SPATIO-TEMPORAL ANALYSIS OF TELECOMMUNICATIONS DATA:

Effects of Residential Shifts and Smartphone Adoption on Mobility Patterns and Communication Behavior

---

Hsiu-Hsuan Yeh (葉秀軒)

Department of Economics, National Taiwan University

Contents

Introduction

- We utilize a large-scale dataset, call detailed records (CDRs).
  - collected for ten months (August 2013 - May 2014)
  - contain over 1.5 billion records per month
  - involve 500,000+ phone numbers
- The effect dynamics of residential shifts and smartphone adoption are examined.
- No comparable dataset combines scale, time depth, and low cost.

- Residential shifts and smartphone adoption are large-scale collective transitions.
  - natural disaster relocations
  - environmental displacements due to industrial pollution
  - technology upgrades in developing countries (Wi-Fi, GPS, and 3G -> 4G)
  - smart devices (smartglasses, Vision Pro)

- Monitoring mobility and communication behaviors provides policy implications.
  - mobile and transportation infrastructure
  - commuter traffic optimization
  - community connectivity evaluation

Findings

- Effects of residential shifts are time-variant.
- During shifting periods, shifting users mostly contact with their existing friends.
  - substantial increases in call duration and contact distances
  - mobility is unpredictable along with fixed direction
- Effects instantly fade back to baseline or become negative.
- Effects of smartphone adoption are relatively static (upward shifts).

Contributions

- CDRs have already extensively employed to study mobility and communication behaviors.
  - commuting flows (Phithakkitnukoon et al. 2012)
  - diseases transmission flows (Wesolowski et al. 2016)
  - social network and mobility interaction (Blumenstock et al. 2025)
- We incorporate the load sharing mechanism, which is more robust to:
  - location inference
  - communication intensity

- Most CDR-based migration studies inadequately utilize temporal data.
  - on monthly basis
  - between two separate periods
- We propose a two-stage home location estimation considering residential shifts.

Data and Methods

Datasets

Client Number	Duration	Start Time	End Time	Calling Number	Called Number	Cell ID
66ak2s5v	62.0	2013-08-31 09:32:08	2013-08-31 09:33:10	66ak2s5v	moyl2k57	3649
ltzkksuv	148.0	2013-08-31 09:33:55	2013-08-31 09:36:23	ltzkksuv	hjo0ksut	3B56
njo45k8v	46.0	2013-08-31 09:36:03	2013-08-31 09:36:49	8yro82d5	njo45k8v	394C
			⋮			

- Datasets include CDRs, clients profiles, and cell coordinates
- CDRs come from Sichuan, China, spanning from August 2013 to May 2014.
- After cleaning, ~0.5 billion monthly records involve 340,000+ numbers.

- We will proceed the content in the following order:

1. identification of residential shifts and smartphone adoption
   - home location estimation

2. construction of mobility and mobile communication outcomes
   - (i) radius of gyration, (ii) movement entropy, and (iii) eccentricity
   - (i) total duration, (ii) contact distance, and (iii) contact entropy

3. Estimation Approach

Home Location Estimation

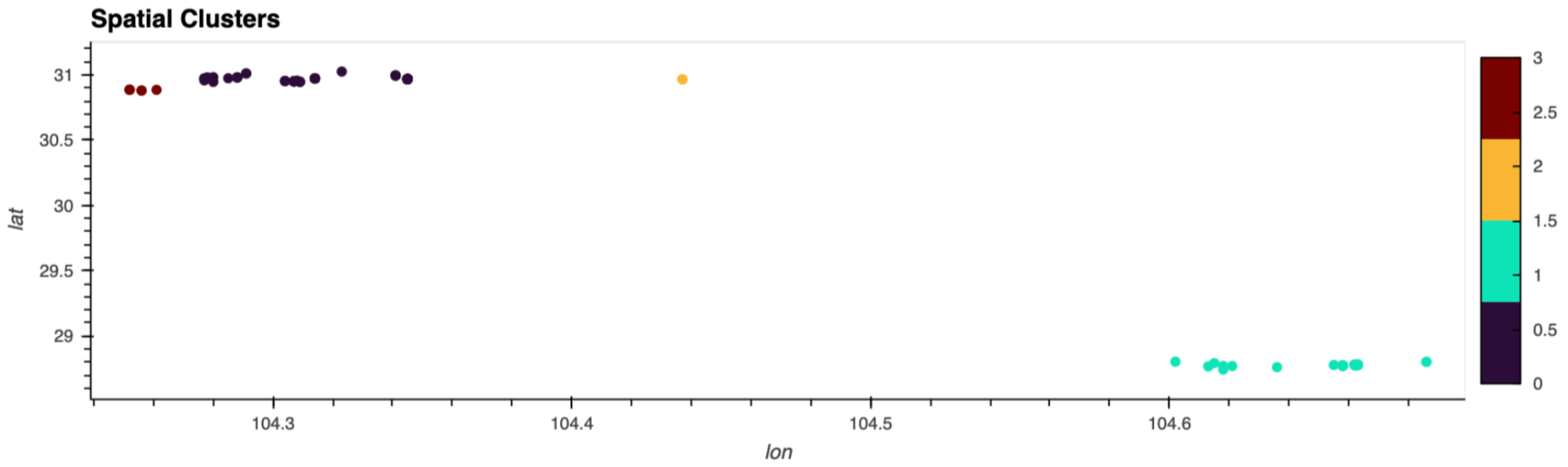
- To identify residential shifts, we need to estimate home locations.
- Previously, home location estimation is not specifically designed to account for relocations.
- We propose a two-stage estimation approach encompassing:
  - first-stage spatial clustering
  - second-stage temporal filtering

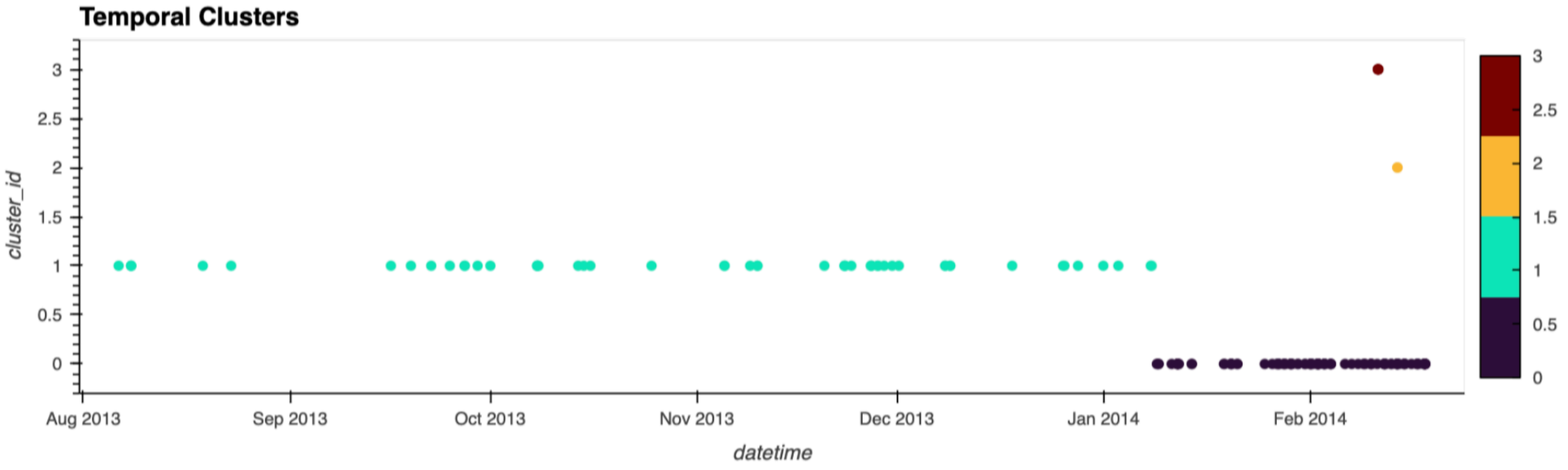
Home Location Estimation: Spatial Clustering

- Spatial clustering group observed locations corresponding to a stay location.
- We assume individuals spend the majority of time at home during nighttime.
- We apply DBSCAN on nighttime CDRs to obtain a set of nighttime clusters.

Home Location Estimation: Temporal Filtering

- Clusters serve calls for phone users.
- The goal of temporal filtering scheme is to identify “representative clusters”.
  - processing substantial calls
  - non-overlapping nature of service time interval
- The home locations are defined as home clusters’ centroids.


Spatial Clusters
31
30.5
30
29.5
29
lat
104.3
104.4
104.5
104.6
lon
3
2.5
2
1.5
1
0.5
0


Temporal Clusters
3
2.5
2
1.5
1
0.5
0
cluster_id
Aug 2013
Sep 2013
Oct 2013
Nov 2013
Dec 2013
Jan 2014
Feb 2014
datetime

Identification of Residential Shifts

- Denote $M := \{1, \cdots, 10\}$ as the set of monthly periods after indexing.
- A user is a shifting user associated with group $\mathcal{G}_g$ where $g \in \{4, 5, 6, 7\} \subset M$ if:

  1. user only changes home locations once throughout the sample period

  2. the two home locations are in different prefectures

  3. the shifting timing occurs in month g.
- Denote a group of non-shifting users as $\mathcal{G}_\infty$.

Identification of Smartphone Adoption

- A user is a smartphone adopter associated with group $\mathcal{G}_g$ if:

  1. user changes from a non-smartphone device to a smartphone model

  2. do not switch back to non-smartphone devices throughout the sample periods

  3. the adoption event occurs in month $g \in \{4, 5, 6, 7\} \subset M$

- Denote a group of non-smartphone users as $\mathcal{G}_\infty$.

Month	Residential Shifts		Smartphone Adoption	
	Shifting	Non-shifting	Adopting	Non-adopting
Nov. 2013	247	291,465	2,186	81,949
Dec. 2013	231	291,465	2,354	81,949
Jan. 2014	338	291,465	2,899	81,949
Feb. 2014	458	291,465	2,058	81,949

Construction of Mobility Features

- We mainly operate on a projected location matrix $L_{i,m} \in R^{|B_{i,m}| \times 2}$.
  - the row is denoted as $\left(l_{i,m}\right)_b$
- Projection transforms curved Earth coordinates to a flat plane.
- We employ AEQD projection with the center set to $(r_c)_{i,m}$ where

$$(r_c)_{i,m} = L'_{i,m}\hat{p}_{i,m}$$

with $\hat{p}_{i,m} \in R^{|B_{i,m}|}$ being the sample probability vector based on relative day counts.

- Radius of gyration $\left(r_g\right)_{i,m}$ quantifies the scope the activity area.

$$\left(r_g\right)_{i,m} = \sqrt{\sum_{b \in B_{i,m}} \hat{p}_{i,m,b} \cdot \| \left(l_{i,m}\right)_b - \left(r_c\right)_{i,m} \|^2}.$$

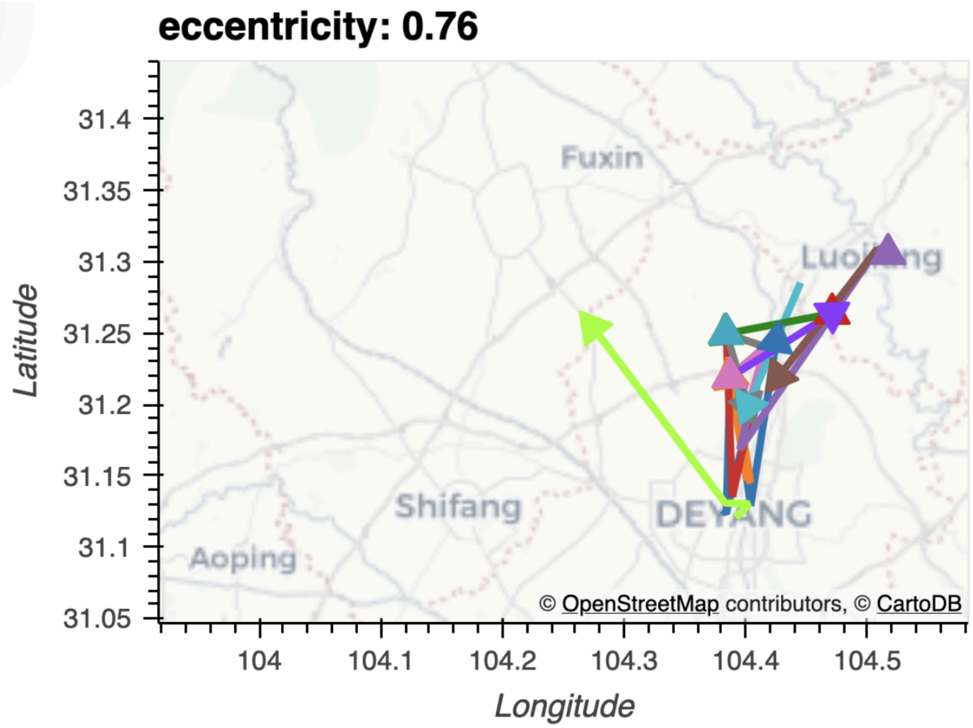
- Movement entropy $\mathrm{me}_{i,m}$ measures how unpredictable (diverse) spatial movement patterns are.

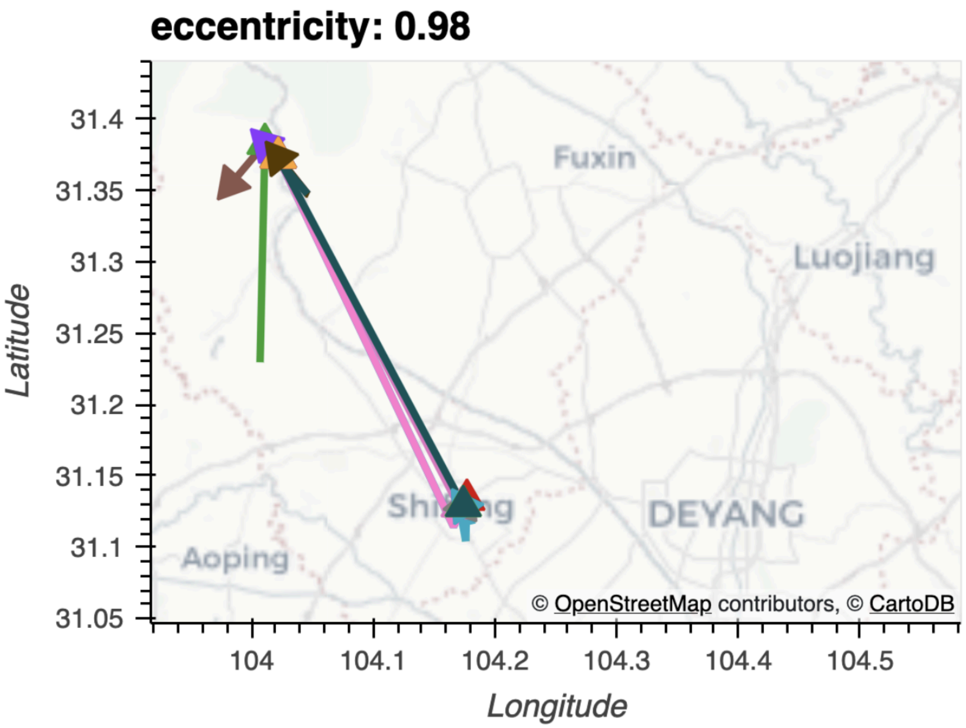
$$\mathrm{me}_{i,m} = -\frac{\sum_{b \in B_{i,m}} \hat{p}_{i,m,b} \log\bigl(\hat{p}_{i,m,b}\bigr)}{\log\bigl(|B_{i,m}|\bigr)}.$$

- Eccentricity $\mathrm{ecc}_{i,m}$ evaluates to what extent the activity area spreads in an elliptical shape.

$$\mathrm{ecc}_{i,m} = \sqrt{1 - \left(\frac{\lambda_{i,m,2}}{\lambda_{i,m,1}}\right)^2}$$

where $\lambda_{i,m,1}$ and $\lambda_{i,m,2}$ are the major and minor eigenvalues of the sample variance-covariance matrix, respectively.


eccentricity: 0.76
Latitude
Longitude
31.4
31.35
31.3
31.25
31.2
31.15
31.1
31.05
104
104.1
104.2
104.3
104.4
104.5
Fuxin
Luojiang
Shifang
DEYANG
Aoping
© OpenStreetMap contributors, © CartoDB


eccentricity: 0.98
Latitude
Longitude
31.4
31.35
31.3
31.25
31.2
31.15
31.1
31.05
104
104.1
104.2
104.3
104.4
104.5
Fuxin
Luojiang
DEYANG
Aoping
© OpenStreetMap contributors, © CartoDB

Construction of Communication Features

- Mobile Communication is directional and we focus on the outgoing side.
- We further require outgoing calls should be reciprocated.
- The randomness of mobile communication is based on relative call duration.

- Contact entropy measures the diversity of mobile interactions.

$$\mathrm{ce}_{i,m}^{\mathrm{out}} = -\frac{\sum_{j \in V_{i,m}^{\mathrm{out}}} \hat{p}_{i,j,m} \log\left(\hat{p}_{i,j,m}\right)}{\log\left(|V_{i,m}^{\mathrm{out}}|\right)}$$

where

$$V_{i,m}^{\mathrm{out}} := \{ j \in V \mid \text{user } i \text{ has once called user } j \text{ in month } m \}.$$

- Contact distance $\mathrm{cd}_{i,m}^{\mathrm{out}}$ quantifies the spatial reach of mobile interactions.

$$\mathrm{cd}_{i,m}^{\mathrm{out}} = \sum_{j \in V_{i,m}^{\mathrm{out}}} \hat{p}_{i,j,m} \cdot d\big(\mathrm{home}_{i,m}, \mathrm{home}_{j,m}\big).$$

- Contact distance is an application of our two-stage home location estimation.

Average Treatment Effects on the Treated

- Ideally, treatment effects should be evaluated under counterfactual scenarios.
- The average treatment effects on the treated (ATT) provides the guidance.

$$\mathbb{E}\left[Y_{i,t} - Y_{i,t}(\infty) \mid D_i = 1\right]$$

where $Y_{i,t}(\infty)$ is the counterfactual untreated outcome.

Difference-in-differences

- Difference-in-differences (DiD) is an intuitive estimator for ATT.

$$\mathbb{E}\big[Y_{i,t} - Y_{i,t-1} \mid D_i = 1\big] - \mathbb{E}\big[Y_{i,t} - Y_{i,t-1} \mid D_i = 0\big]$$

- If trends in counterfactual untreated outcome are parallel, DiD can recover ATT.

$$\mathbb{E}\big[Y_{i,t}(\infty) - Y_{i,t-1}(\infty) \mid D_i = 1\big] = \mathbb{E}\big[Y_{i,t}(\infty) - Y_{i,t-1}(\infty) \mid D_i = 0\big]$$

- Canonical DiD is under settings of two groups and two periods.
- To examine effect dynamics, multiple periods DiD are required.
- We have four treated groups so extension to multi-group in necessary.

Group-Time ATT

- Therefore, our focus of treatment effect shifts to:

$$\mathrm{ATT}(g, m) = \mathbb{E}\big[Y_{i,m} - Y_{i,m}(\infty) \mid G_{i,g} = 1\big].$$

- The DiD estimator becomes:

$$\mathbb{E}\big[Y_{i,m} - Y_{i,g-\delta-1} \mid G_{i,g} = 1\big] - \mathbb{E}\big[Y_{i,m} - Y_{i,g-\delta-1} \mid G_{i,\infty} = 1\big].$$

- The underlying parallel trends assumption:

$$\mathbb{E}\big[Y_{i,m}(\infty) - Y_{i,g-\delta-1}(\infty) \mid G_{i,g} = 1\big] = \mathbb{E}\big[Y_{i,m}(\infty) - Y_{i,g-\delta-1}(\infty) \mid G_{i,\infty} = 1\big].$$

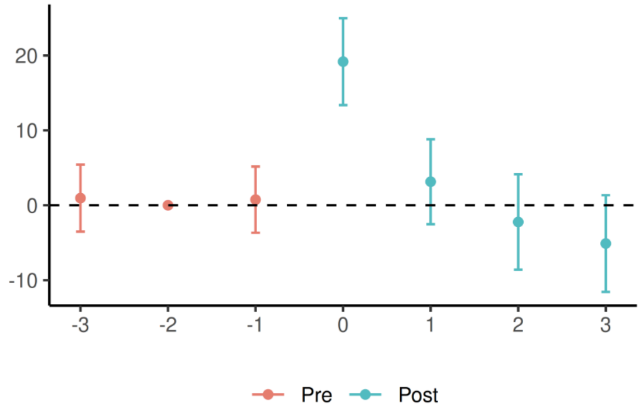
Results

- We adopt the estimation method proposed by (Callaway and Sant'Anna 2021).
- For residential shifts, $\delta = 1$ while regarding smartphone adaption, $\delta = 0$.
- We will use the event study plot to evaluate the plausibility of parallel trends assumption.
- The following formula is used to recover an event study:

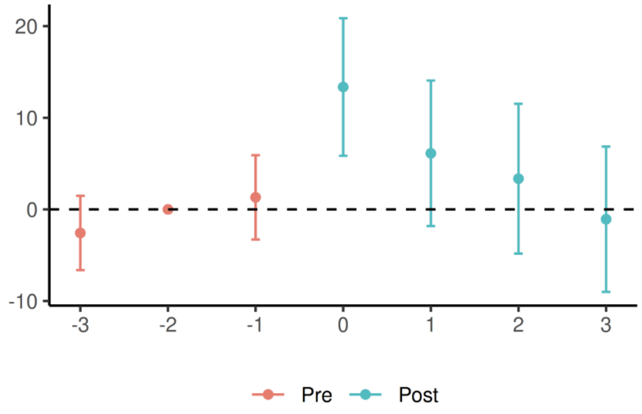
$$\theta(e) = \sum_{g=4}^{7} \mathbb{P}\big(G_{i,g} = 1\big) ATT(g, g+e) \mathbb{1}[g + e \leq 10]$$

Effects of Residential Shifts

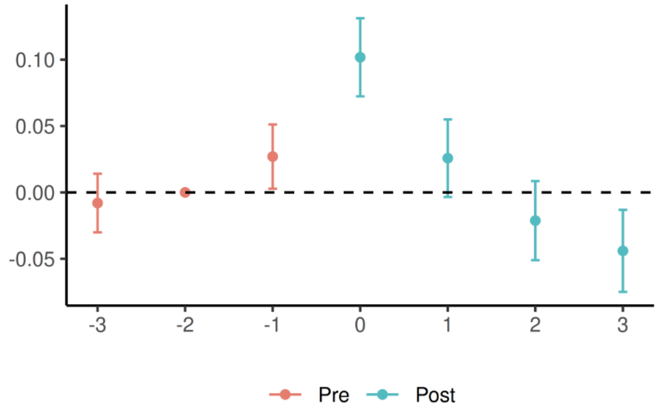
outgoing totoal duration



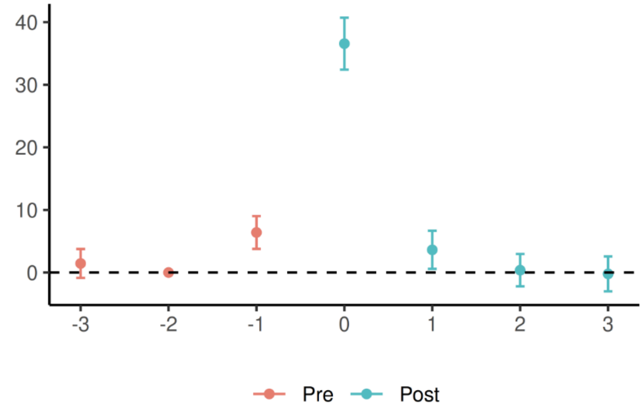
outgoing contact distance



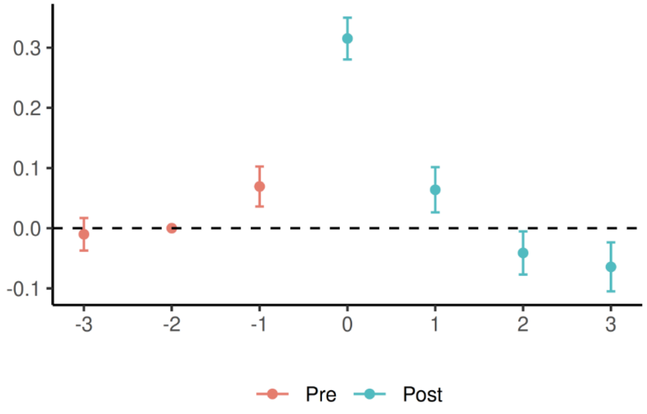
outgoing contact entropy



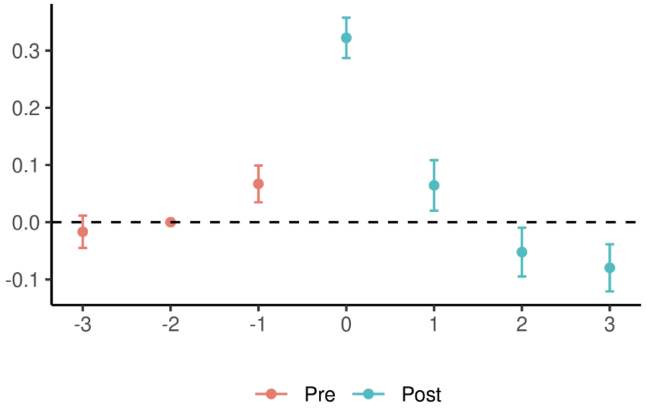
temporal-size-weighted radius of gyration



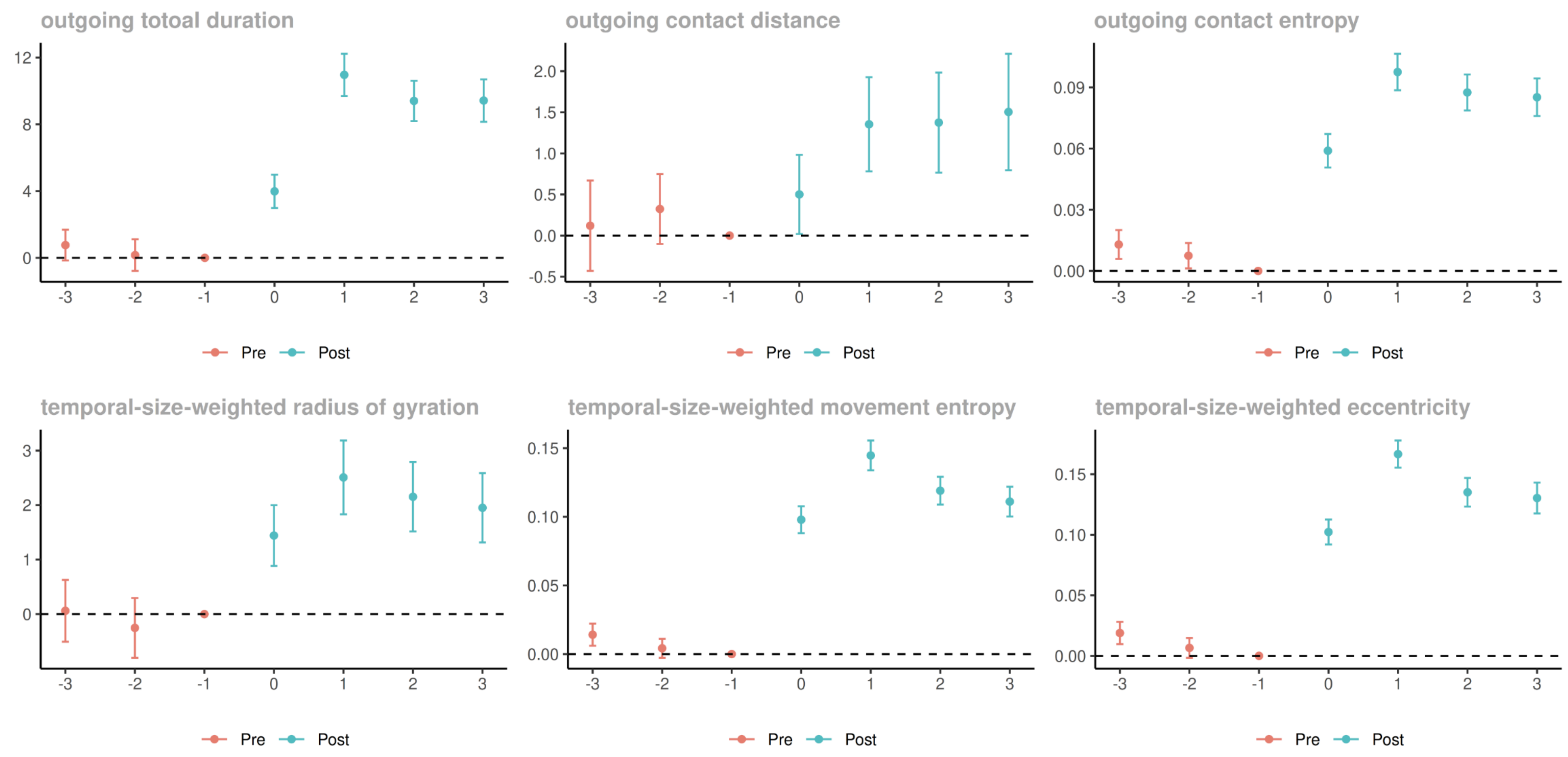
temporal-size-weighted movement entropy



temporal-size-weighted eccentricity



Effects of Smartphone Adoption



Discussions

- Our findings suggest distinct policy implications for two transitions.
- Our methodological contribution lies in accounting for load sharing effects.
- We acknowledge that we did not rigorously test for endogeneity, as both treatments are self-selected.

References

Blumenstock, Joshua E, Guanghua Chi, and Xu Tan. 2025. "Migration and the Value of Social Networks." *Review of Economic Studies* 92 (1): 97–128.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225 (2): 200–230.

Phithakkitnukoon, Santi, Zbigniew Smoreda, and Patrick Olivier. 2012. "Socio-Geography of Human Mobility: A Study Using Longitudinal Mobile Phone Data." *Plos One* 7 (6): e39253.

Wesolowski, Amy, Caroline O Buckee, Kenth Engø-Monsen, and Charlotte Jessica Eland Metcalf. 2016. "Connecting Mobility to Infectious Diseases: The Promise and Limits of Mobile Phone Data." *The Journal of Infectious Diseases* 214 (suppl_4): S414–20.