Structural estimation for route choice models considering link specificity of measurement error variances

Yuki Oyama*, Eiji Hato

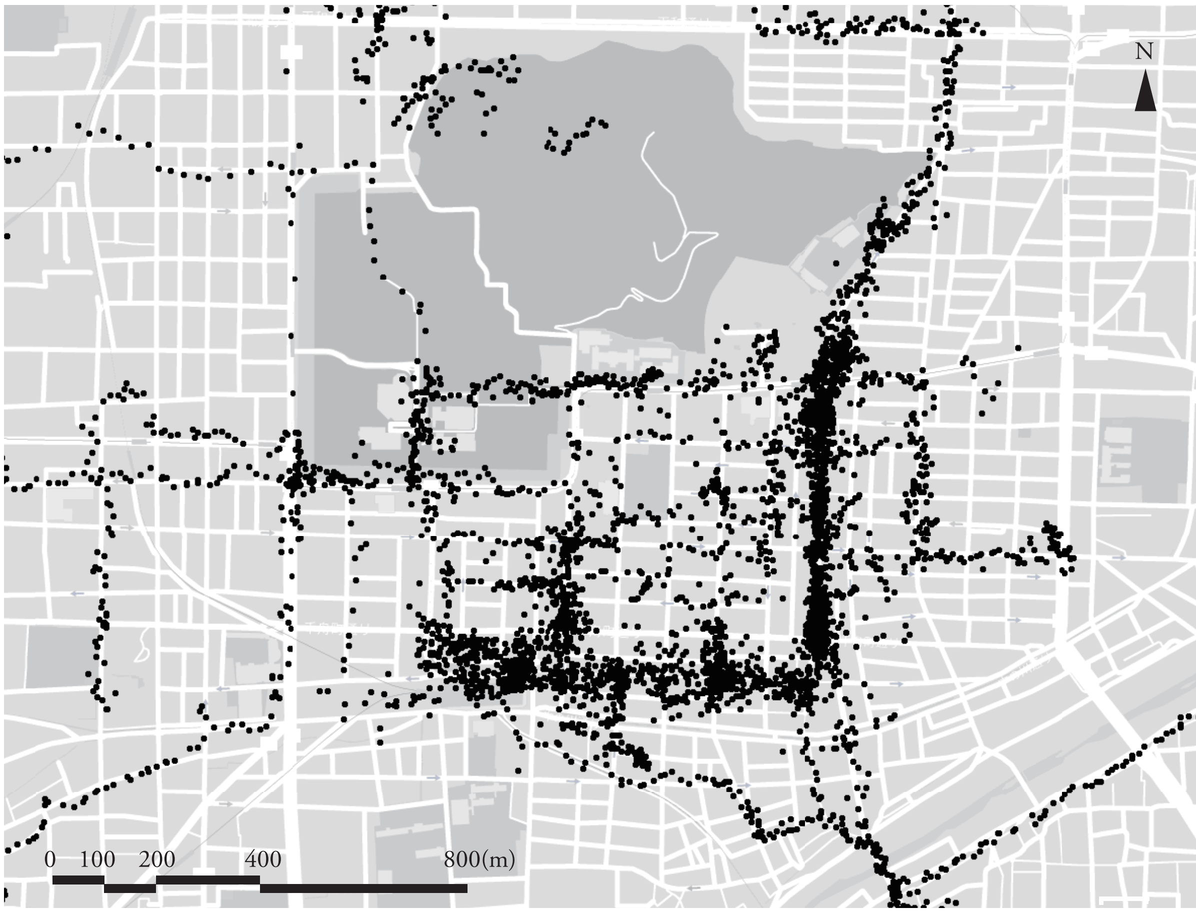
*oymyk.dom@gmail.com
Department of Architecture and Building Engineering
School of Environment and Society
Tokyo Institute of Technology

June 22, 2017

Outline

1. **Introduction**
2. **Methodology** :
   - ***Sequential link measurement model***
   - ***Structural estimation method***
3. **Numerical experiments** :
   - *Simulation data*
   - *Real pedestrian GPS data*
4. **Conclusions**

Measurement uncertainty



- **Dense network**
- **Spatial heterogeneity of Measurement errors**
  - ✓ Along river
  - ✓ Wide street
  - ✓ Narrow street
  - ✓ With arcade




Literature review

Bayesian approach [e.g., Danalet et al., 2014; Chen et al., 2015]

$$p(r|\hat{m}_{1:J};\sigma,\bar{\theta}) \propto \underline{p(\hat{m}_{1:J}|r;\sigma)}\,\underline{p(r|\mathcal{R};\bar{\theta})}$$

$p(\hat{m}_{1:J}|r;\sigma)$: Measurement equation

$p(r|\mathcal{R};\theta)$: Route choice model (prior information)

Joint estimation [Bierlaire and Frejinger, 2008]

$$p(\hat{m}_{1:J};\bar{\sigma},\theta) = \sum_{r\in\mathcal{R}} \underline{p(\hat{m}_{1:J}|r;\bar{\sigma})}\,\underline{p(r|\mathcal{R};\theta)}$$

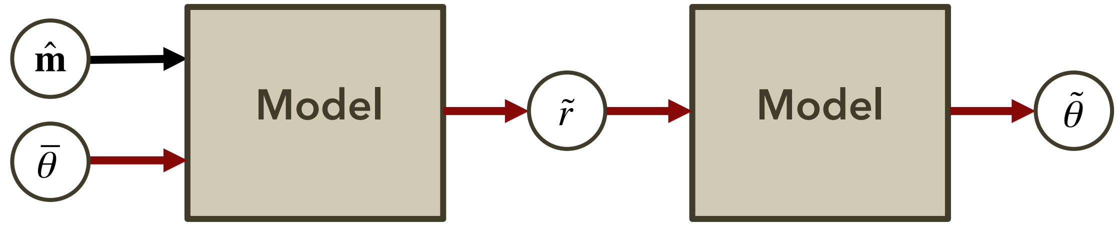
$\mathcal{R}$: Path candidate set defined by σ

Objective

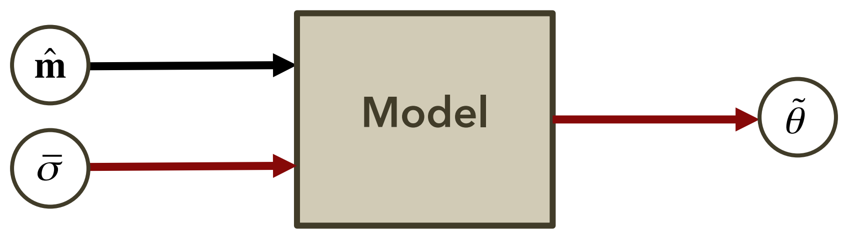
Bayesian approach:

: Biased

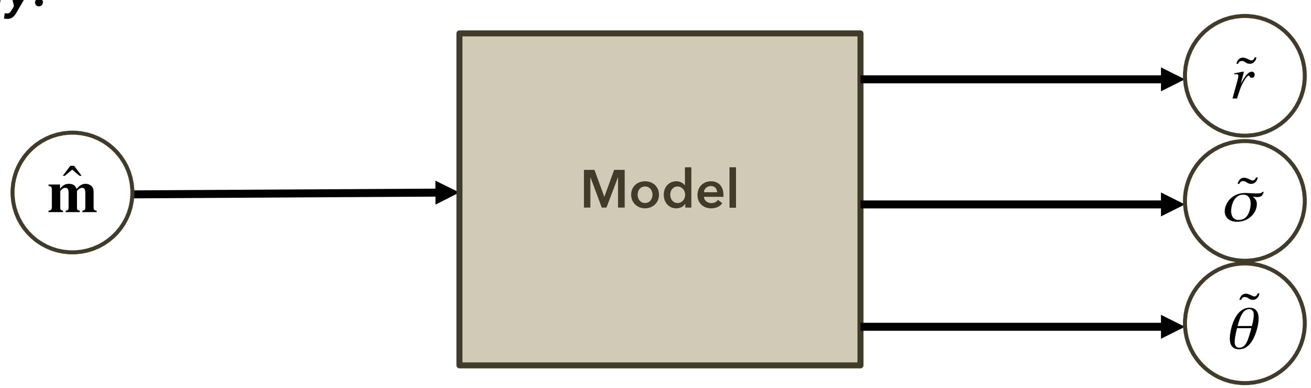
: Unbiased



Joint estimation:



This study:



Problem definition & notations

1. Network

$\mathcal{G} = (\mathcal{N}, \mathcal{A})$: Network graph

$i \in \mathcal{N}$: Node (coordinates: $x_i = \{x_{i\text{lat}}, x_{i\text{lon}}\}$)

$a \in \mathcal{A}$: Link (up/down nodes: (u_a, d_a), attributes: y_a)

2. Route choice behavior

Sequential link choices (a Markovian route choice model)

3. Observation

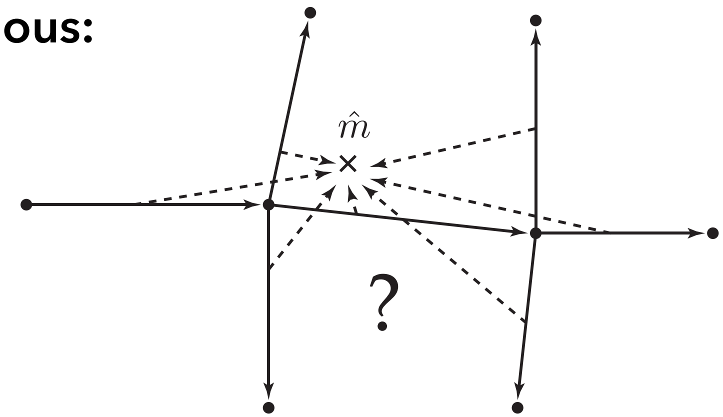
$\hat{m} = (\hat{x}, \hat{\tau})$: GPS measurement (coordinates, timestamp)

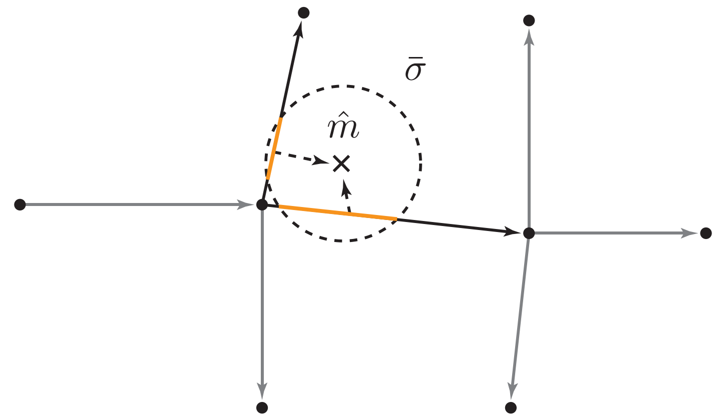
$p(\hat{x}|x; \sigma)$: Probability distribution of GPS measurement error

$\sigma = \sigma_a$: Variance assumed to be ***link-specific*** and ***estimated***

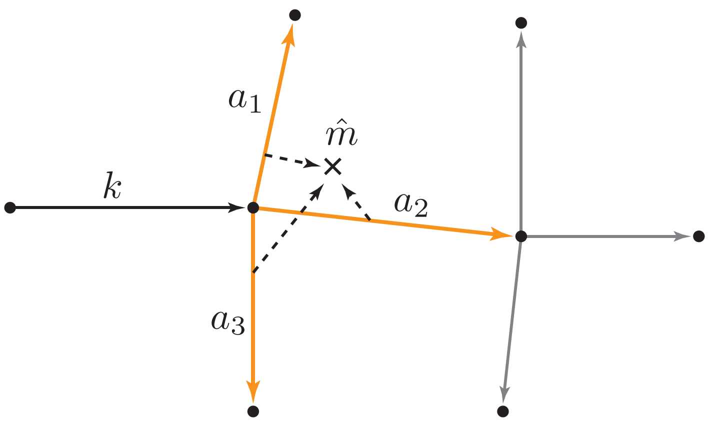
Identification of link-specific variance σ_a

Previous:

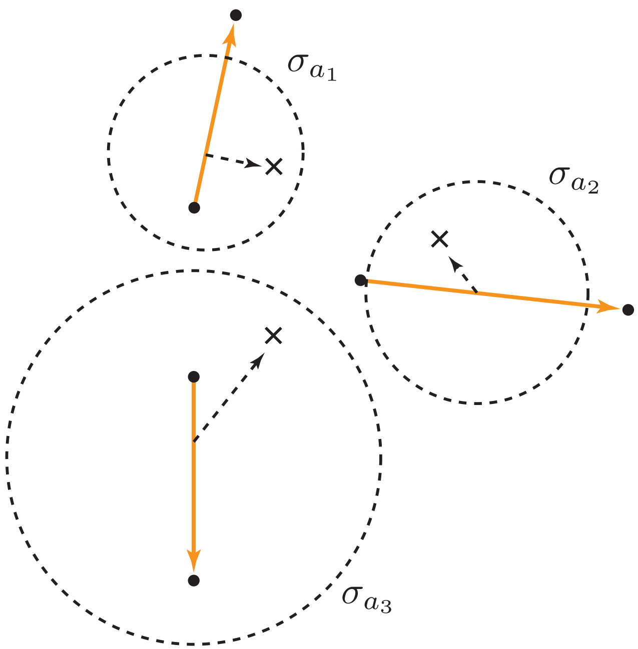




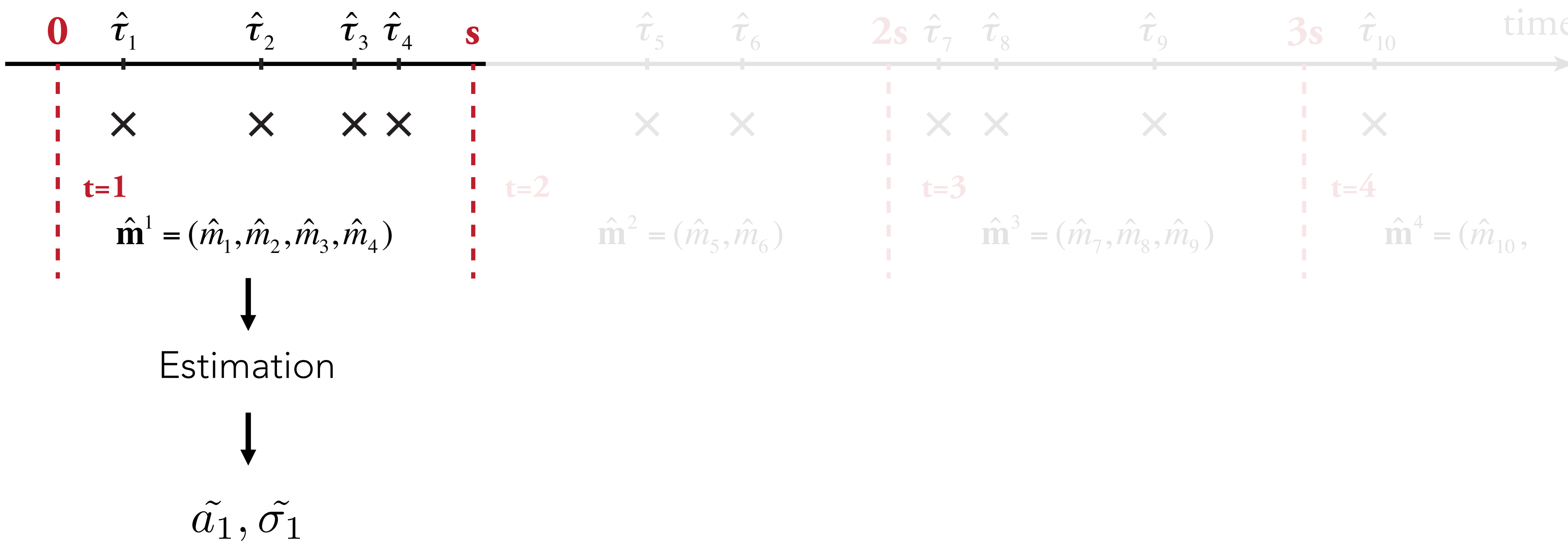
This study:



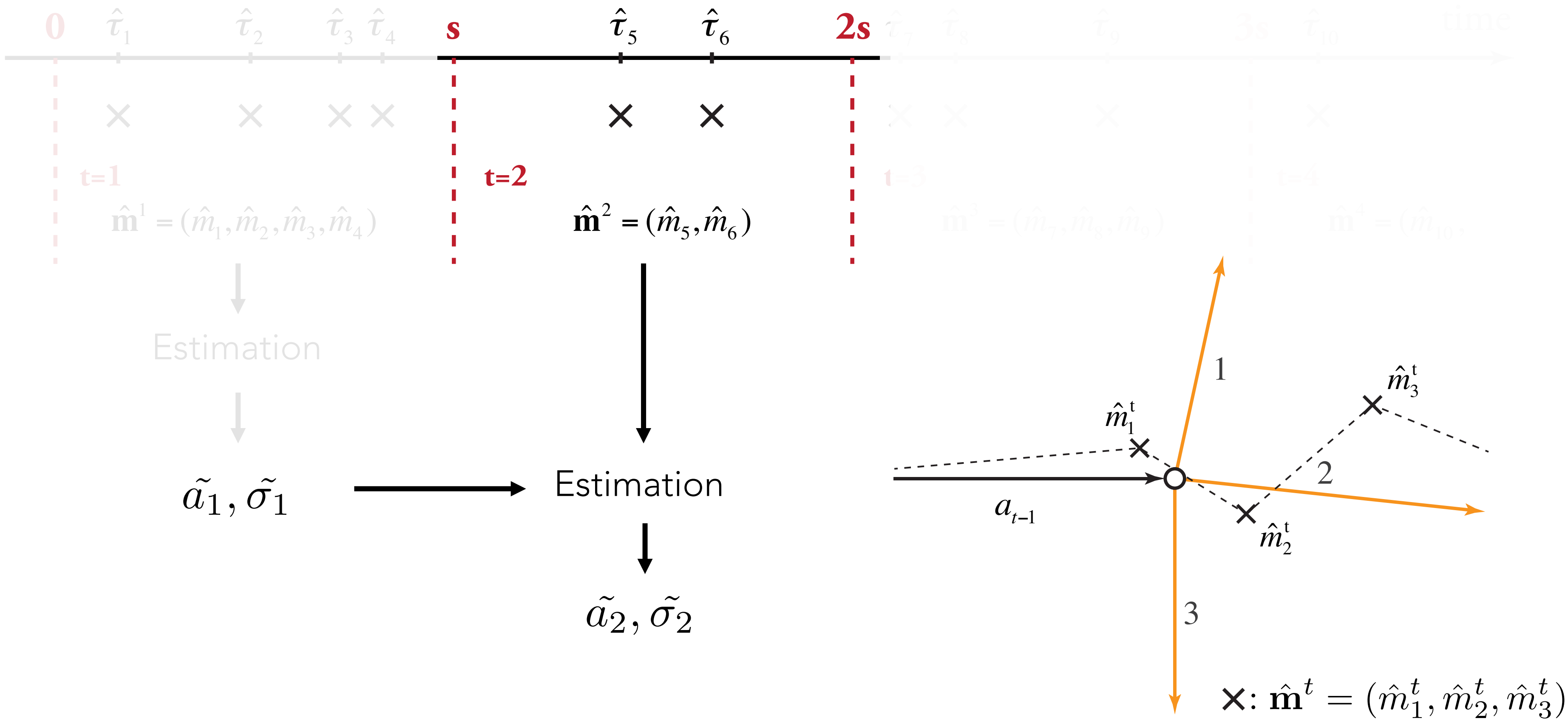
$$\mathcal{A}(k) = \{a | \delta(a|k) = 1\}$$



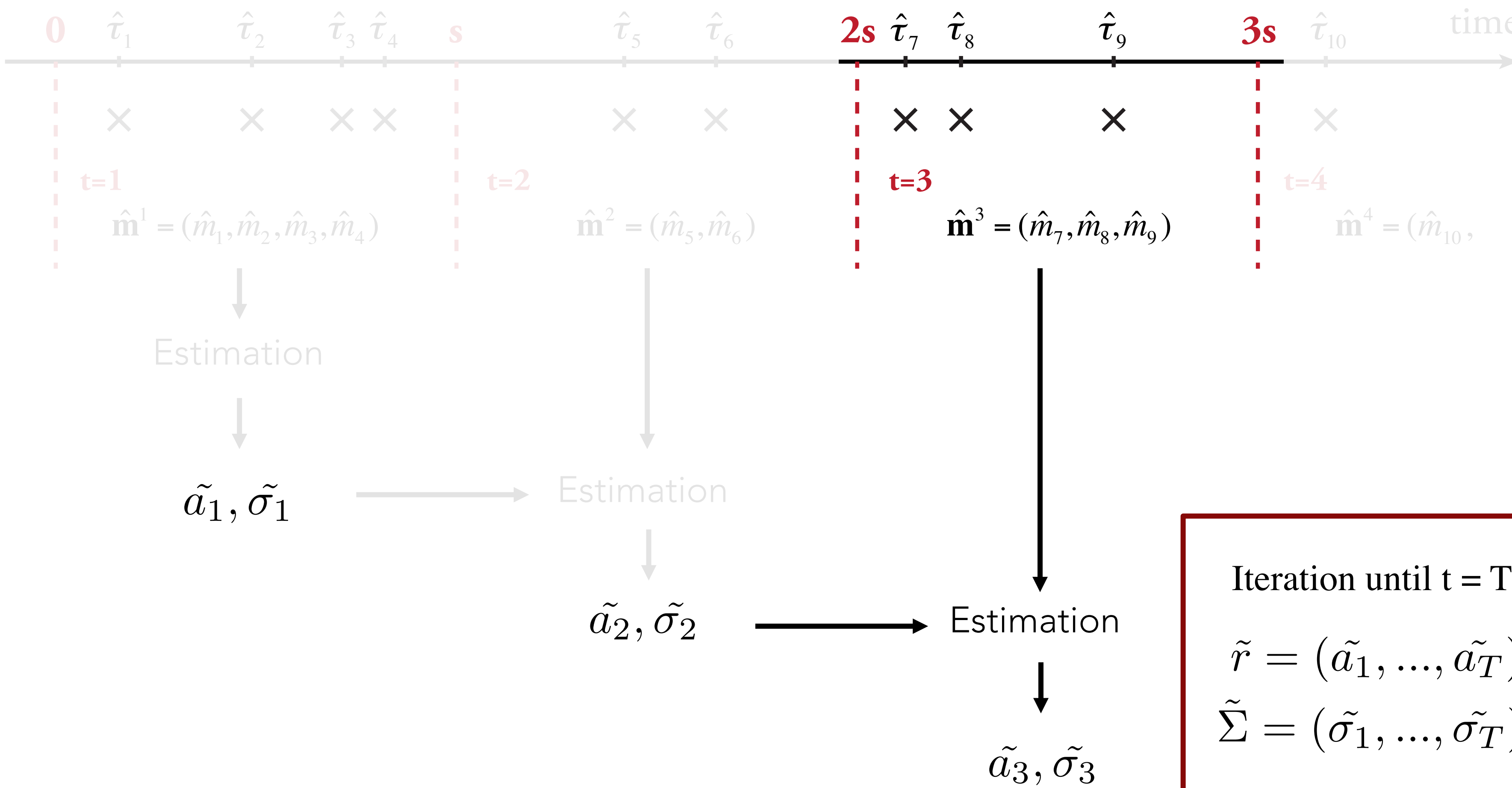
Sequential link measurement



Sequential link measurement



Sequential link measurement



Iteration until t = T

$$\tilde{r} = (\tilde{a}_1, \ldots, \tilde{a}_T)$$

$$\tilde{\Sigma} = (\tilde{\sigma}_1, \ldots, \tilde{\sigma}_T)$$

Estimation of the variance

Maximizing measurement likelihood

$$\tilde{\sigma}_{a_t} = \arg\max_{\sigma} p(\hat{\mathbf{m}}^t | a_{t-1}; \sigma)$$

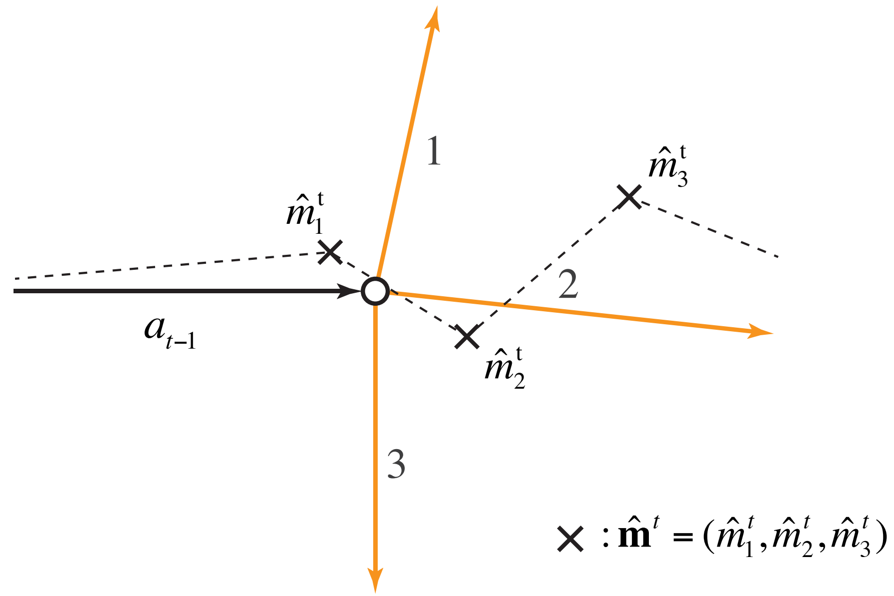
where $$p(\hat{\mathbf{m}}^t | \tilde{a}_{t-1}; \sigma) = \sum_{a_t \in \mathcal{A}(\tilde{a}_{t-1})} \underline{p(\hat{\mathbf{m}}^t | a_t; \sigma)} \, \underline{p(a_t | \tilde{a}_{t-1}; \theta)}$$

$p(\mathbf{m}|a)$: Measurement equation

$p(a'|a)$: Markovian route choice model

$\tilde{a}$: Inferred link

$\mathcal{A}(a)$: Set of subsequent links of a



Measurement equation

Giving probability that $\hat{\mathbf{m}}^t$ *is observed if* a_t *is the true link*

Measurement equation

$$p(\hat{m}_1^t, ..., \hat{m}_J^t | a_t; \sigma_{a_t}) = p(\hat{x}_1^t, ..., \hat{x}_J^t | a_t; \sigma_{a_t})$$ Timestamp has no error

$$= \prod_{j=1}^{J} p(\hat{x}_j^t | a_t; \sigma_{a_t})$$ Probabilities are independent

$$= \prod_{j=1}^{J} \int_{x_j \in a_t} p(\hat{x}_j^t | x_j, a_t; \sigma_{a_t}) p(x_j | a_t) dx_j$$

Rayleigh distribution [van Diggelen, 2007]

$$p(\hat{x}_j^t | x_j, a_t; \sigma_{a_t}) = \frac{\|\hat{x}_j^t - x_j\|}{\sigma_{a_t}^2} \exp\left(-\frac{\|\hat{x}_j^t - x_j\|^2}{2\sigma_{a_t}^2}\right)$$

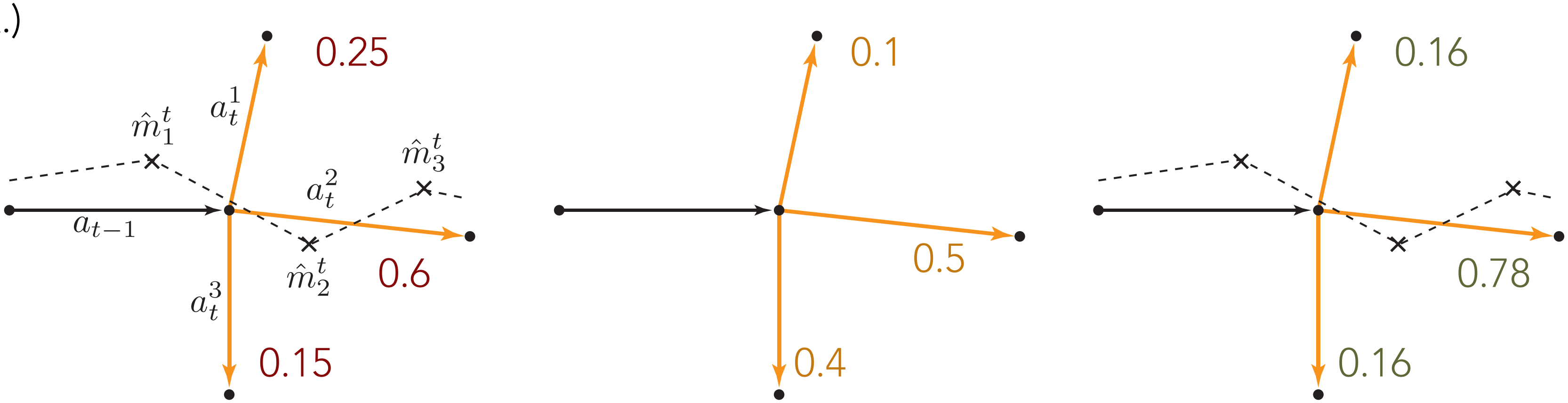
Identifying links

Maximizing posterior probability

$$a_t = \arg \max_{a_t \in \mathcal{A}(a_{t-1})} \underline{p(a_t | \hat{\mathbf{m}}^t, a_{t-1})} = \frac{\underline{p(\hat{\mathbf{m}}^t | a_t; \sigma_{a_t})}\underline{p(a_t | a_{t-1}; \theta)}}{\sum_{a_t \in \mathcal{A}(a_{t-1})} \underline{p(\hat{\mathbf{m}}^t | a_t; \sigma_{a_t})}\underline{p(a_t | a_{t-1}; \theta)}}.$$

By repeating until t=T, we obtain a path: $\tilde{r} = [a_1, ..., a_t, ..., a_T]$

Ex.)



Measurement probability: $\underline{p(\hat{\mathbf{m}}^t | a_t; \sigma_{a_t})}$

Prior probability: $\underline{p(a_t | a_{t-1})}$

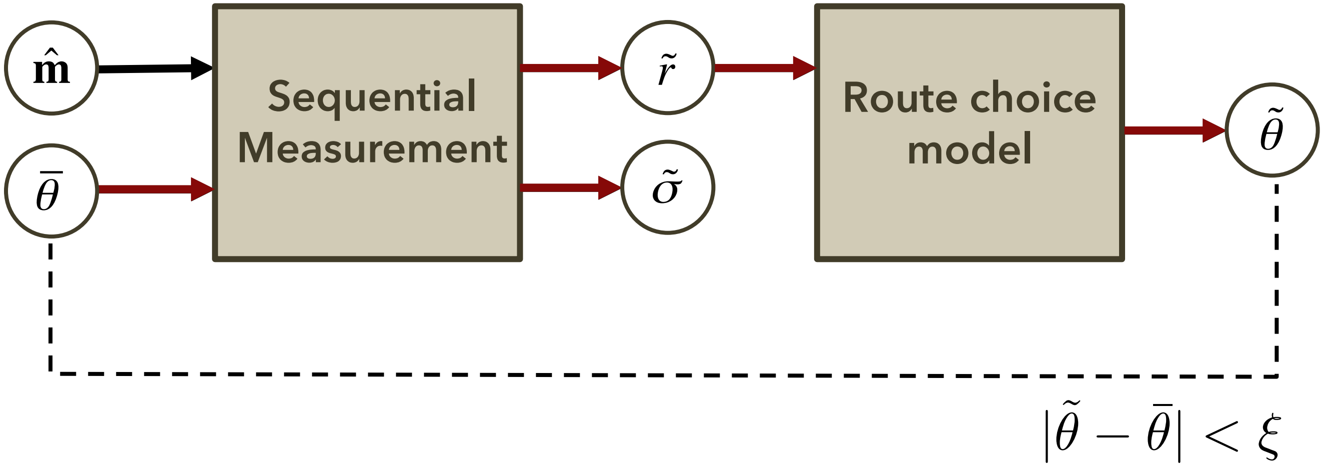
Posterior probability: $\underline{p(a_t | \hat{\mathbf{m}}^t, a_{t-1})}$

Estimation of route choice model

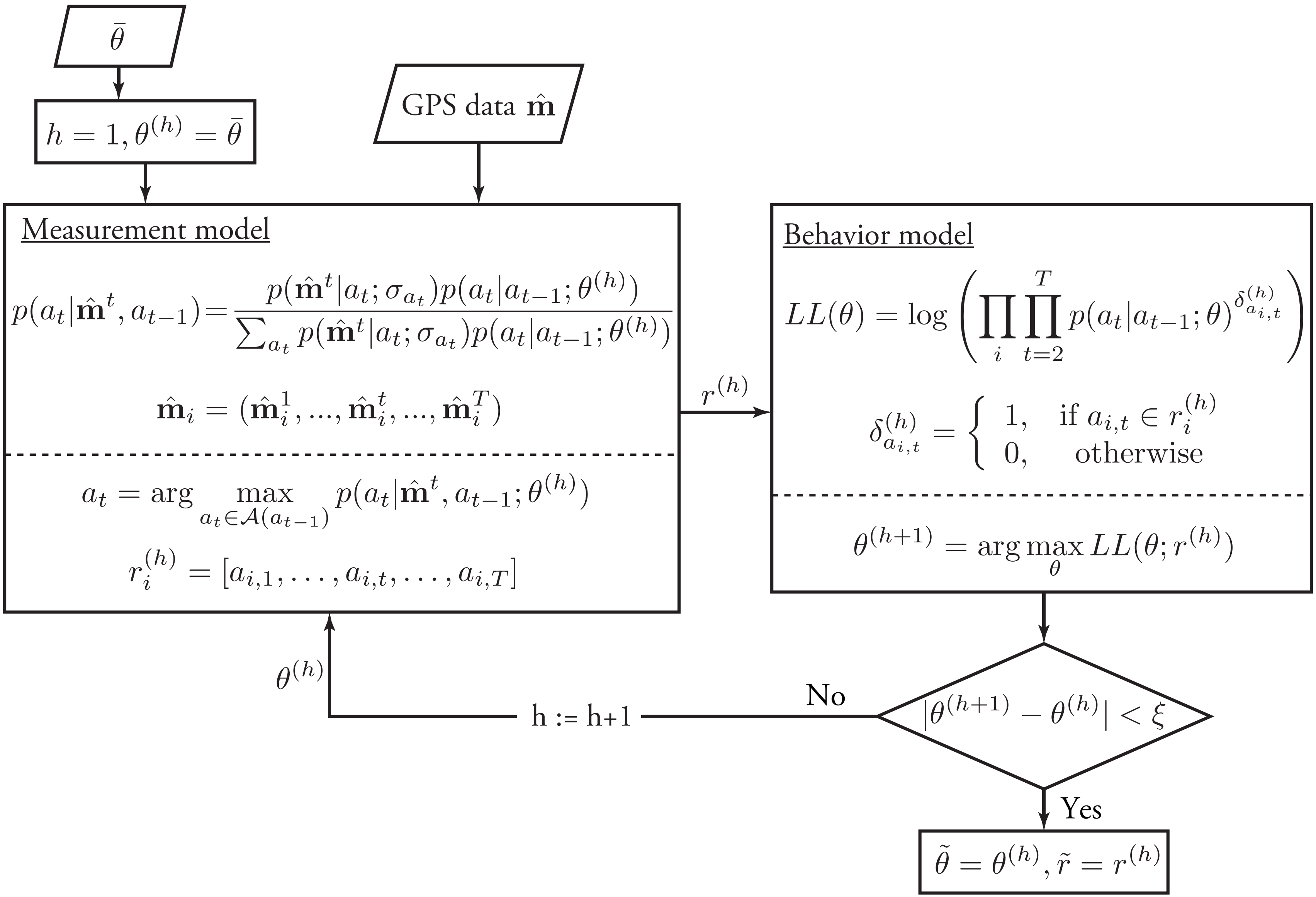
Maximum likelihood estimation

$$\theta = \arg\max_{\theta} LL(\theta; \tilde{r})$$

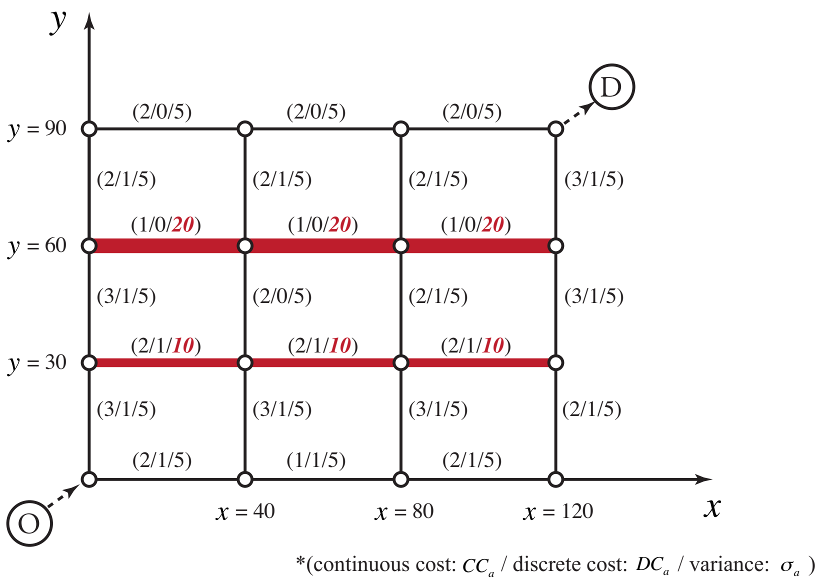
where
$$LL(\theta; \tilde{r}) = \sum_{i} \sum_{t=2}^{T} \delta^{i}_{a_t} \log\left(p(a_t | a_{t-1}; \theta)\right)$$



Structural estimation



Numerical experiment



1. Sampling paths

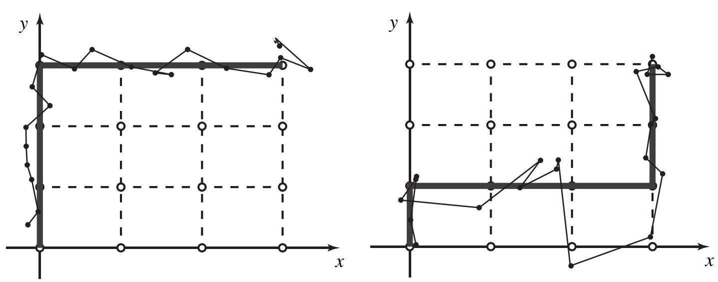
Model:

$$p(a|k) = \frac{\exp(u(a|k))}{\sum_{a \in \mathcal{A}(k)} \exp(u(a|k))}$$

$$u(a|k) = \theta_1 \mathrm{TT}_a + \theta_2 \mathrm{CC}_a + \theta_3 \mathrm{DC}_a + \theta_4 \mathrm{UT}_{a|k}$$

True parameter: $\bar{\theta} = [-0.1, -2, -1.5, -4]$

Examples:



(a) (b)

2. Sampling trajectories

Time discretization: $s = 30\text{sec.}$

Sampling interval: $\hat{\tau}_j - \hat{\tau}_{j-1} = 10\text{sec.}$

Distribution: $\hat{x} - x \sim N(0, \sigma_a)$

$\hat{y} - y \sim N(0, \sigma_a)$

Result | link measurement accuracy

- **Estimating variance refines** the measurement accuracy
- The effect of incorporating route choice models is **largely dependent on the initial parameter**

			Link accuracy(%)		Ave.$\|\tilde{\sigma}_a - \sigma_a^*\|$	
Model	σ	θ	-	Switching	-	Switching
1	Given	$[0, 0, 0, 0]$	54.57	68.86	-	-
2	Estimated	$[0, 0, 0, 0]$	76.86	82.86	5.848	4.40
3	Estimated	$[-1.5, -0.1, -2, -10]$	4.86	38.29	41.99	21.21
4	Estimated	$[-0.1, -2, -1.5, -4]$	76.86	91.71	7.58	4.06

$^*\theta = [0, 0, 0, 0]$: no prior route choice information

Result | structural estimation

- **Structural estimation reduces biases** in estimated parameters and gives estimates close to the true values (total diff.: 3.643 vs. 1.244)
- **Convergence values are equivalent regardless of the initial values**, which are examined by the following settings:

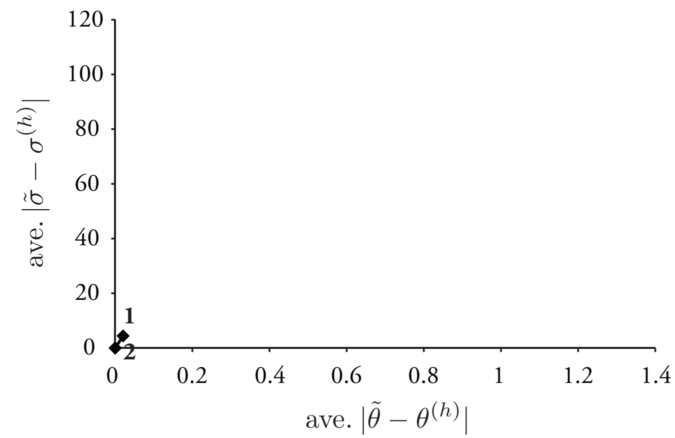
$$\bar{\theta} = [-0.1, -2, -1.5, -4], [-10, -10, -10, -10], [-100, 0, 0, 0], [-100, -100, -100, -100], [10, 10, 10, 10]$$

Input: $\bar{\theta} = [0, 0, 0, 0]$ (No information)

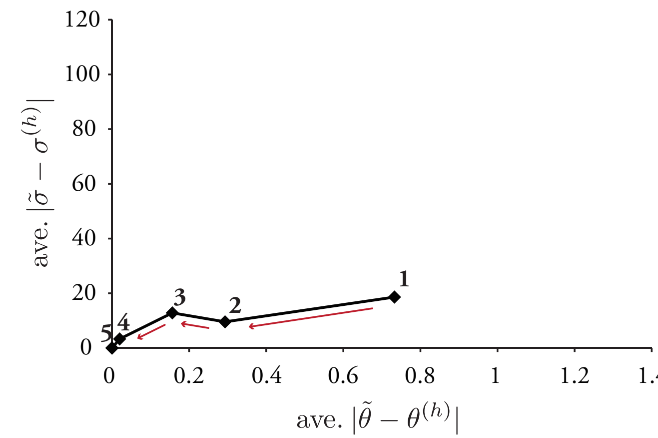
		One-way			Structural Estimation		
	TRUE	Estimates	abs(diff.*)	t-value	Estimates	abs(diff.)	t-value
θ_1	-0.1	0.002	0.102	0.101	-0.064	0.036	-2.562
θ_2	-2	-0.755	1.245	-4.164	-1.727	0.273	-6.882
θ_3	-1.5	-1.312	0.188	-4.772	-1.046	0.454	-3.519
θ_4	-4	-1.892	2.108	-8.864	-3.519	0.481	-9.739
total error			3.643			1.244	
sample				350			350
L0				-373.221			-371.887
LL				-269.872			-211.308
ρ^2				0.266			0.421
iteration							6

Result | convergence process

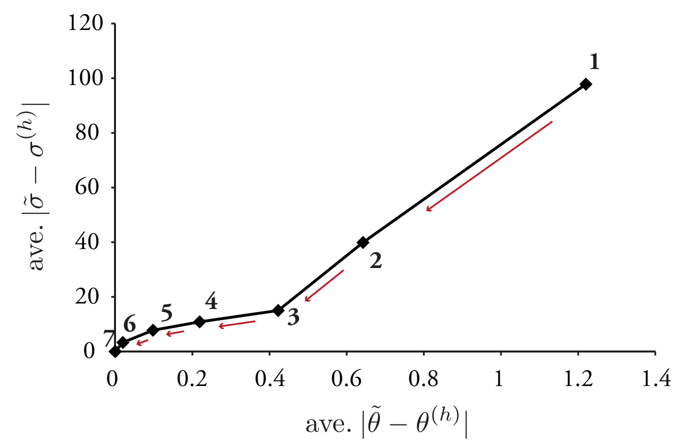
- **Less vibration, and small number of iteration in any cases**



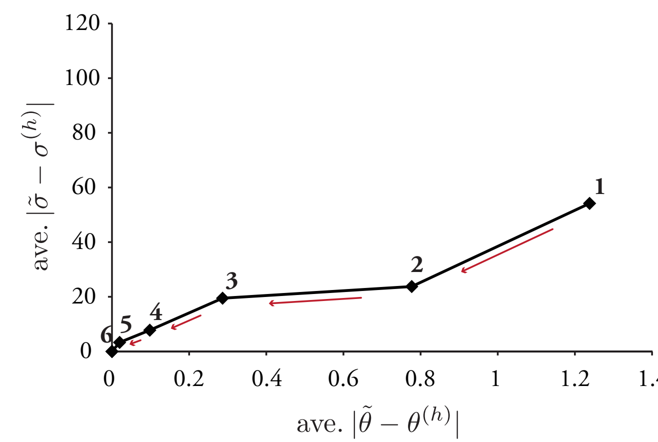
(a) $\theta^{(0)} = [-0.1, -2, -1.5, -4]$



(b) $\theta^{(0)} = [0, 0, 0, 0]$



(c) $\theta^{(0)} = [-1.5, -0.1, -2, -10]$



(d) $\theta^{(0)} = [10, 10, 10, 10]$

Case study

GPS data

- Matsuyama city, Japan
- Pedestrian trip in the city center
- Randomly sampled 30 walking trips

Specification

$$v(a|k) = \theta_1 TT_a + \theta_2 CU_a + \theta_3 DU_a + \theta_4 UT_{a|k}$$

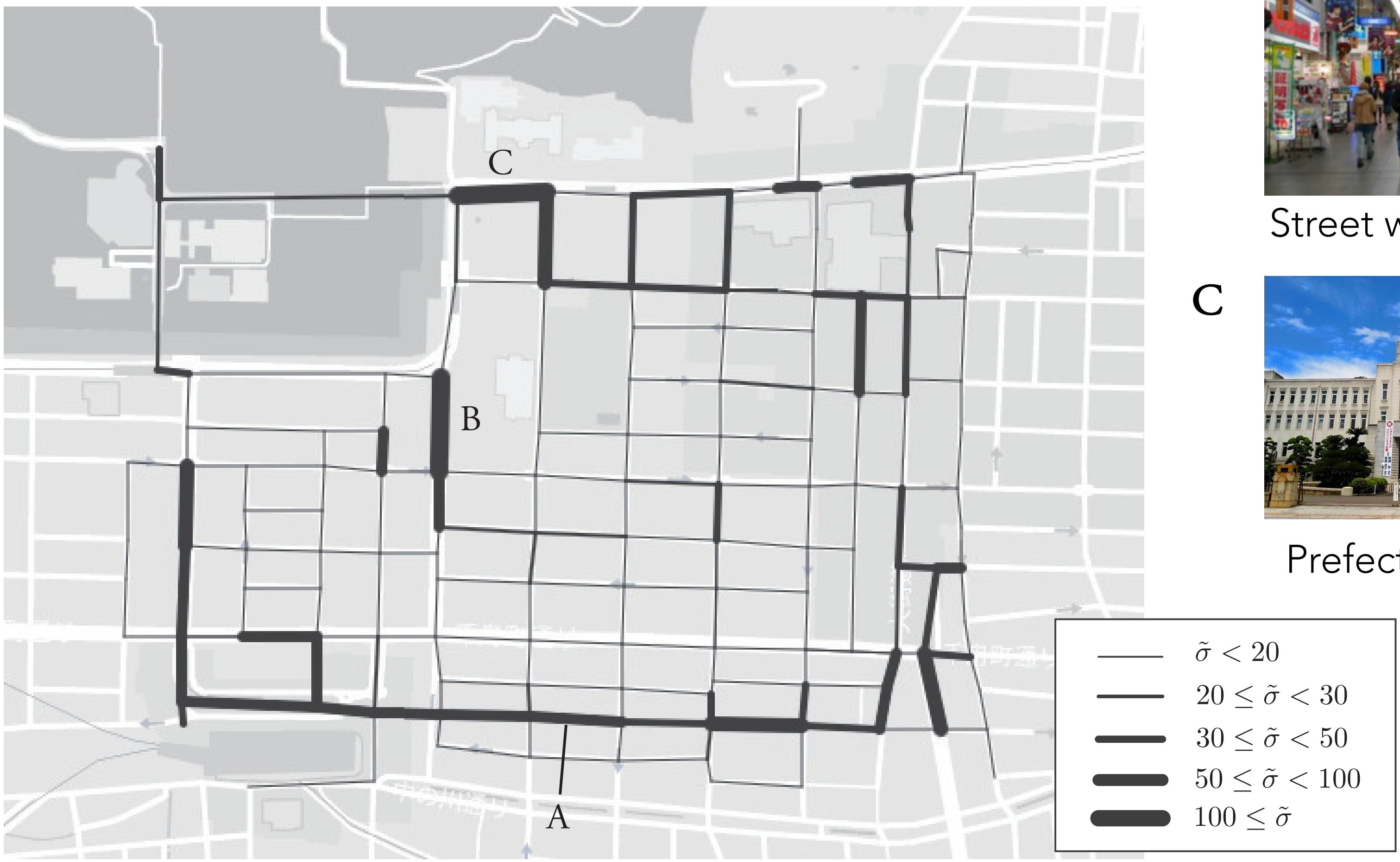
TT : Travel time [min.]

CU : Sidewalk width [m]

DU : Arcade dummy variable

UT : U-turn dummy variable

Estimation result of measurement model



A

Street with arcade

C

Prefecture hall

Estimation result of route choice model

- Travel time (θ_1) seems to be significant from the result of ***one-way model***, however,
- ***Structural estimation*** result shows that links with arcade (θ_3) are the most likely to be passed by pedestrians; travel time (θ_1) is not significant
- Other t-values and rho-square (ρ^2) indicate that the ***structural estimation*** improves parameter estimation results

Input: $\tilde{\theta} = [0, 0, 0, 0]$ (No information)

		One-way		Structural Estimation	
		Estimates	t-value	Estimates	t-value
Travel time (min.)	θ_1	-0.007	-2.473	-0.001	-0.428
Sidewalk width (m)	θ_2	0.088	1.497	0.134	1.582
With arcade	θ_3	-0.004	-0.011	2.760	4.288
U-turn	θ_4	0.774	0.532	0.469	3.344
	sample		270		270
	L0		-307.608		-309.066
	LL		-302.174		-225.162
	ρ^2		0.005		0.259
	iteration				11

Conclusion

Contributions

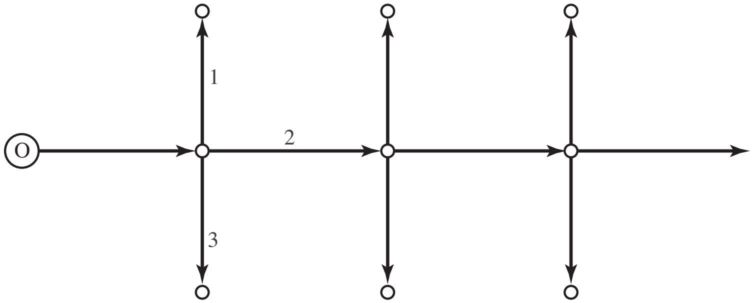
- **Sequential link measurement model**
  - Estimating ***link-specific variance*** of GPS measurement error
- **Structural estimation**
  - ***Reducing biases*** caused from the initial parameter settings
- **Validation**
  - Structural estimation achieves to ***refine the results*** and ***quickly converge***.
- **Application**
  - The methods are effective in a real pedestrian network and ***bring hidden preferences to light***.

Future work

Joint estimation

$$p(\hat{m}_{1:T}; \sigma, \theta) = \sum_{a_{1:T} \in \mathcal{R}} p(\hat{m}_{1:T} | a_{1:T}; \sigma) p(a_{1:T}; \theta)$$

- **Algorithm for preserving path-based measurement probabilities**
  - Generating path candidates by re-sampling at each time using probabilities



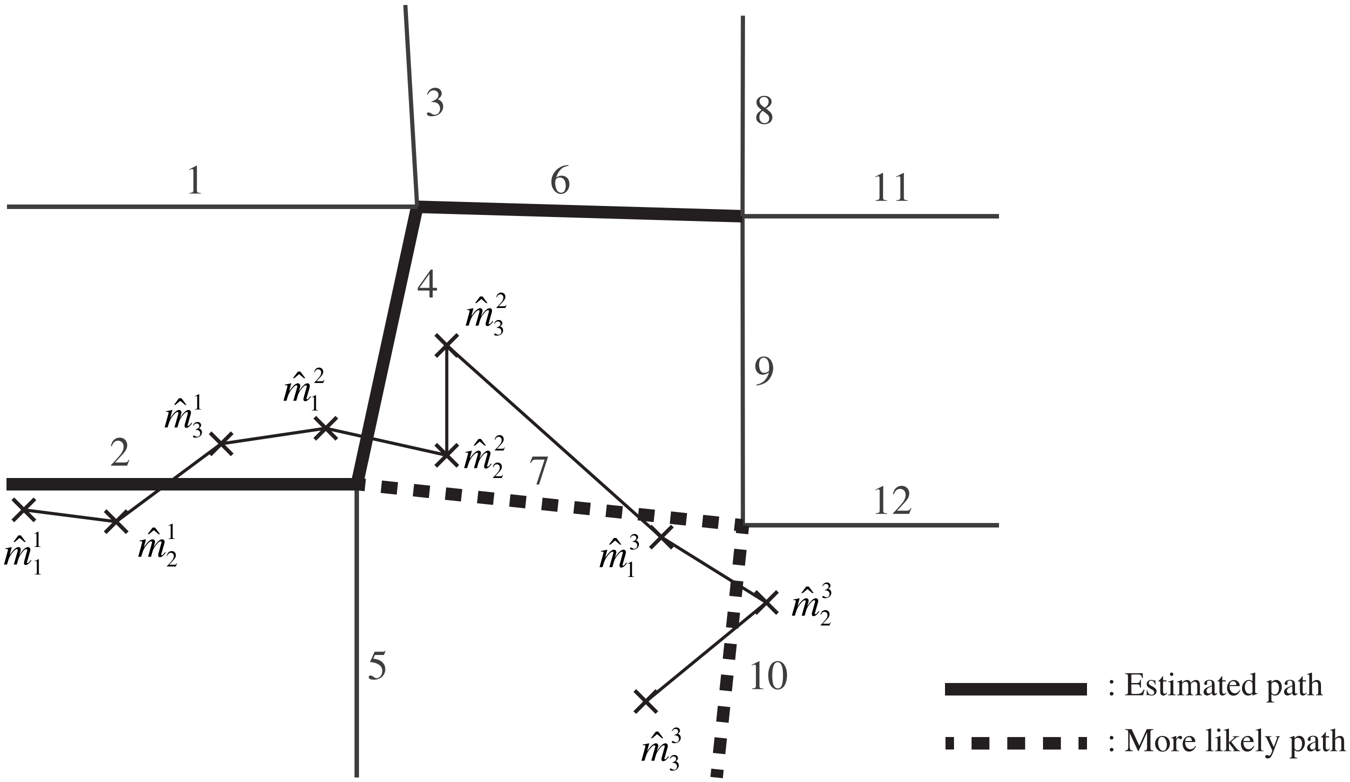
$$|\mathcal{R}| = 3^T$$

- **Estimation method for reducing computational burden**
  - EM algorithm, variational Bayes method

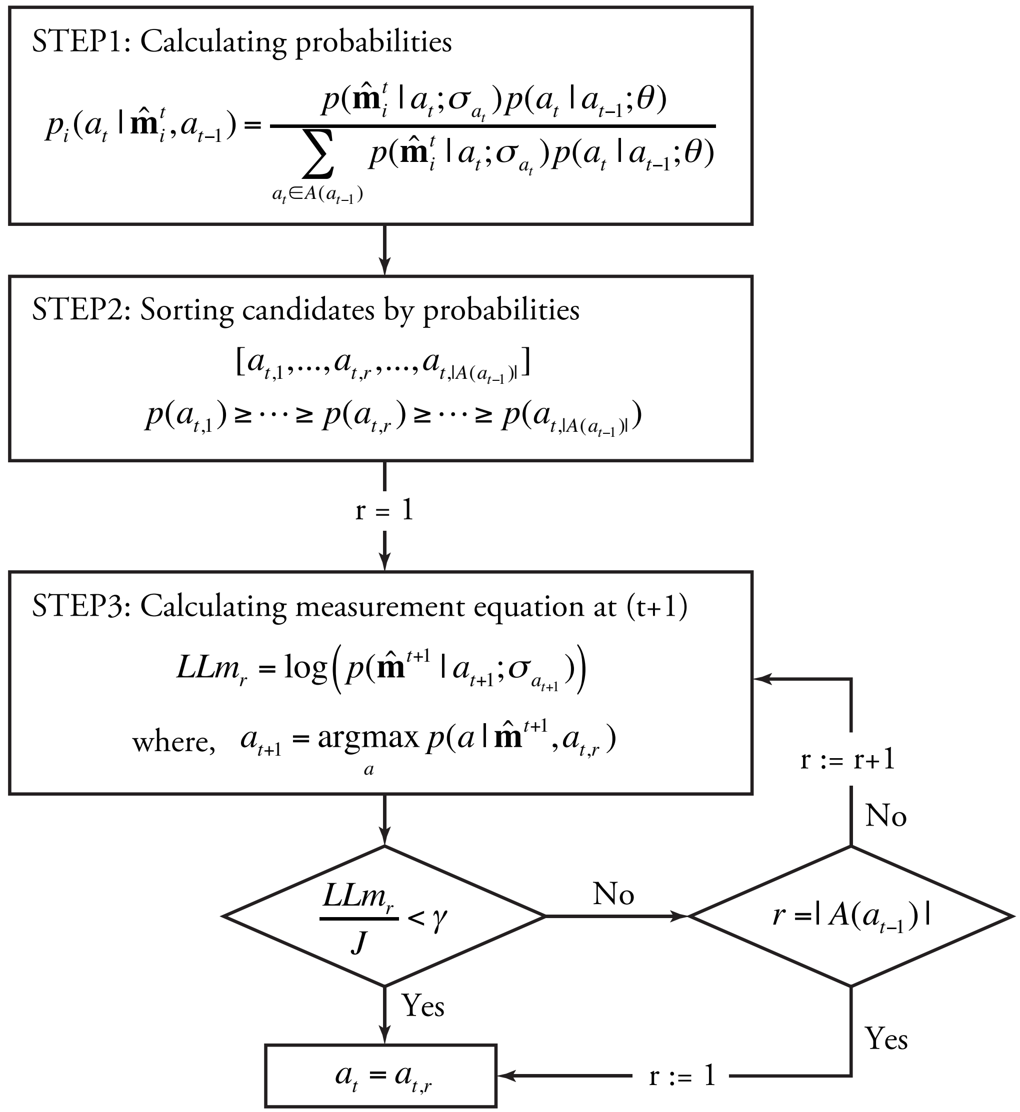
Thank you for attention.

Link switching

Difficulties regarding link connectivity because of myopic optimization



Link switching



Result | structural estimation

Table:
Average and standard deviation of estimated parameters, the number of iterations and computational time of 100 structural estimations

	$\tilde{\theta}_1$	$\tilde{\theta}_2$	$\tilde{\theta}_3$	$\tilde{\theta}_4$	Iteration	CPU time (s)
Ave.	-0.112	-1.140	-1.006	-2.916	4.086	858.179
Std.	0.023	0.663	0.413	4.006	1.007	229.737

***True parameter:** $\bar{\theta} = [-0.1, -2, -1.5, -4]$