12th Behavior Model Summer School

Combined estimation of activity generation models incorporating unobserved small trips using probe person data

Sep. 22nd, 2013

The University of Tokyo
Sohta Itoh

Contents

- Research background
- Comparison between PT and PP data
- Combined estimation model
- Correcting sampling bias
- Conclusion

Research background

1960s **Person Trip survey** (Paper-based)

(1955 CATS, 1967 Hiroshima)

1980s **Activity based model** – disaggregate data

2000s **Probe Person survey** (GPS-based)

(Zitto and D'este, 1995; Murakami and Wagner, 1999; Asakura and Hato, 2004; Hato et al., 2006; Stopher et al., 2011)

Non-response activities

Short trips and activities are often underreported

(Wolf et al., 2001; Bricka and Bhat, 2006; Itsubo and Hato, 2006)

↓

Non-response bias

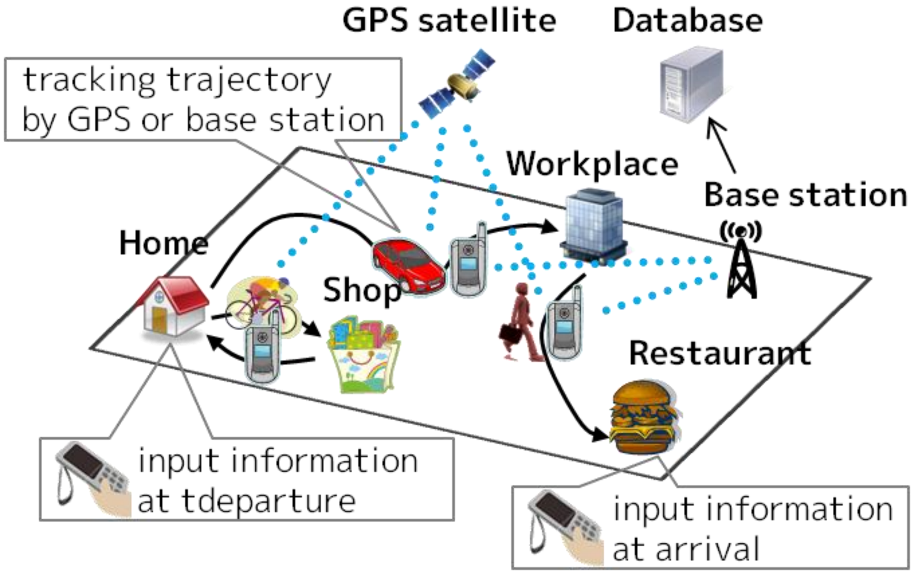
Changes of activity patterns

- Aging society
- Inner-city problems

↓

Short activities becomes important

Methods of PP survey



PP Survey ex) observe 3 activities

PP data

GPS
- Timestamp
- Latitude
- Lontitude

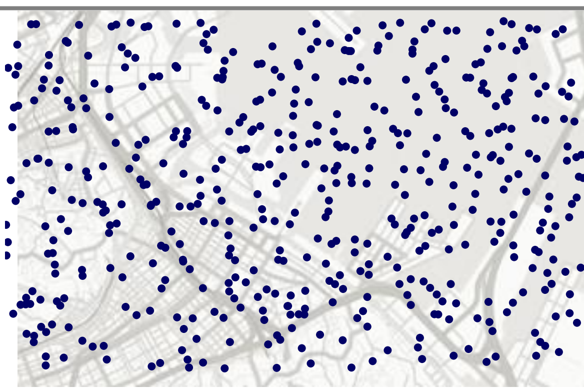
Web diary
- Trip purpose
- Transportation mode

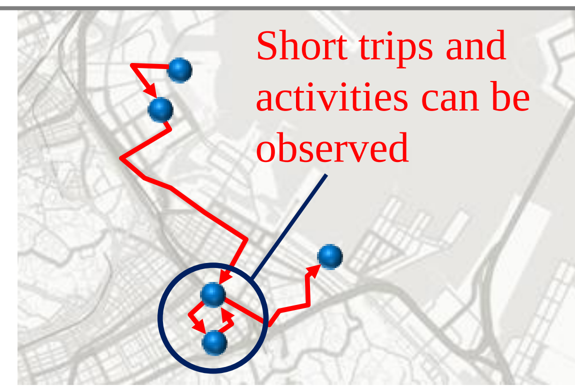
+ personal information

2010/7/5 17:01:52
Walk
Car
Bike
Motorcycle
Bus
Train
(PM)

Comparison between survey data

PT survey data	Massive location data	PP survey data
Paper-based (Rely on respondents' memories)	GPS (Automatical but fragmentary)	GPS (Automatical) **+ Web diary**
Zone-based	Dot data (High-resolution)	Dot-based (High-resolution)
Large sample	Large sample	Small sample





Combined Estimation using both PT and PP data

Outline of PP and PT survey data

- Both data are obtained in Yokohama, Japan
- Respondents are resided in Yokohama

■PT survey

Surveillance period	2008/10 - 2008/11 (each respondent answers his/her travel behavior of 1 day in surveillance period)
Method	Paper questionnaire
The number of all trips	1,906,032 trips
The number of trips in Yokohama	253,737 trips

■PP survey

Surveillance period	35 days (2010/07/05 - 2010/08/08)
Survey methods	Probe Person survey with GPS cell phone + Web diary
The number of samples	40 people
The number of Trips	3,617 trips
The number of location data	789,074 points

Elementary analysis

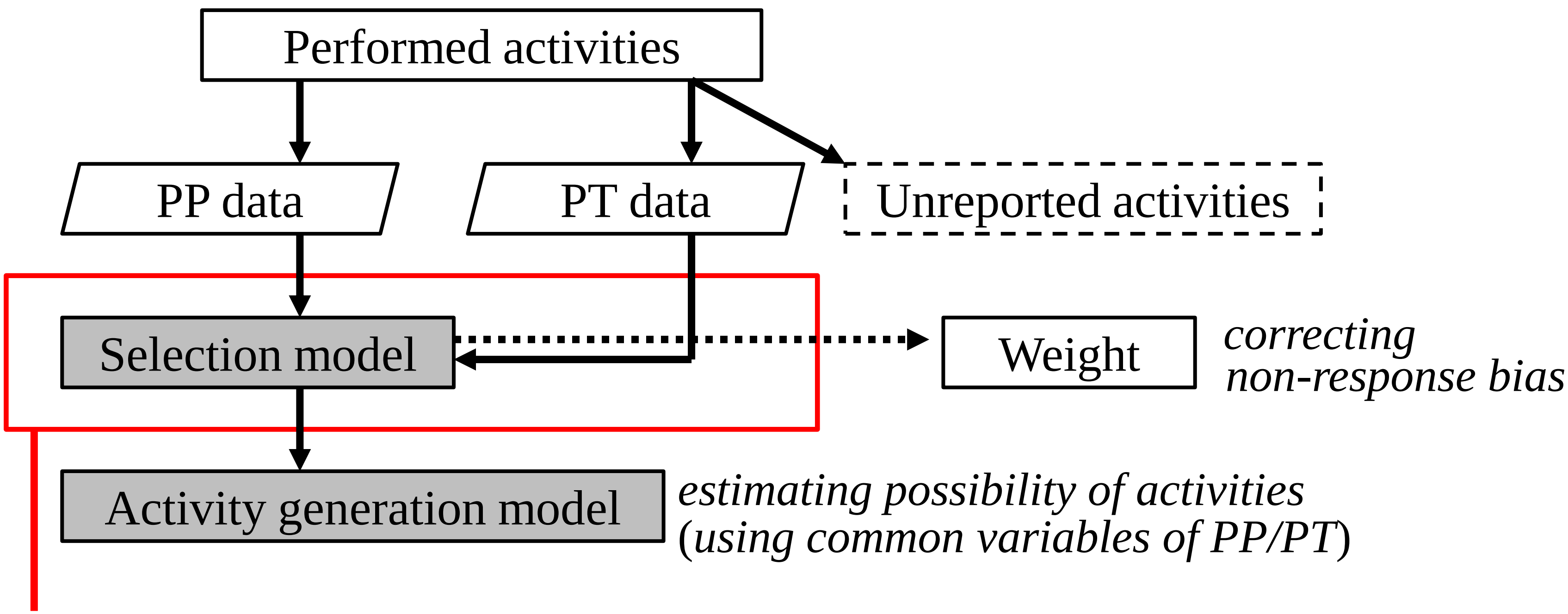
・In almost all of categories, the number of activities of PT data is smaller than that of PP data

	The number of activities			The sum of activity duration		
	mean		t-statistics	mean (min.)		t-statistics
	PT	PP		PT	PP	
age 20s	1.26	1.39	2.62*	457.0	544.0	5.95*
age 30s	1.40	1.60	3.12*	426.9	389.0	1.84
age 40s	1.53	1.74	2.63*	445.0	288.5	8.60*
age 50s	1.55	1.80	1.98*	412.2	325.9	3.73*
age 60s+	1.56	1.58	0.19	233.3	298.2	1.63
male	1.49	1.78	4.86*	459.7	497.9	2.90*
female	1.43	1.43	0.00	309.1	281.7	2.14*
total	1.46	1.60	5.39*	383.0	389.5	0.65

* : reject the null hypothesis of no difference between the mean of PT data and that of PP data at 5% significant level

Estimation model framework

- It is assumed that PP data does not have unreported activities.
- If missing activities have some characteristics in common, sampling bias affects the estimation result



Detecting the factors influencing the propensities to record activities

Introducing selection model

Apply Tobit selection model to activity generation and its observation

● Activity generation model

$$y^*_{in1} = \beta_1 x_{in1} + \varepsilon_{in1}$$

Latent variable **about activity generation** of individual *i* and zone *n*

$$\begin{cases} y_{in1} = 1 & \text{if } y^*_{in1} > 0 \\ y_{in1} = 0 & \text{if } y^*_{in1} \le 0 \end{cases}$$

generate / not generate

x_{in1} : explanatory variables of individual *i* and zone *n*
ε_{in1} : error term of individual *i* and zone *n*

● Selection model

$$y_{in2} = \beta_2 x_{in2} + \varepsilon_{in2}$$

Latent variable **about observation** of individual *i* and zone *n*

$\text{if } y_{in2} > 0$ — y_{in1} is observed

$\text{if } y_{in2} \le 0$ — y_{in1} is not observed

x_{in2} : explanatory variables of individual *i* and zone *n*
ε_{in2} : error term of individual *i* and zone *n*
y_{in2} : unobserved variable of individual *i* and zone *n*

$$\begin{pmatrix} \varepsilon_{in1} \\ \varepsilon_{in2} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1 \\ \rho\sigma_1 & 1 \end{pmatrix} \right)$$

Introducing selection model

- Activity generation model $y^*_{in1} = \beta_1 x_{in1} + \varepsilon_{in1}$
- Selection model $y_{in2} = \beta_2 x_{in2} + \varepsilon_{in2}$

Expected value of latent variable y_{in1} after considering selection bias

$$E(y_{in1} \mid y_{in2} > 0) = \beta_1 x_{in1} + E(\varepsilon_{in1} \mid \varepsilon_{in2} > -\beta_2 x_{in2})$$

$$= \beta_1 x_{in1} + \boxed{\rho \sigma_1 \frac{\phi(\beta_2 x_{in2})}{\Phi(\beta_2 x_{in2})}}$$

Correction term
(apply only for PT data)

Φ : cumulative distribution function of the standard normal distribution
φ : probability density function of the standard normal distribution

Estimation results

Independent variables	The normal activity generation model			The sample selection model		
	Parameter	t score		Parameter	t score	
For activity generation model						
Constant	-1.902	-76.64	*	-1.808	-79.24	*
Male	0.091	12.59	*	0.069	7.51	*
Age ≧ 60	-0.116	-15.37	*	-0.106	-10.89	*
Single-member household	0.090	8.79	*	0.100	7.73	*
Car ownership	-0.003	-0.42		-0.002	-0.17	
Distance from home (km)	-0.108	-98.83	*	-0.117	-58.97	*
Distance from workplace (km)	-0.025	-43.52	*	-0.028	-35.70	*
Store space (ha) [1)]	0.043	71.31	*	0.035	39.55	*
γ	0.125	5.09	*	-	-	
ρ	-	-		0.435	16.94	*
For selection model						
Male	-	-		0.466	14.18	*
Age 20-39 years	-	-		-0.545	-7.07	*
Age ≧ 60	-	-		0.355	4.20	*
Distance from home (km)	-	-		0.071	0.66	
Distance from workplace (km)	-	-		0.020	0.23	
Stay Duration (min.)	-	-		0.044	4.99	*
μ	-	-		3.557	17.67	*
Observations (PT)	1,780,164			1,780,164		
Observations (PP)	23,000			23,000		
Initial log-likelihood	–1,249,858			–1,249,858		
Final log-likelihood	–65,013			–64,272		
Rho-squared $\bar{\rho}^2$	0.948			0.949		

- Not relevant; * Significant at 5% level.

1) : The sum of space about retail stores in the zone

Following attributes associate with activity under-reporting at the significant level

- male
- stay duration
- age 20-39 years
- age 60+

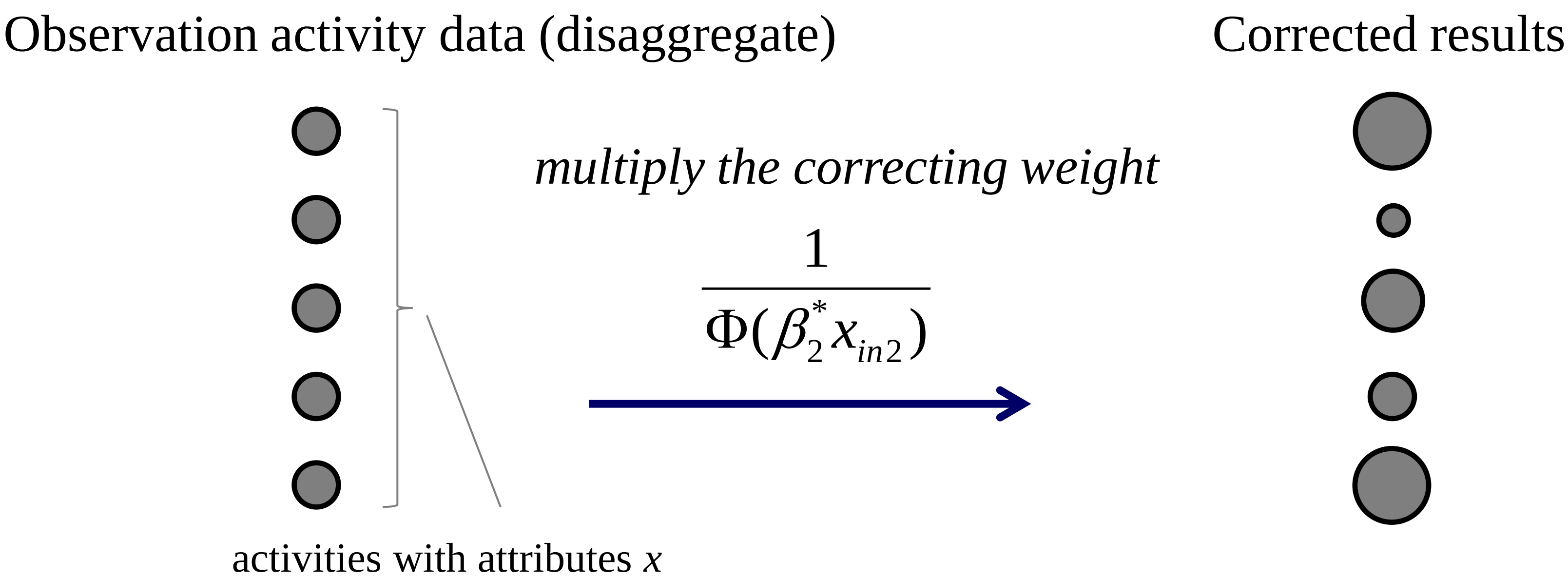
Correcting sampling bias

To correct the bias, the inverse of observation probability is considered the weight as:

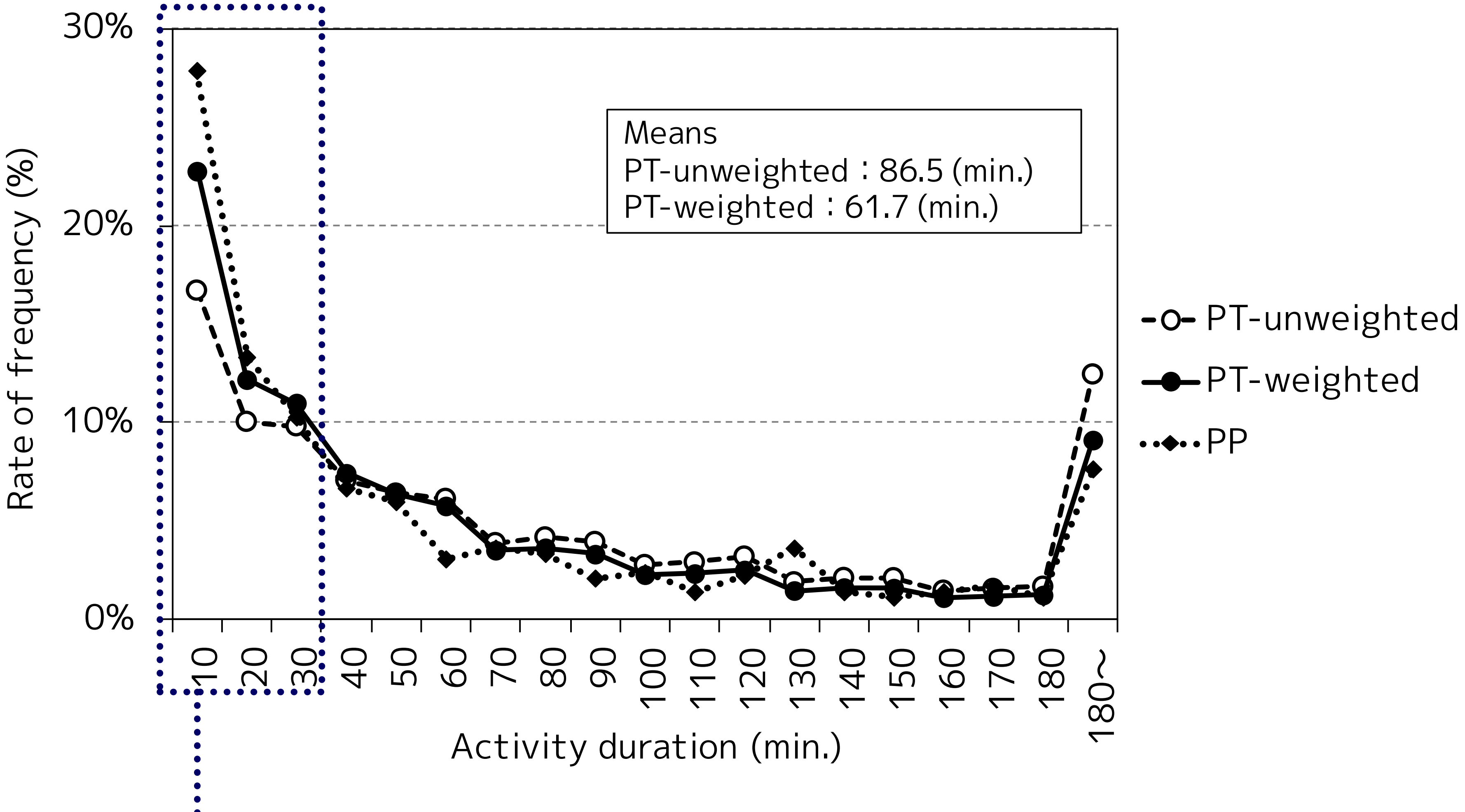
$$w_{in} = \frac{1}{p(y_{in2} > 0 \mid x_{in2})} = \frac{1}{\Phi(\beta_2^* x_{in2})}$$

β^* : the parameter estimated in the model

comes from the estimation results

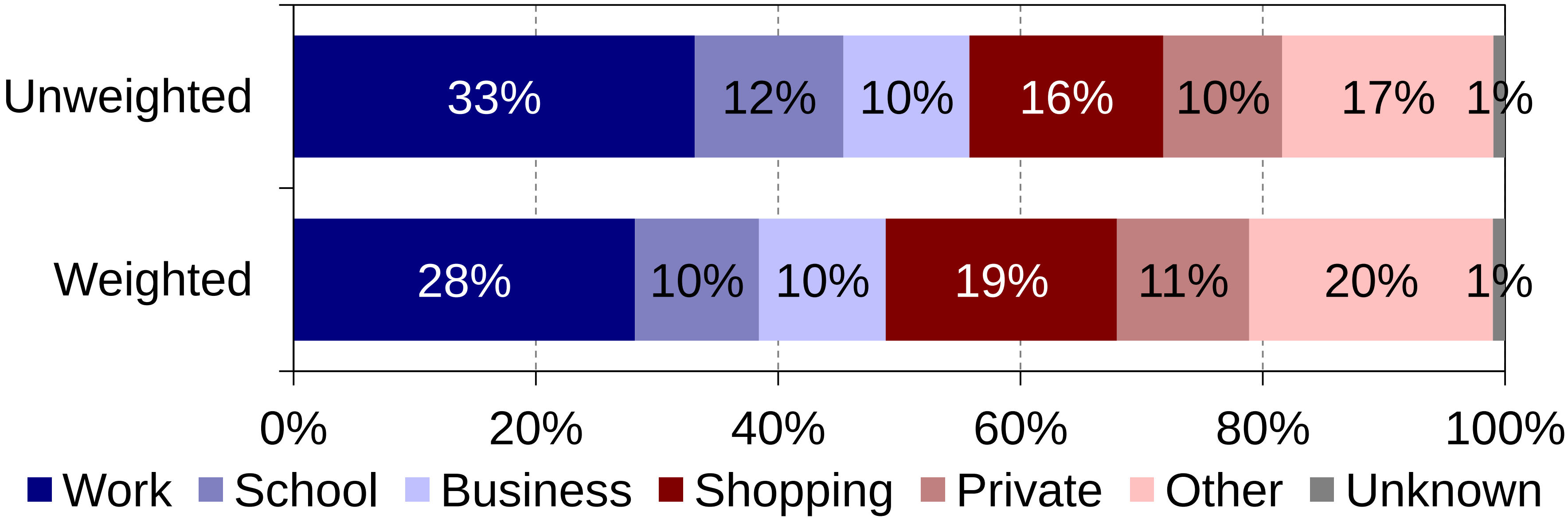


Correcting sampling bias



The rate of frequency of weighted PT is similar to PP, which represents the bias of short activities is corrected

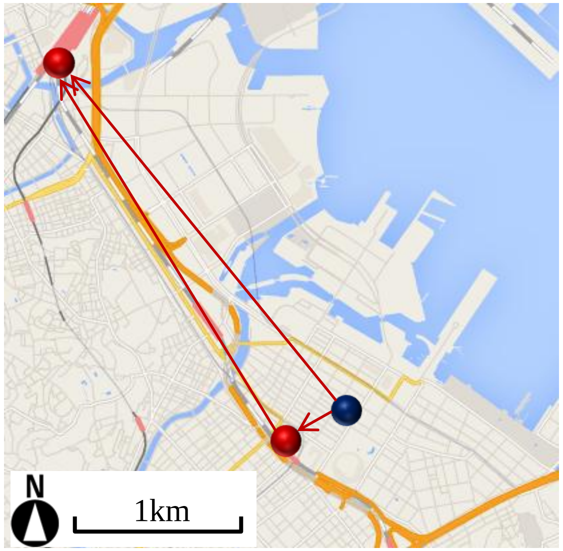
Correcting sampling bias



The rate of discretionary activities is expanded by weighting.

Work 240 min. — Shopping 15 min. — Private 30 min.

Adding activities stochastically



Conclusion

Comparison between PT and PP

We have discussed the advantages of both new GPS-based PP surveys and conventional PT surveys

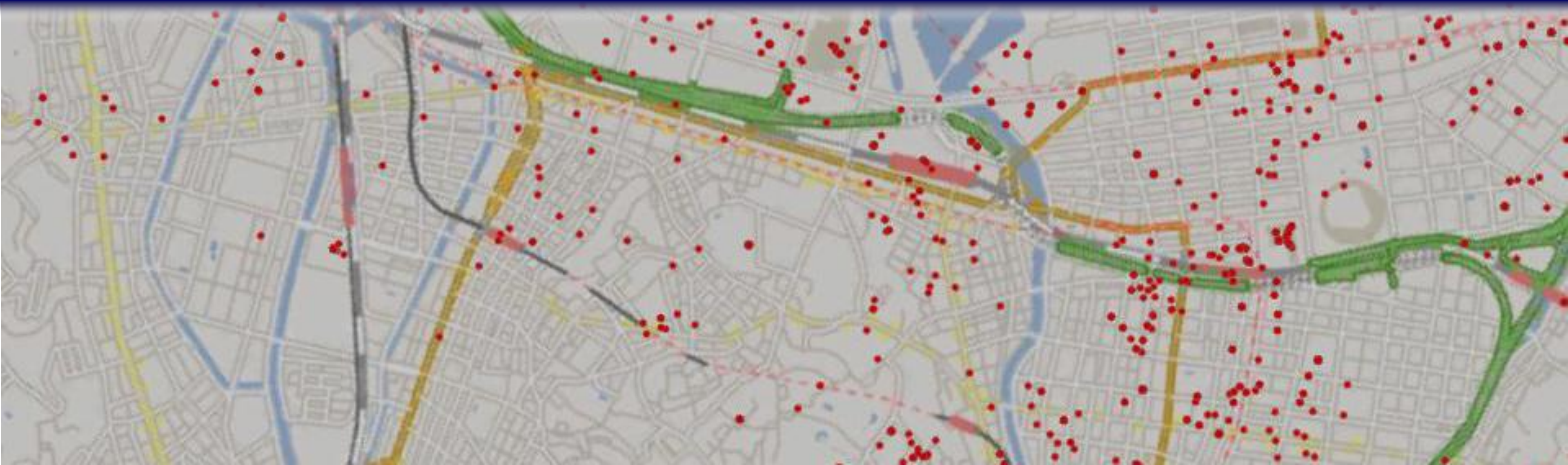
Combined estimation using PT and PP data

Introducing the selection model, we show several demographic attributes and activity characteristics associate if activities are missed or not and consider the selection bias

Correcting the sampling bias

By multiplying the inverse of probabilities of observation obtained from the selection model, the bias is appropriately assessed and corrected

Thank you for your attention!



References

▪Asakura, Y., and Hato, E. (2004) Tracking Survey for Individual Travel Behavior Using Mobile Communication Instruments, Transportation Research C, 12, 273-291.
▪Bricka, S., Bhat, C. (2006) Comparative Analysis of Global Positioning System-Based and Travel Survey-Based Data, Transportation Research Record No. 1972, 9-20.
▪Brog, W., Erl, E., Meyburg, A. H., Wermuth, M. J. (1982) Problems of Nonreported Trips in Surveys of Nonhome Activity Patterns, Transportation Research Record, Vol. 891, 1-5.
▪Itsubo, S., Hato, E. (2006) A Study of the Effectiveness of a Household Travel Survey Using GPS-Equipped Cell Phones and a Web Diary through a Comparative Study with a Paper Based Travel Survey, TRB Annual Meeting in Washington DC (CDROM).
▪Kitamura, R., Bovy, P. (1987) Analysis of Attrition Biases and Trip Reporting Errors for Panel Data, Transportation Research A, 21, 287-302.
▪Kitamura, R. (1990) Panel Analysis in Transportation Planning: An Overview, Transportation Research A, 24, 401-415.
▪Hato, E., Itsubo, S., Mitani, T. (2006) Development of MoALs (Mobile Activity Loggers Supported by GPS-Phones) for Travel Behavior Analysis, TRB Annual Meeting in Washington DC (CDROM).
▪Hato, E. (2006) Evaluation of Trip-Activity Pattern Variability Using Probe Person Data, TRB Annual Meeting in Washington DC (CDROM).
▪Hato, E. (2010) Development of Behavioral Context Addressable Loggers in the Shell for Travel-Activity Analysis, Transportation Research C, 18, 55-67.
▪Murakami, E., Wagner, D. P. (1999) Can Using Global Positioning System (GPS) Improve Trip Reporting?, Transportation Research C, 7, 149-165.
▪Rubin, D. B. (1976) Inference and Missing Data, Biometrika, 63, 581-590.

References

- Sermons, M. W., Koppelman, F. S. (1996) Use of Vehicle Positioning Data for Arterial Incident Detection, Transportation Research C, 4, No. 2, 87-96.
- Sneade, A. (2011) Using Accelerometer Equipped GPS Devices in Place of Paper Travel Diaries to Reduce Respondent Burden in a National Travel Survey, 9th International Conference on Transport Survey Methods.
- Stopher, P., Greaves, S. (2010) Missing and Inaccurate Information from Travel Surveys: Pilot Results, Working paper (University of Sydney. Institute of Transport and Logistics Studies).
- Stopher, P. R., Prasad, C., Wargelin, L., Minser, J. (2011) Conducting a GPS-only Household Travel Survey, 9th International Conference on Transport Survey Methods.
- Morikawa, T. (1994) Correcting state dependence and serial correlation in the RP/SP combined estimation method, Transportation, 21, 153-165.
- Timmermans, H. J. P., Hato, E. (2009) Electronic Instrument Design and User Interfaces for Activity-Based Modeling, Transport Survey Methods keeping Up With a Changing World, Emerald Group Publishing Ltd., 437-462.
- Wolf, J. (2004) Applications of New Technologies in Travel Surveys, 7th International Conference on Transport Survey Quality and Innovation.
- Wolf, J., Loechl, M., Meyers, J., Arce, C. (2001) Trip Rate Analysis in GPS-Enhanced Personal Travel Surveys, International Conference on Transport Survey Quality and Innovation.
- Zitto, R., D’este, G., Taylor, A. P. (2007) Global Positioning System in the Time Domain: How Useful a Tool for Intelligent Vehicle-Highway Systems?, Transportation Research C, 3, 193-209.