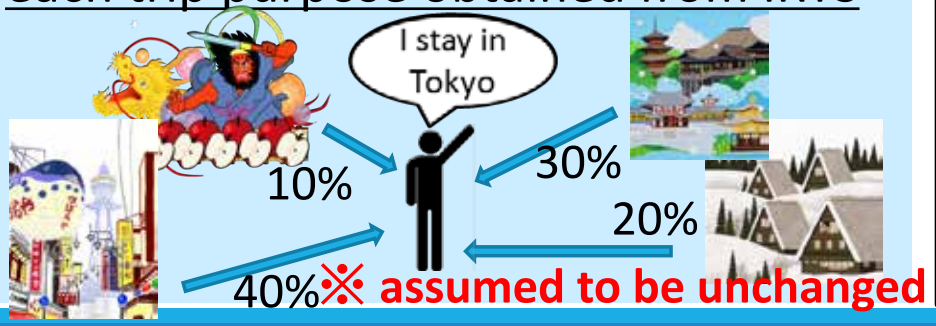
Response Variable Q: # staying people from MSS at 4 a.m.

(Assumption: staying place at 4 a.m. is accommodation place)

Explanatory Variable from IRTS

$\boldsymbol{R}$ $(0 \leq R \leq 1)$

share of selecting **residential Pref.** by each trip purpose obtained from IRTS



※ **assumed to be unchanged**

Parameter to be estimated

(constrained least squares method)

$\boldsymbol{V}(\geq 0)$

passenger for each trip purpose from each **staying night Pref.**

e.g.)
for sightseeing
Stay in Tokyo

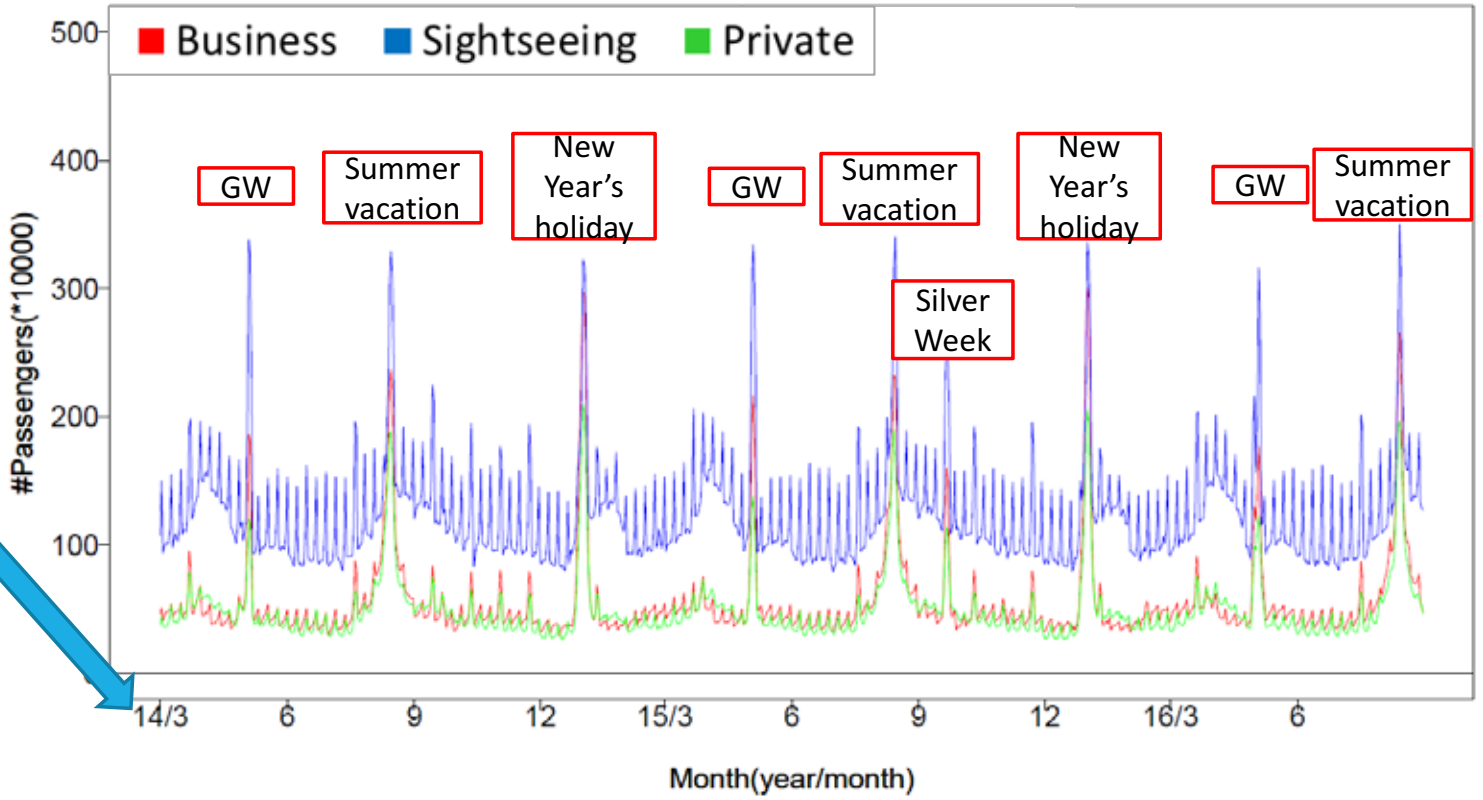
3. Estimated "Total trip concentrations" $\widehat{Vs}$ across different days

Total trip concent-ration by purposes

$$\sum_s V$$

Data
IRTS:2010
MSS : Mar, 2014
~ Aug, 2016
(915 days)

Spatial Unit
Prefecture



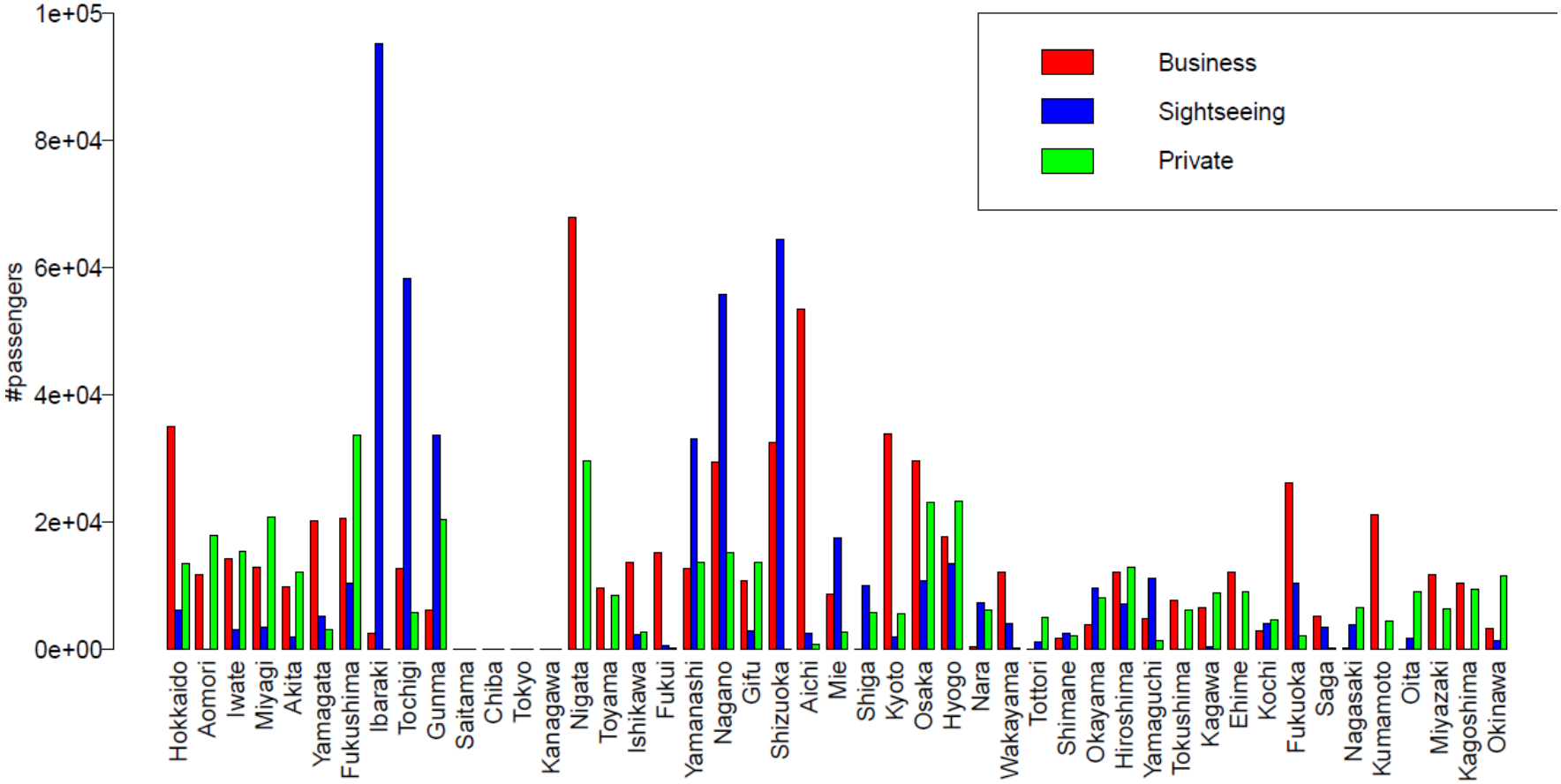
The coefficient of determination is 0.57(Summer vacation) ～0.88(around survey date of IRTS)

In long vacation, there are many passengers for not only sightseeing and private, but also business.

3. Estimated inter regional passenger flow

~ Residence : Tokyo(metropolitan), date : New Year's day ~

Result of Jan. 1, 2015

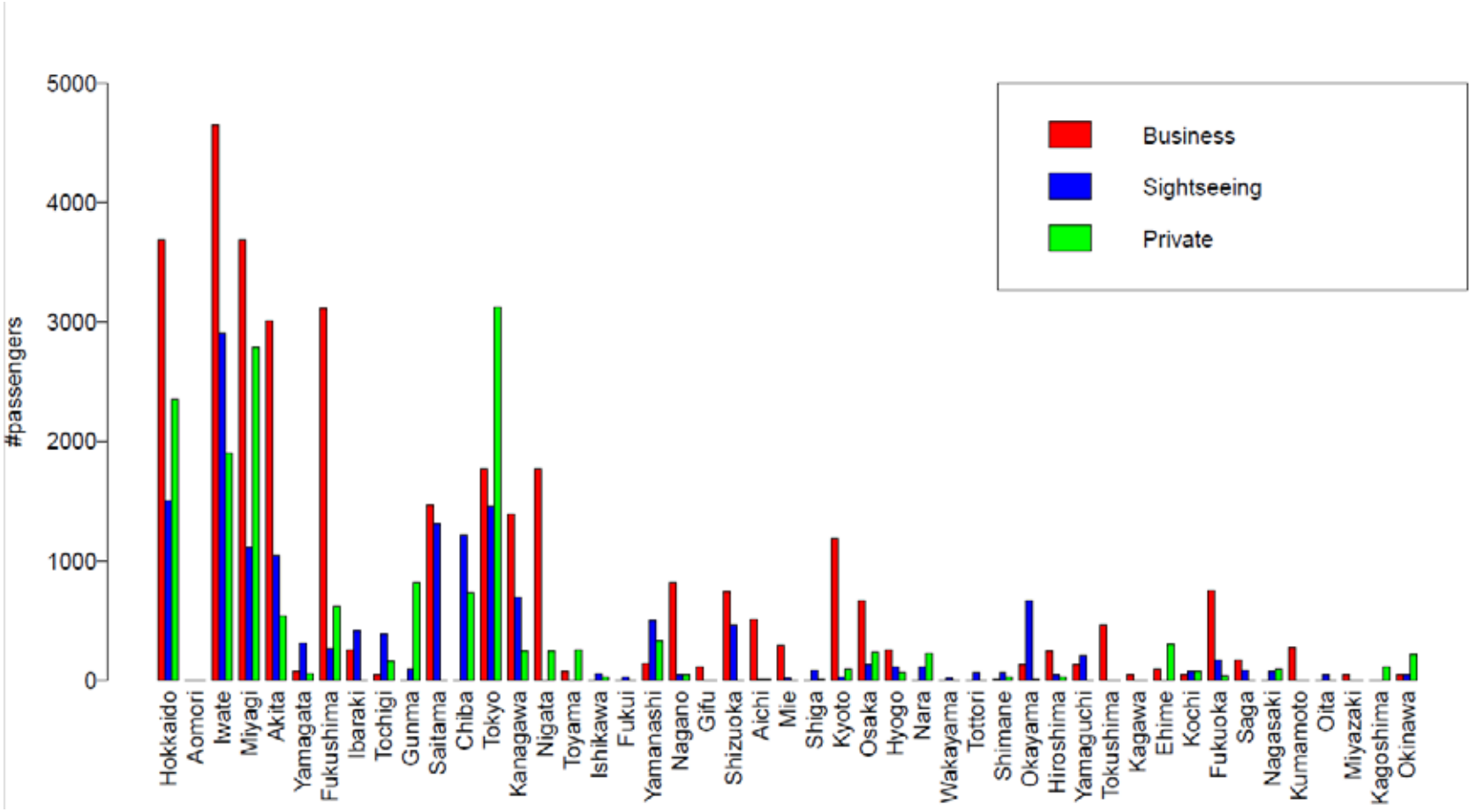


Many business passengers in spite of New Year's day

3. Estimated inter regional passenger flow

~ Residence : Aomori(rural Pref.), date : New Year's day ~

Result of Oct. 21, 2015 (Wed)



Unreasonable results in both metropolitan and local area departure

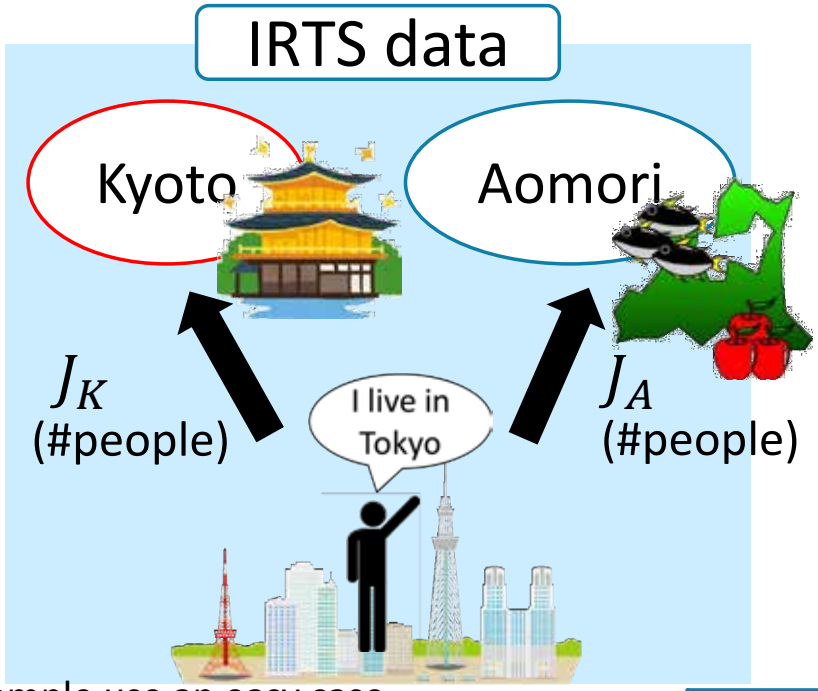
Appendix 4

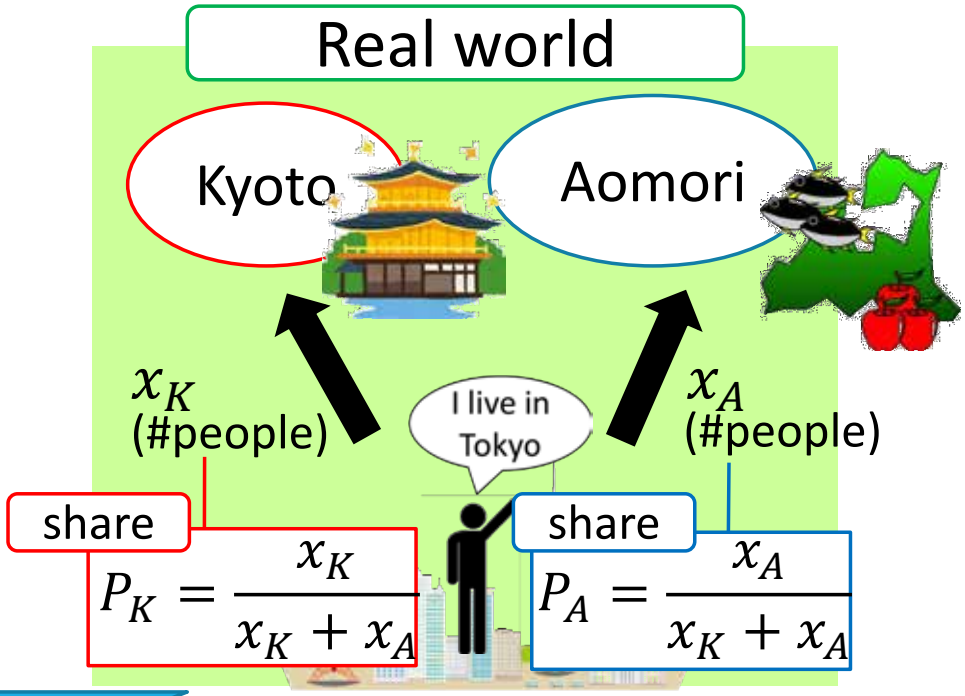
INTEGRATING 3 DATA SOURCES(MSS, IRTS, JNTS)

~MAXIMUM LIKELIHOOD ESTIMATION~

Model Formulation (Log-likelihood function)

Ex) Residential Pref. = Tokyo, Trip-purpose = sightseeing, in one day,





※Example use an easy case
(only 3 Pref.: Tokyo, Kyoto, Aomori)

The probability of getting survey results of IRTS (J_K, J_A) ⋯ $(P_k)^{J_K} \times (P_A)^{J_A}$

Log-likelihood function ⋯ $\boldsymbol{L(x) = J_K \log(P_K) + J_A \log(P_A)}$

with some constraints

Model Formulation (The constraints)

- Constraints 1: sum up estimated results by purposes = ***MSS***

$x_{Tokyo,Kyoto,Business}$

$x_{Tokyo,Kyoto,Sightseeing}$

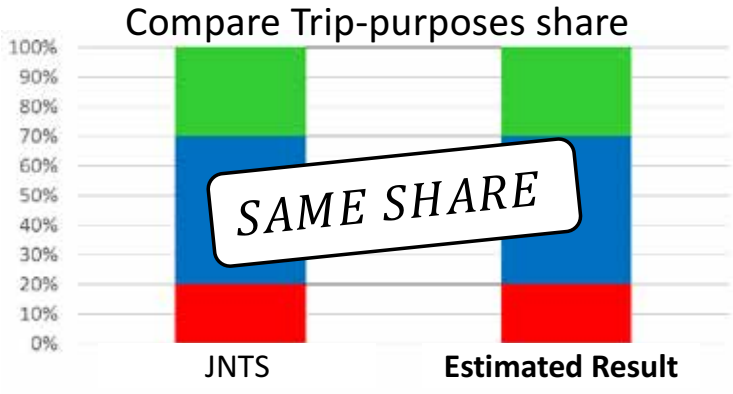
$x_{Tokyo,Kyoto,Private}$

MSS

$Q_{Tokyo,Kyoto}$

$$\sum_{p \in P} x_{h,s,p} = M_{h,s}$$

- Constraints 2: Trip-purposes share of estimated results every 3 month = ***Share of trip purposes from JNTS***



$$\frac{\sum_{s \in Z} \sum_{d \in D} x_{h,s,d,p}}{\sum_{p \in P} \sum_{s \in Z} \sum_{d \in D} x_{h,s,d,p}} = \frac{T_{h,s,d,p}}{\sum_{p} T_{h,s,d,p}}$$

Model Formulation (Overall - Tentative)

- Thinking of all date (d), all residences (h), all staying Pref. (s)
- Estimation by 3 months (the data term of JNTS) and residential Pref. (47 Pref.)

$$max. \sum_{s\in Z}\sum_{d\in D}\sum_{p\in P} J_{h,s,p} \ln\left(\frac{x_{h,s,d,p}}{\sum_{s\in Z} x_{h,s,d,p}}\right)$$

Sub.to

$$\sum_{p\in P} x_{h,s,d,p} = M_{h,s,d} \qquad \forall h, \forall s, \forall d$$

$$\frac{\sum_{s\in Z}\sum_{d\in D} x_{h,s,d,p}}{\sum_{p\in P}\sum_{s\in Z}\sum_{d\in D} x_{h,s,d,p}} = \frac{T_{h,s,d,p}}{\sum_{p} T_{h,s,d,p}} \qquad \forall h, \forall s, \forall p$$

x :estimated parameter,
M :# staying people from MSS at 4 a.m,
J : # people from JNTS,
T :# people from JNTS,
D:dates in 3 month
Z:47 Pref.

P:All purposes,
h: residential Pref.
s: staying night Pref.
d: date,
p : purposes
(Business, Sightseeing, Private)