The Cost of Privacy: Welfare Effects of the Disclosure of COVID-19 Cases

David Argente
Penn State

Chang-Tai Hsieh
University of Chicago

Munseob Lee
UC San Diego

July 2020

Outline

1. Motivation
2. Data and Empirical Facts
3. Spatial Model with Spread of Disease
4. Calibration and Simulation
5. Counter-factual Exercises

South Korea's Case

- In the fight against COVID-19, South Korea stands out:
  - Despite a large outbreak, South Korea was able to flatten the curve.
  - South Korea has not shut down its economy.
- South Korea (and other Asian countries)'s strategy:
  - Massive testing, combined with contact tracing
  - Disclosure of detailed route information of confirmed cases
- Disclosure possible under social acceptance of surveillance using mobile phone and credit card data.
- Allowed people to engage in targeted social distancing by avoiding places visited by those that tested positive.

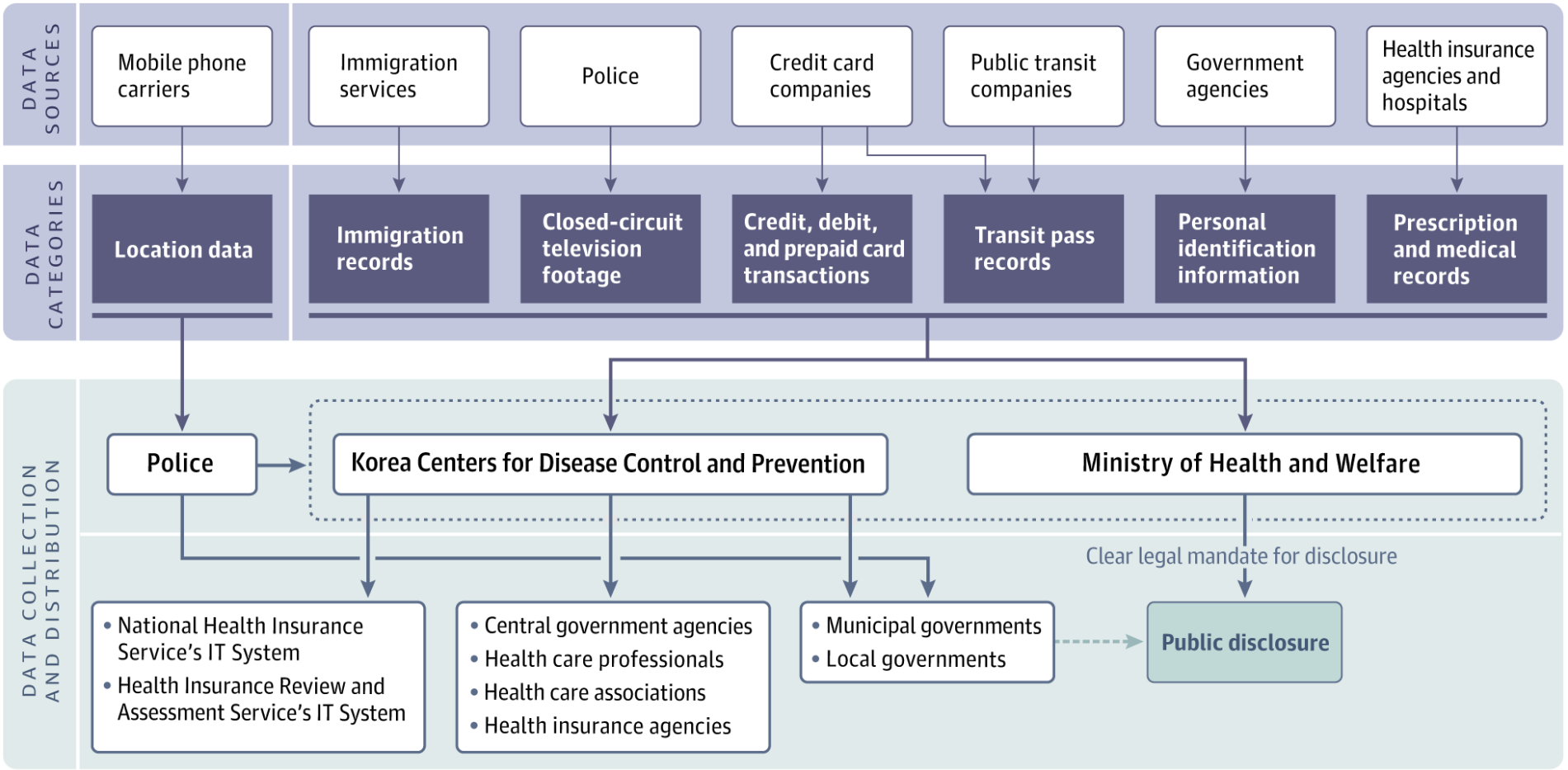
Public Disclosure: Official Website

Korean, male, born in 1987, living in Jungnang district.
Confirmed on January 30. Hospitalized in Seoul Medical Center.

January 24	Return trip from Wuhan without symptoms.
January 26	Merchandise store* at Seongbuk district at 11 am, fortune teller* at Seongdong district by subway at 12 pm, massage spa* by subway in the afternoon, two convenience stores* and two supermarkets*.
January 27	Restaurant* and two supermarkets* in the afternoon.
January 28	Hair salon* in Seongbuk district, supermarket* and restaurant* in Jungnang district by bus, wedding shop* in Gangnam district by subway, home by subway.
January 29	Tested at a hospital in Jungnang district.
January 30	Confirmed and hospitalized.

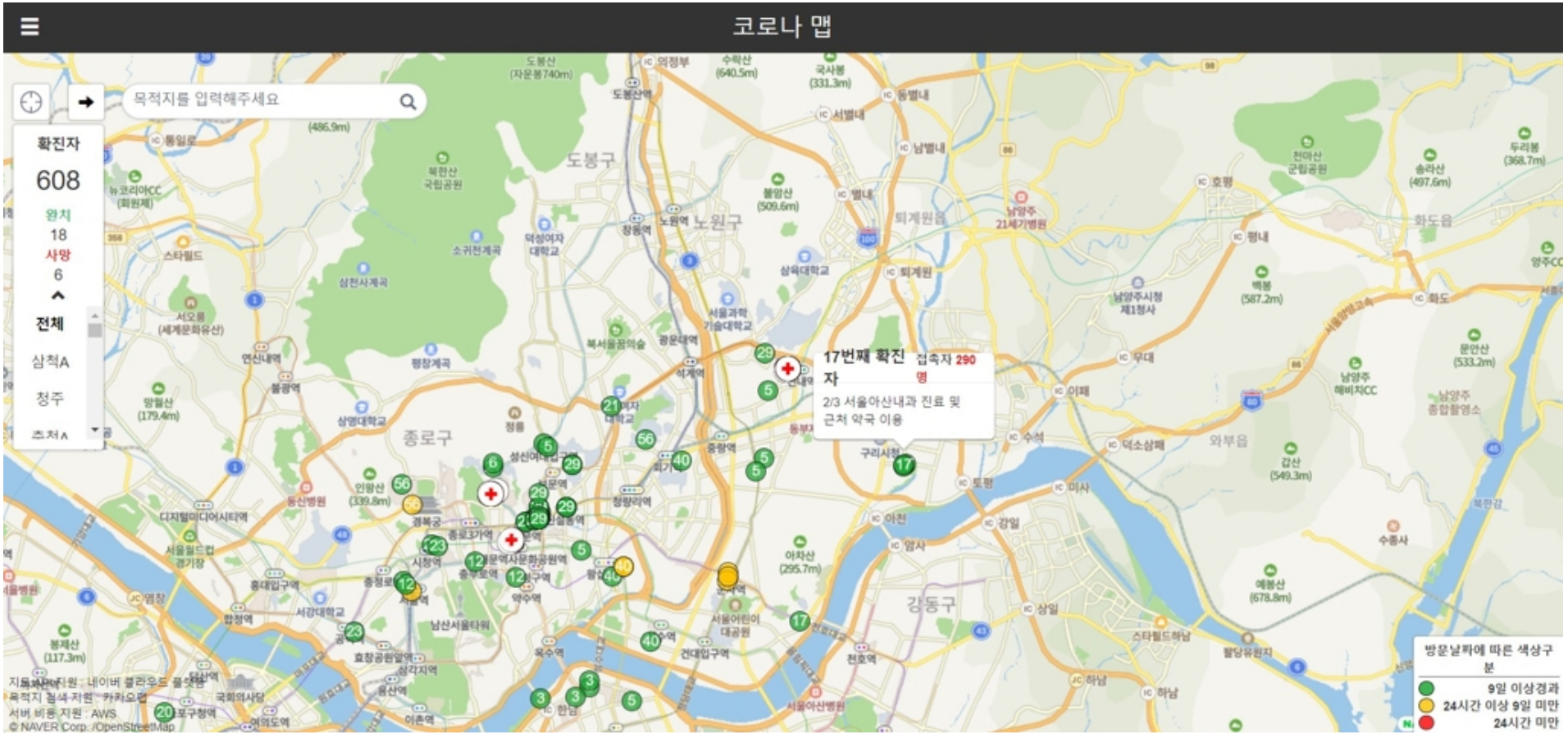
Note: The * denotes establishments whose exact names have been disclosed.

Public Disclosure: Technology



- Park, Choi and Ko, *JAMA*, 2020.

Public Disclosure: Mobile App



- February 24, 2020.

Effects of Public Disclosure

- On March 28th, a patient had briefly stopped by to order a drink at a coffee shop.
- On March 30th, district officials fumigated a coffee shop and the coffee shop was "disclosed" as a place visited by the patient.
  - That day, not a single customer showed up.

This Paper

- This paper: quantify the effect of public disclosure on the transmission of the virus and economic losses in Seoul.
  - Use detailed foot-traffic data to document the change in the flows of people across neighborhoods in Seoul in response to information.
  - Analyze the effect of the change in commuting flows in a SIR meta-population model
  - Endogenize these flows in a model of urban neighborhoods with commuting decisions.
- Important caveat: COVID-19-specific parameters are still uncertain.

Results

- **Public disclosure:** 602K cases, 12.5K deaths, and losses of about 0.6% of welfare loss over the next 2 years.

- Three alternative scenarios
  - No disclosure: 200K more infected people and 6.5K additional deaths, welfare losses lower.
  - Lockdown: welfare cost is 40 percent larger to achieve same number of cases as the full disclosure case.
  - Shutdown: costly in terms of output/welfare and not effective at delaying the spread of the disease relative to no disclosure case.

Outline

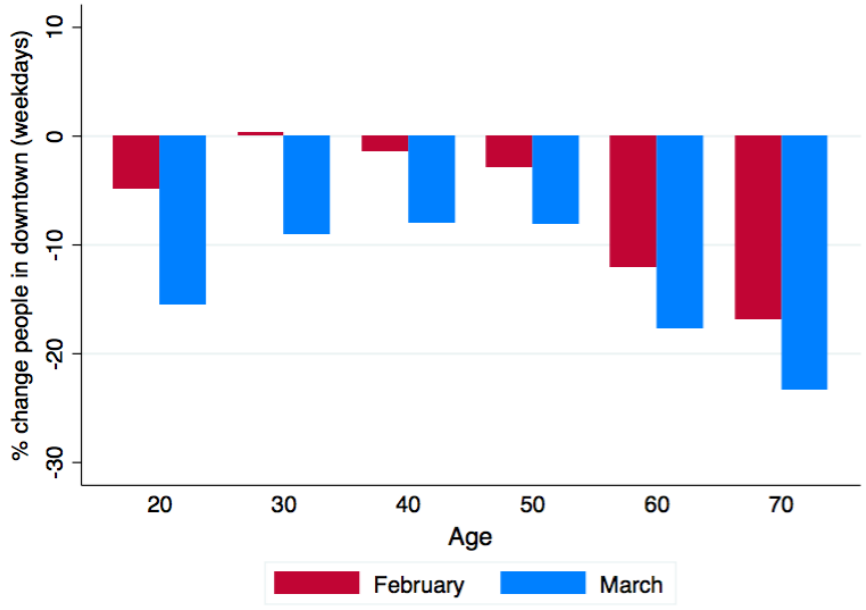
1. Motivation
2. Data and Empirical Facts
3. Spatial Model with Spread of Disease
4. Calibration and Simulation
5. Counter-factual Exercises

Data

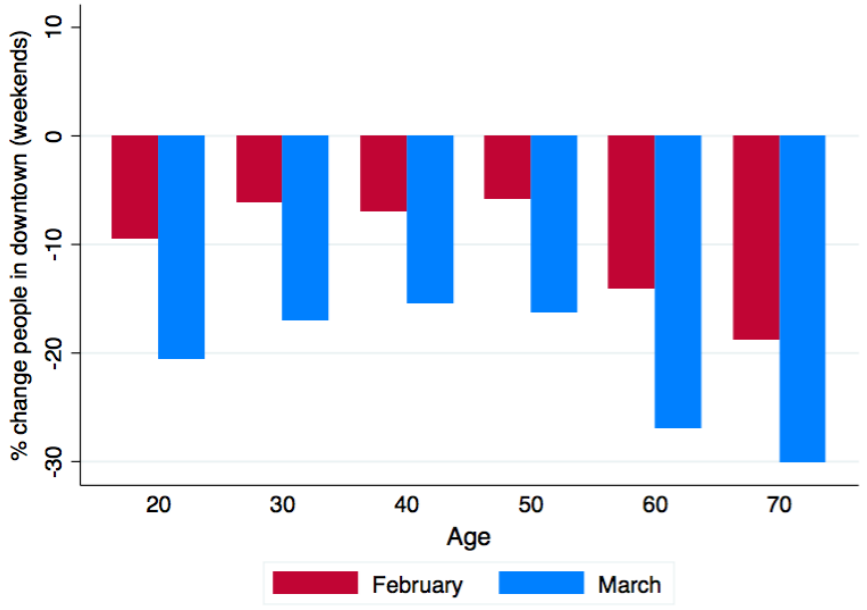
- Mobile Phone Data
  - Korean largest telecommunication company, SK Telecom.
  - Number of subscribers present at any point in time in 25 districts.
  - Location of a user is inferred by cellular tower triangulation.
  - The data splits users by the gender and by age group.
  - Two formats
    1. Hourly Foot Traffic Data ▸ Example
    2. Daily Bilateral Flow Data

- Seoul Survey
  - Labor market information.

Percent Change in # of People in Downtown during the Day: February and March compared to January



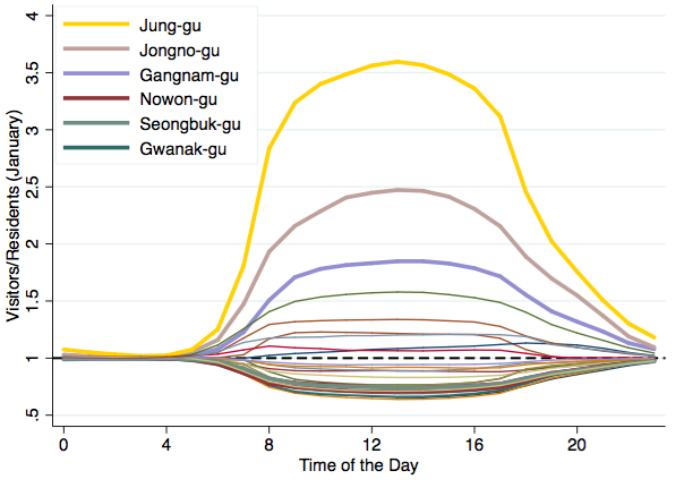
(a) Weekdays



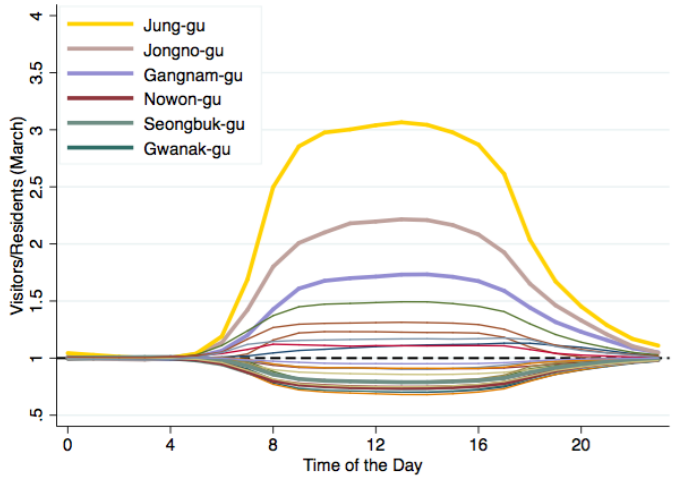
(b) Weekends

Share of Visitors over Residents by District

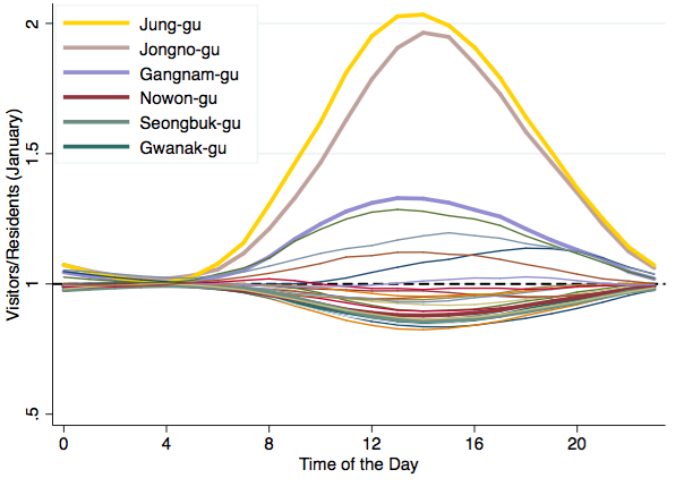
▸ Maps



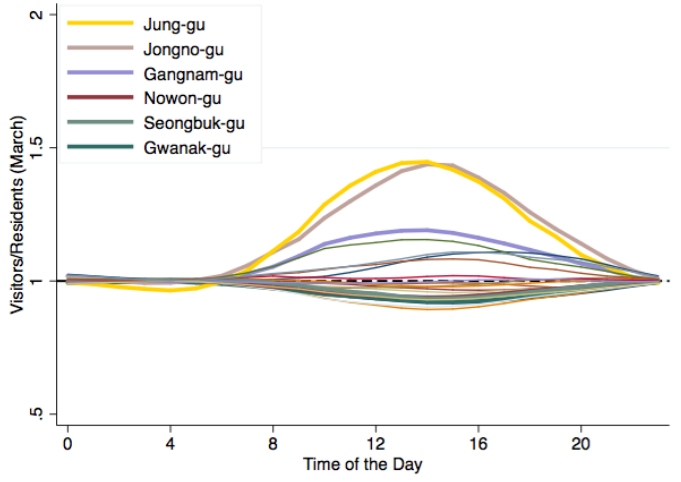
(a) Weekdays: January



(b) Weekdays: March

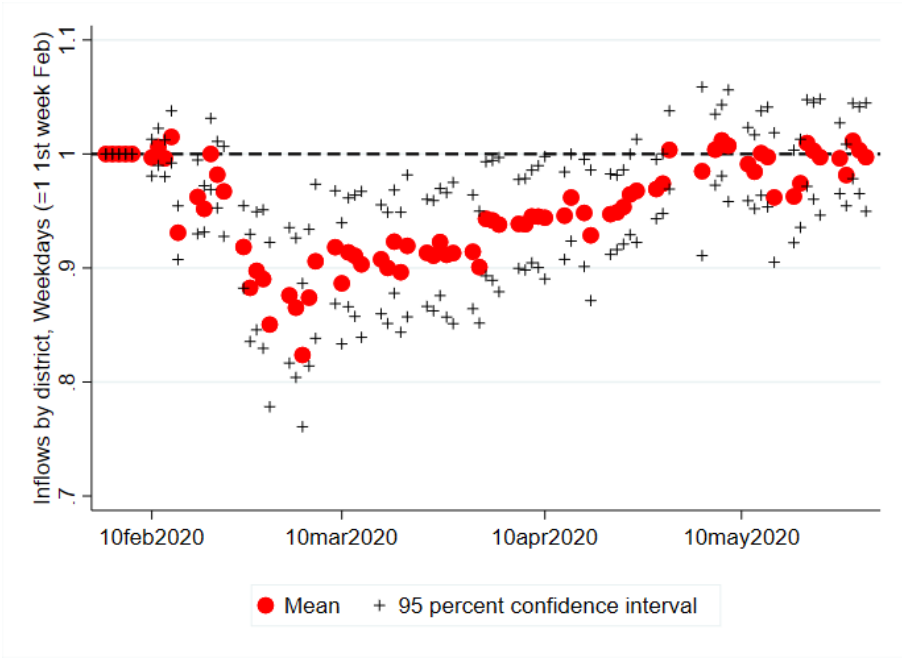


(c) Weekends: January

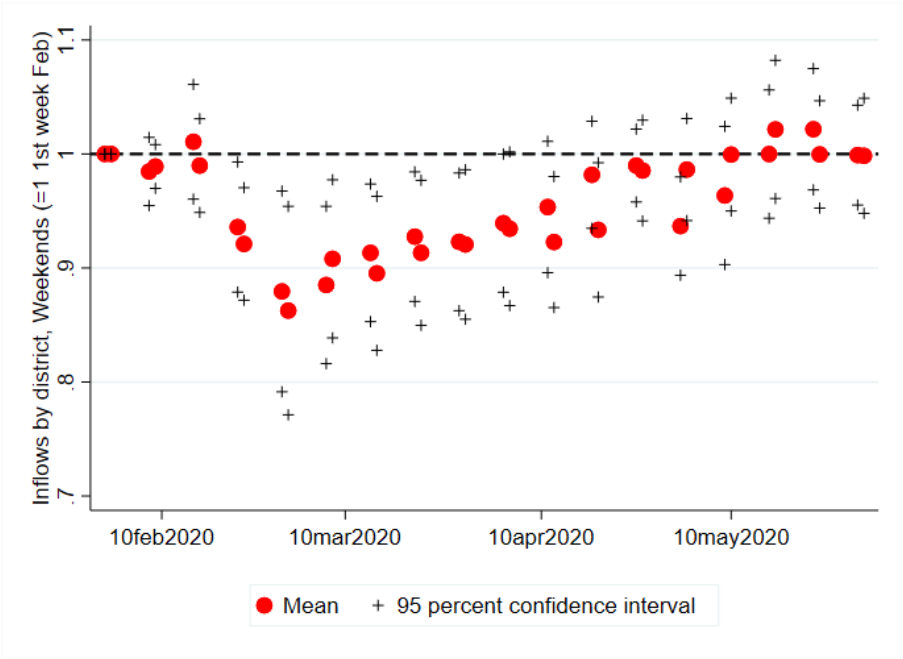


(d) Weekends: March

Change in Inflows by Districts



(a) Weekdays



(b) Weekends

Commuting Changes: Aggregate Flows

$$\ln \pi_{ijt} = \alpha + \iota \ln C_t + \gamma_i + \nu \tau_{ij} + \theta_j + e_{ijt}$$

where τ_{ij} is travel distance and C_t is the number of total cases for past 14 days in Seoul.

	(1)	(2)	(3)	(4)
$\ln C_t$	-0.0253***	-0.0433***	-0.0559***	-0.0689***
	(0.0021)	(0.0012)	(0.0021)	(0.0021)
Observations	67,500	67,500	27,500	27,500
R-squared	0.8048	0.8090	0.7762	0.7813
Days	Weekdays	Weekdays	Weekends	Weekends
Group	Under 60	Above 60	Under 60	Above 60

Commuting Changes: Destination Choices

$$\ln \pi_{ijt} = \alpha + \varphi_1 \ln C_{jt} + \varphi_2 \ln V_{jt} + \nu \tau_{ij} + \gamma_i + \theta_j + e_{ijt}$$

where τ_{ij} is travel distance and $C_{j,t}$ is the number of cases for past 14 days in district j and $V_{j,t}$ is the number of visits for past 14 days in district j.

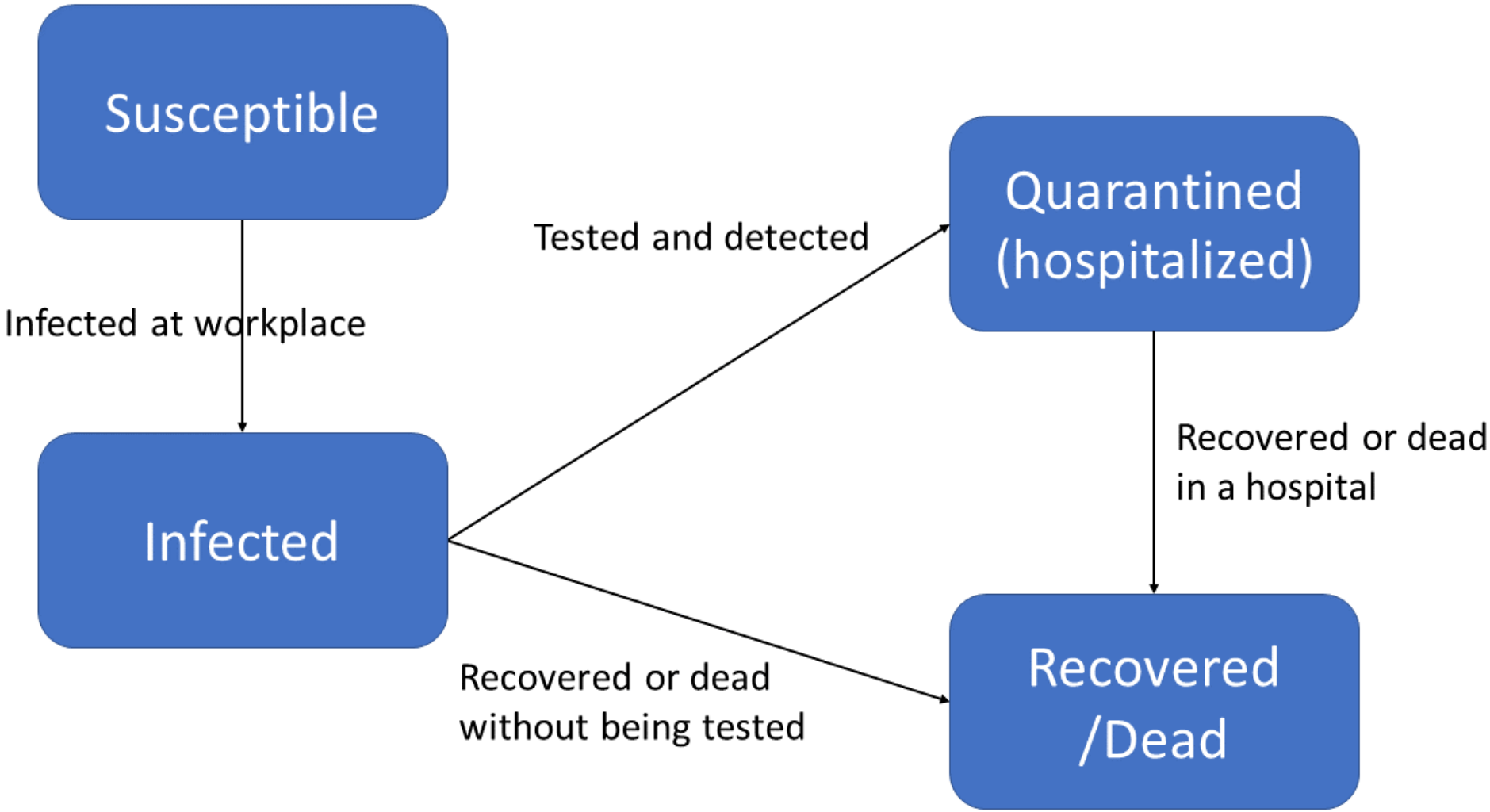
	(1)	(2)	(3)	(4)
$\ln C_{jt}$	-0.0181***	-0.0252***	-0.0293***	-0.0350***
	(0.0037)	(0.0037)	(0.0057)	(0.0058)
$\ln V_{jt}$	-0.0213***	-0.0347***	-0.0570***	-0.0666***
	(0.0024)	(0.0024)	(0.0037)	(0.0038)
Observations	67,500	67,500	27,500	27,500
R-squared	0.8046	0.8077	0.7752	0.7792
Days	Weekdays	Weekdays	Weekends	Weekends
Group	Under 60	Above 60	Under 60	Above 60

Outline

1. Motivation
2. Data and Empirical Facts
3. Spatial Model with Spread of Disease
4. Calibration and Simulation
5. Counter-factual Exercises

Susceptible, Infected, Quarantined, Recovered

Call Center Example



Susceptible, Infected, Quarantined, Recovered

$$
\begin{aligned}
\Delta S_i^a(t) &= -\beta \sum_j \left[\frac{\sum_{a,s} \pi_{sj}^a(t) I_s^a(t)}{\sum_{a,s} \pi_{sj}^a(t) N_s^a(t)} \times \pi_{ij}^a(t) S_i^a(t) \right] \\
\Delta I_i^a(t) &= \beta \sum_j \left[\frac{\sum_{a,s} \pi_{sj}^a(t) I_s^a(t)}{\sum_{a,s} \pi_{sj}^a(t) N_s(t)} \times \pi_{ij}^a(t) S_i^a(t) \right] - \gamma I_i^a(t) - d_I I_i^a(t) \\
\Delta Q_i^a(t) &= d_I I_i^a(t) - \tau^a Q_i^a(t) \\
\Delta R_i^a(t) &= \gamma I_i^a(t) + \tau^a Q_i^a(t)
\end{aligned} \tag{1}
$$

- $\pi_{ij}^a(t)$: people of age group a living in i's probability of working in j at time t.
- β: transmission rate.
- γ: daily recovery rate.
- d_I: daily rate at which infectious individuals are detected.
- $1/\tau^a$: average days spent in isolation.

Spatial Model: Workers

- Each district is populated by an exogenous measure of H^a_{Ri} workers.
  - Abstract from residential location choice.
  - Only traded goods ($p = 1$).
- Worker o of age a residing in i decides where to commute after observing her idiosyncratic heterogeneity (productivity/utility) $z^{a,k}_{jo}$.
- Weeks are divided into weekdays ($k = wd$) and weekends ($k = wn$).
- The aggregate consumption index is:

$$C^{a,wd}_{ijo} = \frac{c^{a,wd}_{ijo}}{d^a_{ij}}, \quad C^{a,wn}_{ijo} = \frac{z^{a,wn}_{jo}}{d^a_{ij}}$$

Indirect Utility

- Idiosyncratic component of productivity/utility ($z_{jo}^{a,k}$) is drawn from an independent Fréchet distribution:

$$F^{a,k}(z_{jo}^{a,k}) = e^{E_j^{a,k}(z_{jo}^{a,k})^{\varepsilon^k}}, \quad E_j^{a,k} > 0, \varepsilon^k > 1$$

- The indirect utility from residing in district i and commuting to district j:

$$u_{ijo}^{a,wd} = \frac{w_j^a z_{jo}^{a,wd}}{d_{ij}^a}, \quad u_{ijo}^{a,wn} = \frac{z_{jo}^{a,wk}}{d_{ij}^a}$$

  where w_j^a is wage per efficiency unit.

- Distance: $d_{ij}^a = e^{\kappa\tau_{ij} + \delta^a \ln C_j + \xi^a \ln V_j + \alpha^a \ln C}$
  - τ_{ij}: distance in kilometers
  - C_j: number of confirmed cases living in j
  - V_j: number of infected people who visited j
  - C: total number of infected people

Home Sector

- Workers can choose to work or stay at the home sector.
- The home sector has its own $E^{a,k}_{\text{home}}$ and w^{a}_{home}.
- No commuting cost or risk of being infected $\rightarrow d^{a,k}_{i,\text{home}} = 1$ for all i.

Weekdays

- Probability that a worker commutes for work to district j is

$$\pi_{ij|i}^{\text{a,wd}} = \frac{E_j^{a,wd}(w_j^a/d_{ij}^a)^{\varepsilon^{wd}}}{\sum_s E_s^{a,wd}(w_s^a/d_{is}^a)^{\varepsilon^{wd}}}$$

- Commuting market clearing condition:

$$H_{Mj}^{\text{a,wd}} = \sum_i \pi_{ij|i}^{\text{a,wd}} H_{Ri}^a = \sum_i \frac{E_j^{a,wd}(w_j^a/d_{ij}^a)^{\varepsilon^{wd}}}{\sum_s E_s^{a,wd}(w_s^a/d_{is}^a)^{\varepsilon^{wd}}} H_{Ri}^a$$

- Residence employment (H_{Ri}^a) is defined as:

$$H_{Ri}^a \equiv N_i^a - Q_i^a - D_i^a.$$

where N_i^a is population of age a, Q_i^a is the number of quarantined people and D_i^a is total death in district i.

Weekends

- Probability that a worker visits district j is

$$\pi_{ij|i}^{\text{a,wn}} = \frac{E_j^{a,wn}\left(d_{ij}^a\right)^{-\varepsilon^{wn}}}{\sum_s E_s^{a,wn}\left(d_{ij}^a\right)^{-\varepsilon^{wn}}}$$

- Commuting market clearing condition:

$$H_{Mj}^{\text{a,wn}} = \sum_i \pi_{ij|i}^{\text{a,wn}} H_{Ri}^a = \sum_i \frac{E_j^{a,wn}\left(d_{ij}^a\right)^{-\varepsilon^{wn}}}{\sum_s E_s^{a,wn}\left(d_{ij}^a\right)^{-\varepsilon^{wn}}} H_{Ri}^a$$

Production

- The production technology for the traded good is:

$$Y_j^a = A_j^a H_{Mj}^{\mathrm{a,wd}} \mathbb{E}[z_j^{a,wd} | \text{People choose j}]$$

- Under perfect competition, $w_j^a = A_j^a$.

- The aggregate production in the economy is defined as:

$$Y \equiv \sum_{a,j} Y_j^a = \sum_{a,j} A_j^a H_{Mj}^{\mathrm{a,wd}} \mathbb{E}[z_j^{a,wd} | \text{People choose j}]$$

where the sum across j incorporates all districts including the home sector.

Outline

1. Motivation
2. Data and Empirical Facts
3. Spatial Model with Spread of Disease
4. Calibration and Simulation
5. Counter-factual Exercises

Gravity

- From the commuting probabilities, before the outbreak of the virus:

$$ln\pi_{ij}^{k} = -\nu^{k}\tau_{ij} + \theta_i + \theta_j + e_{ij}^{k}$$

- π_{ij}^{k}: commuting probabilities from cell phone data.
- τ_{ij}: travel distances from the data.
- e_{ij}^{k}: stochastic error capturing measurement error in travel distances.
- $\nu^{k} = \varepsilon^{k}\kappa$ is the semi-elasticity of commuting flows wrt travel distances.
  - $\nu^{wd} = 0.1413$. $\nu^{wn} = 0.1666$.

Calibration of ε

- The coefficient of variation in wages within a region is:

$$\frac{\text{Variance}}{\text{Mean}^2} = \frac{\Gamma(1-\frac{2}{\varepsilon})}{\Gamma(1-\frac{1}{\varepsilon})^2} - 1$$

  where Γ is a Gamma function.

- $\varepsilon^{wd} = 4.1642$.
- $\kappa = \nu^{wd} \times \varepsilon^{wd} = 0.0339$.
- $\varepsilon^{wn} = \nu^{wn} / \kappa = 4.9144$

Calibration of $E_i^{a,k}$

- We estimate $E_j^{a,wd}$ and $E_j^{a,wn}$ using the following conditions:

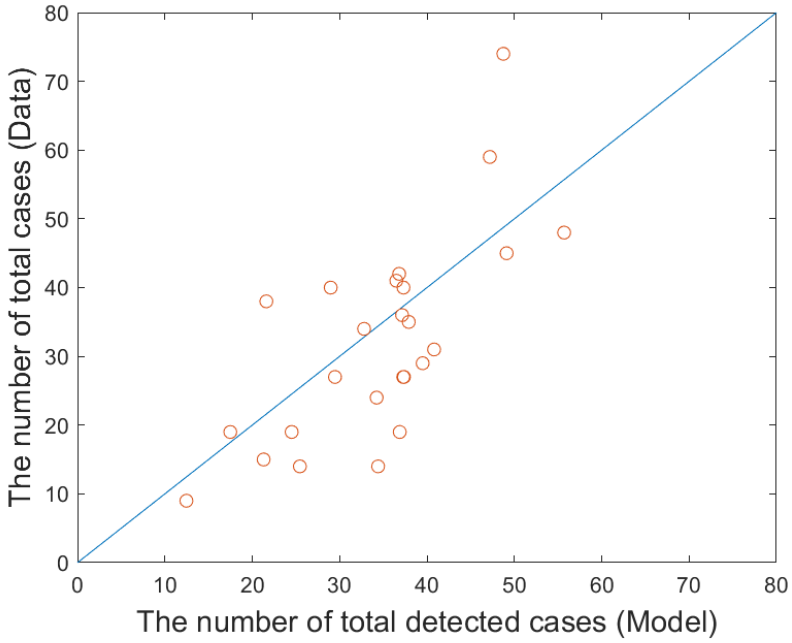
$$\mathbb{E}\left[H_{Mj}^{a,wd} - \sum_{i=1}^{S} \frac{E_j^{a,wd}(w_j^a)^{\varepsilon^{wd}}/e^{\nu^{wd}\tau_{ij}}}{\sum_{s=1}^{S} E_s^{a,wd}(w_s^a)^{\varepsilon^{wd}}/e^{\nu^{wd}\tau_{is}}} H_{Ri}^a\right] = 0$$

$$\mathbb{E}\left[H_{Mj}^{a,wn} - \sum_{i=1}^{S} \frac{E_j^{a,wn}/e^{\nu^{wn}\tau_{ij}}}{\sum_{s=1}^{S} E_s^{a,wn}/e^{\nu^{wn}\tau_{is}}} H_{Ri}^a\right] = 0$$

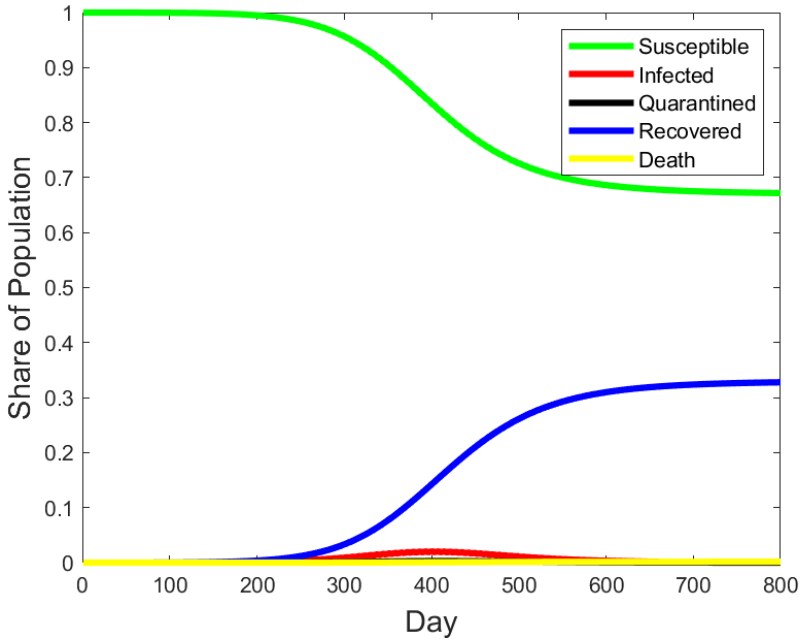
Calibration of COVID-19-specific Parameters

Parameter	Value (young, old)	Definition/Target Moments
Externally Calibrated		
γ	1/18	Daily rate at which active cases recover.
τ^a	1/8.5, 1/10.2	Mean duration of hospitalization.
ψ^a	0.21%, 2.73%	Case fatality rate.
α^a	0.007986, 0.011717	Elasticity to total confirmed cases by age.
δ^a	0.004665, 0.006219	Elasticity to local confirmed cases by age.
ξ^a	0.007726, 0.010460	Elasticity to local visits by infected by age.
Internally Calibrated		
β	0.15513	Total detected cases by May 31th.
d_I	0.01632	Fraction of undetected infections (Stock *et al.*, 2020).

Predicted Spread of Disease

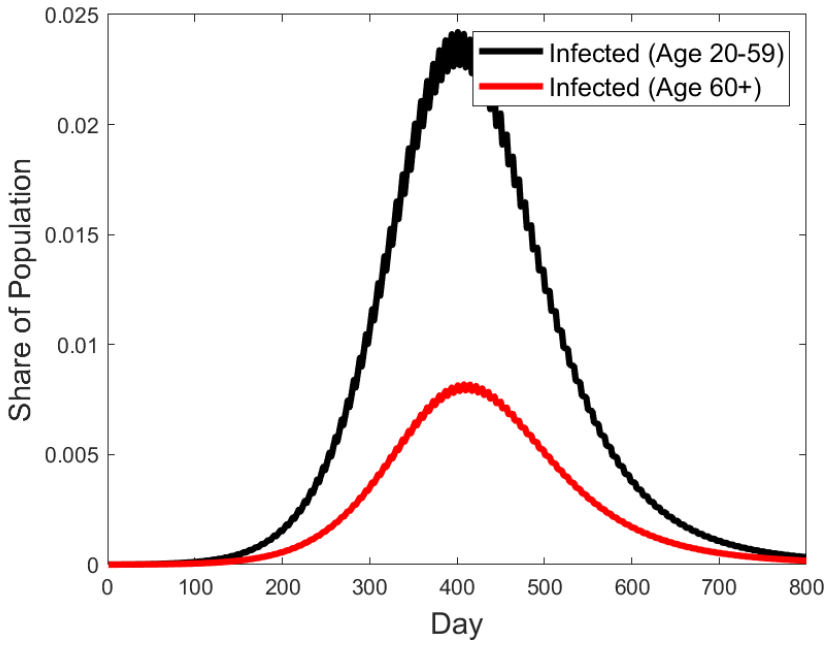


(a) Data vs Model

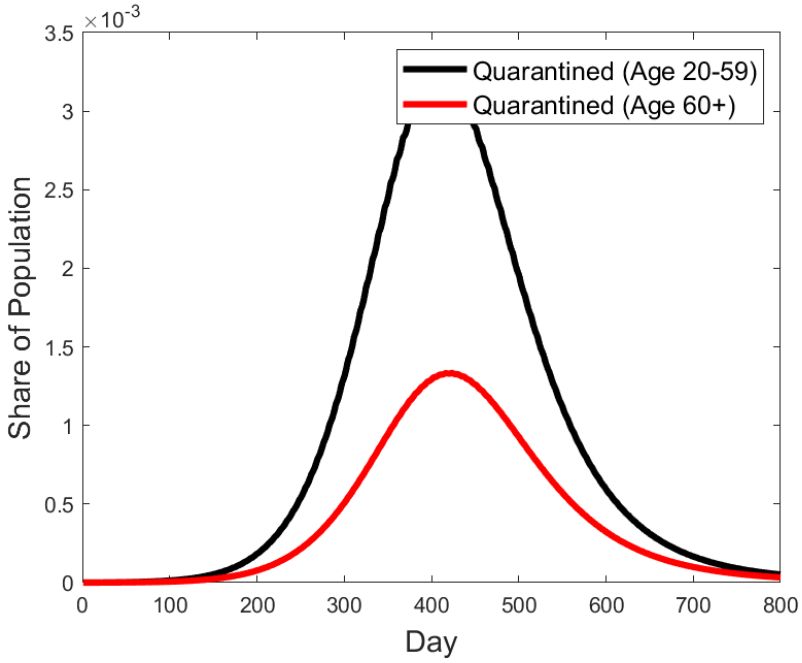


(b) SIR Dynamics

Infected and Quarantined Population by Age Groups

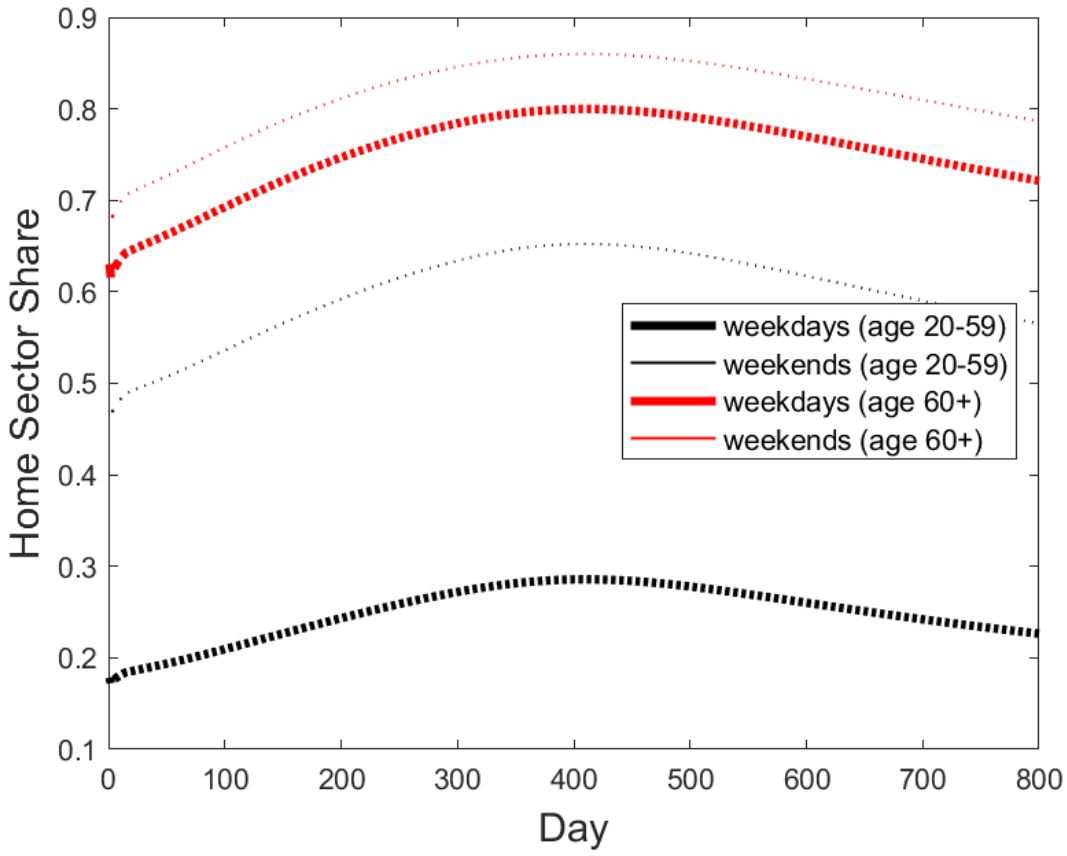


(a) Share of Infected People

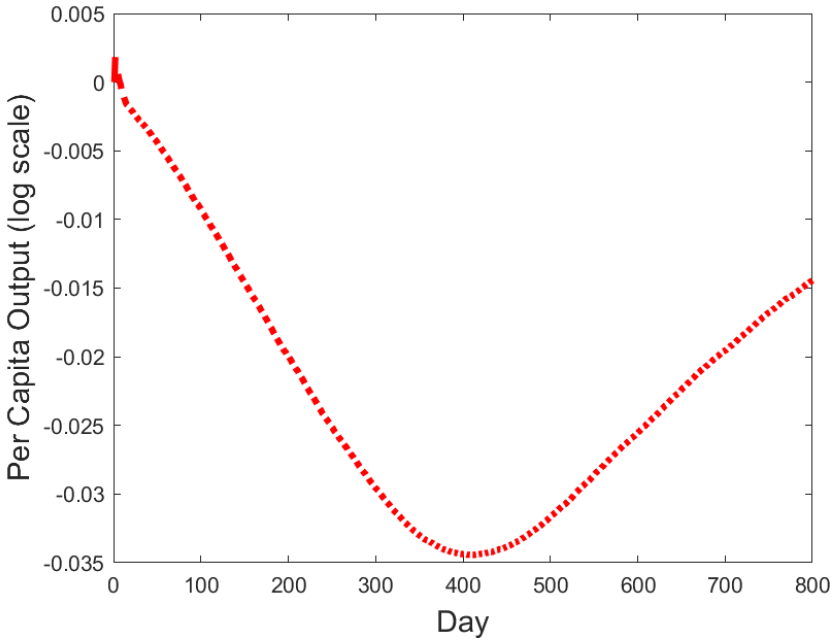


(b) Share of Quarantined People

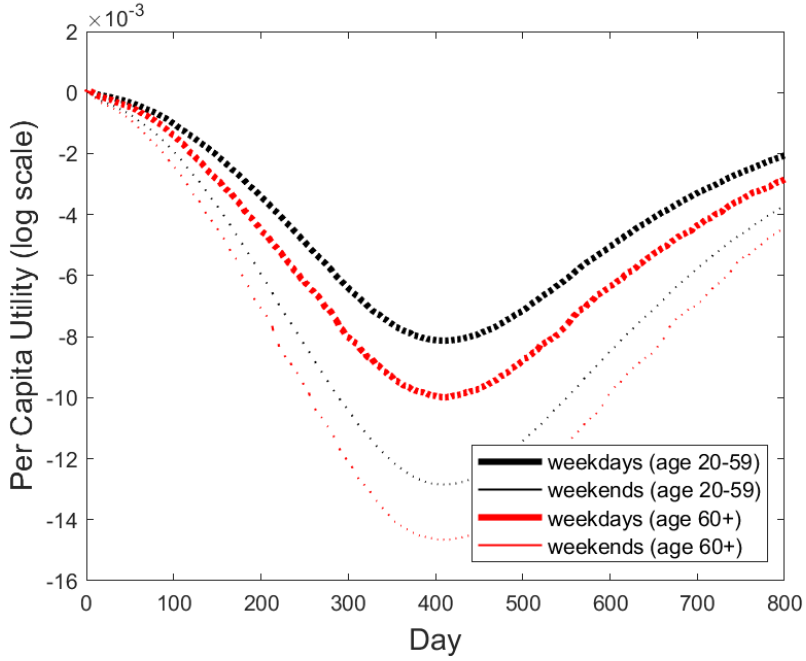
Home Sector Share



Aggregate Output and Welfare



(a) Aggregate Output

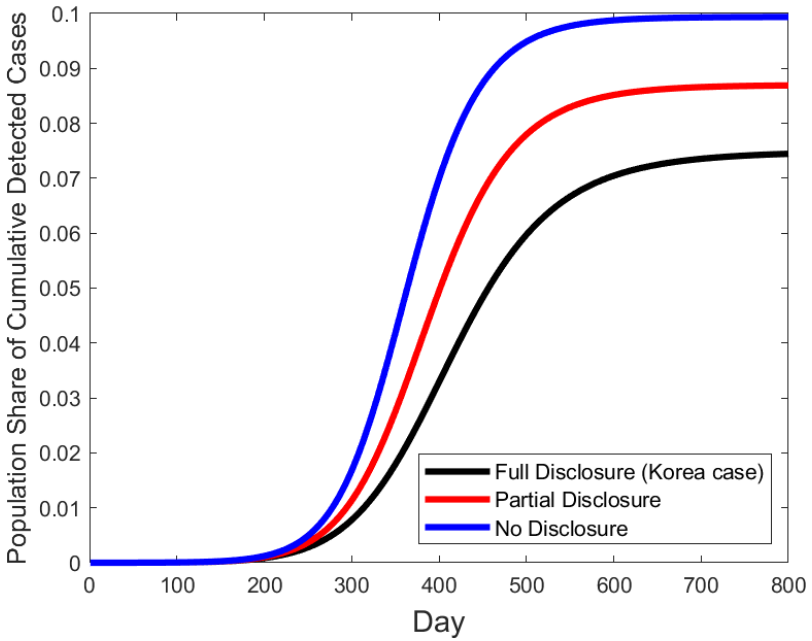


(b) Aggregate Welfare

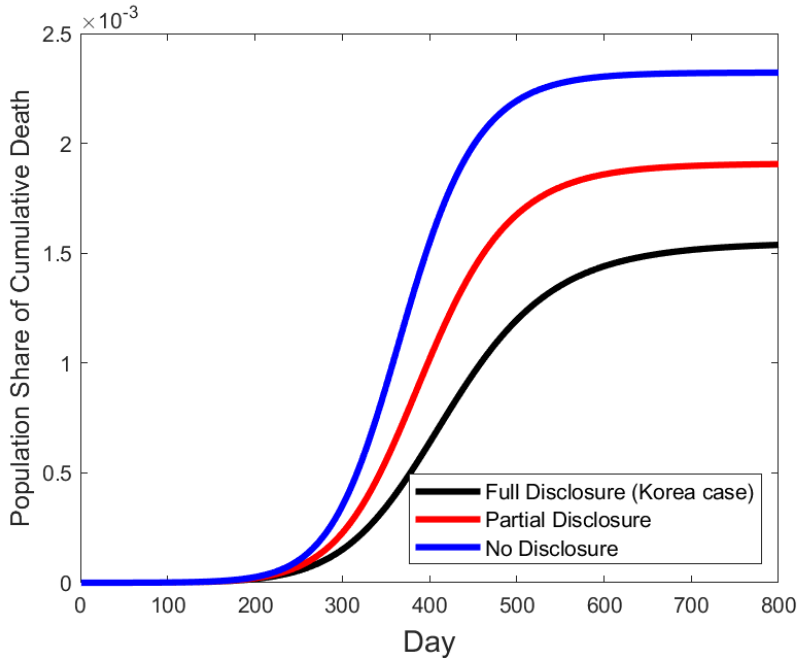
Outline

1. Motivation
2. Data and Empirical Facts
3. Spatial Model with Spread of Disease
4. Calibration and Simulation
5. Counter-factual Exercises

Disclosure Policy: Detected Cases and Deaths



(a) Cumulative Detected Cases



(b) Cumulative Death

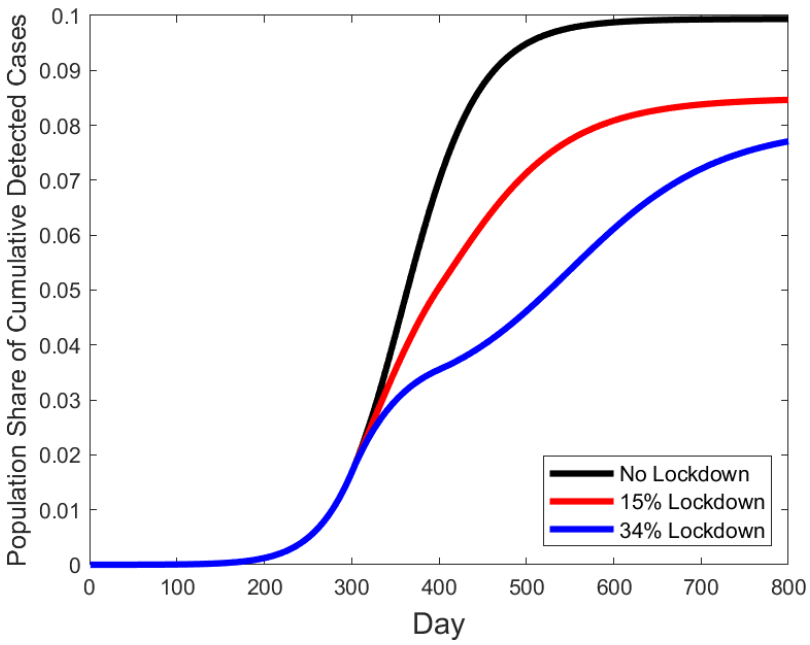
Disclosure Policy: Cases, Output and Welfare

	Original	Partial Disclosure	No Disclosure
Total # of Cases	602,999	707,475	810,158
Total # of Death (age 20-59)	5,003	5,796	6,540
Total # of Death (age 60+)	7,432	9,709	12,388
Output Loss per day	-2.29	-1.60	-0.98
Welfare Loss per day (age 20-59)	-0.55	-0.31	-0.13
Welfare Loss per day (age 60+)	-0.67	-0.39	-0.18

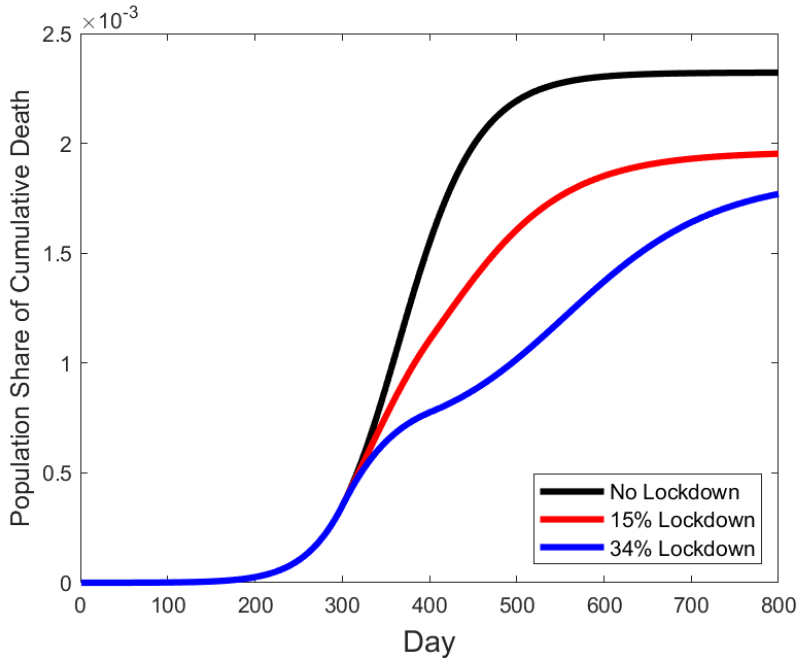
* in percent

- Trade-off between public health and output/welfare

Lockdown Policy for three months: Detected Cases and Deaths



(a) Cumulative Detected Cases



(b) Cumulative Death

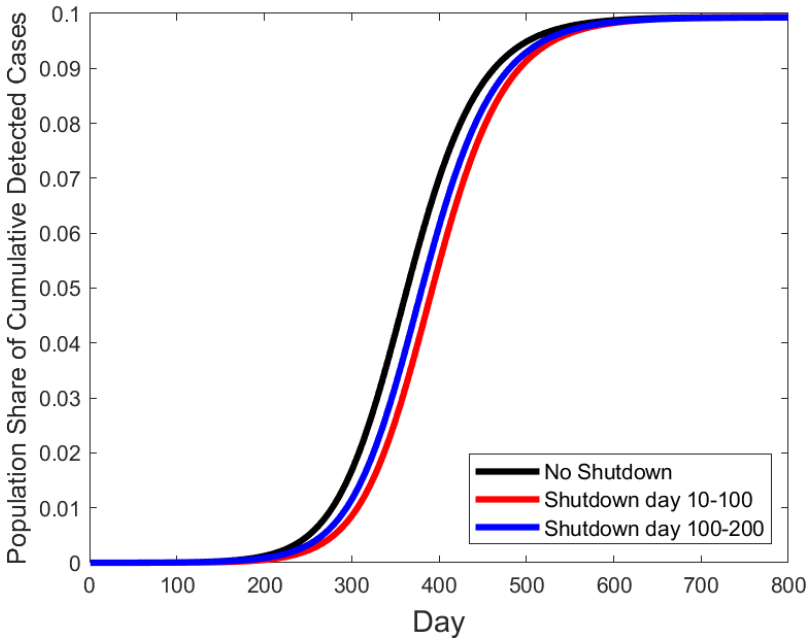
Lockdown Policy: Cases, Output and Welfare

Lckdwn day 300-400 (100 days)	No Lckdwn	15% Lckdwn	34% Lckdwn
Total # of Cases	810,158	686,495	604,223
Total # of Death (age 20-59)	6,540	5,558	4,894
Total # of Death (age 60+)	12,388	10,264	8,903
Output Loss per day	-0.98	-1.33	-1.93
Welfare Loss per day (age 20-59)	-0.13	-0.56	-1.15
Welfare Loss per day (age 60+)	-0.18	-0.23	-0.28

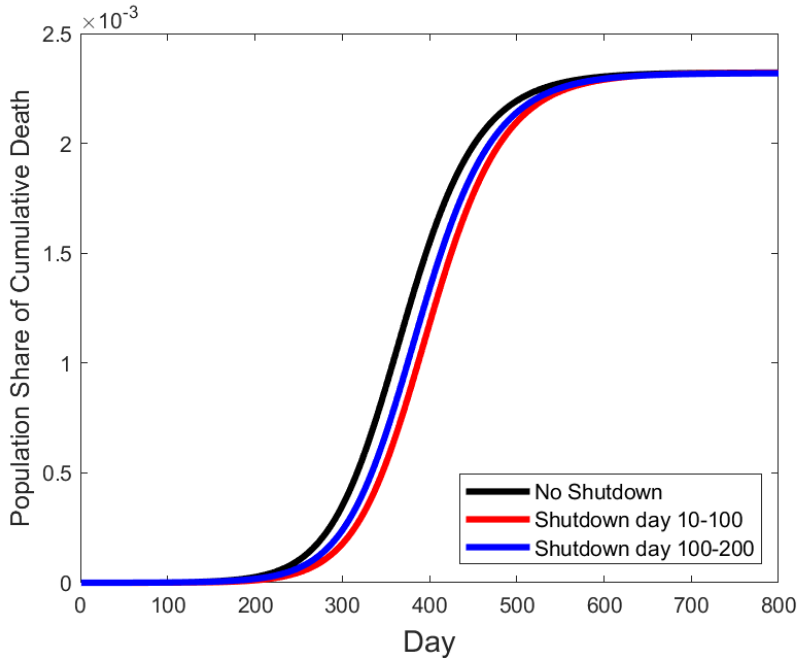
* in percent

- Note that we assume no disclosure here
- Higher welfare loss compared to disclosure

Shutdown Policy: Detected Cases and Deaths



(a) Cumulative Detected Cases



(b) Cumulative Death

Shutdown Policy: Cases, Output and Welfare

	No Shtdwn	Shtdwn 10-100	Shtdwn 100-200
Total # of Cases	810,158	809,823	809,304
Total # of Death (age 20-59)	6,540	6,537	6,533
Total # of Death (age 60+)	12,388	12,381	12,369
Output Loss per day	-0.98	-1.13	-1.20
Welfare Loss per day (age 20-59)	-0.13	-0.76	-0.83
Welfare Loss per day (age 60+)	-0.18	-0.46	-0.49

* in percent

Summary

- Meta-population SIR model and a model of commuting choice to measure the effect of the disclosure of information about COVID-19.
- Compared to a scenario without disclosure, public disclosure reduces the number of cases by 140k and deaths by 7k in Seoul over two years.
- Compared to a lockdown with the same number of cases as the disclosure strategy, welfare losses are 40 percent lower.
- We do not attempt to measure the cost of the loss of privacy.
- Targeted social distancing more effective way to reduce the spread of the disease while minimizing economic costs of isolation.

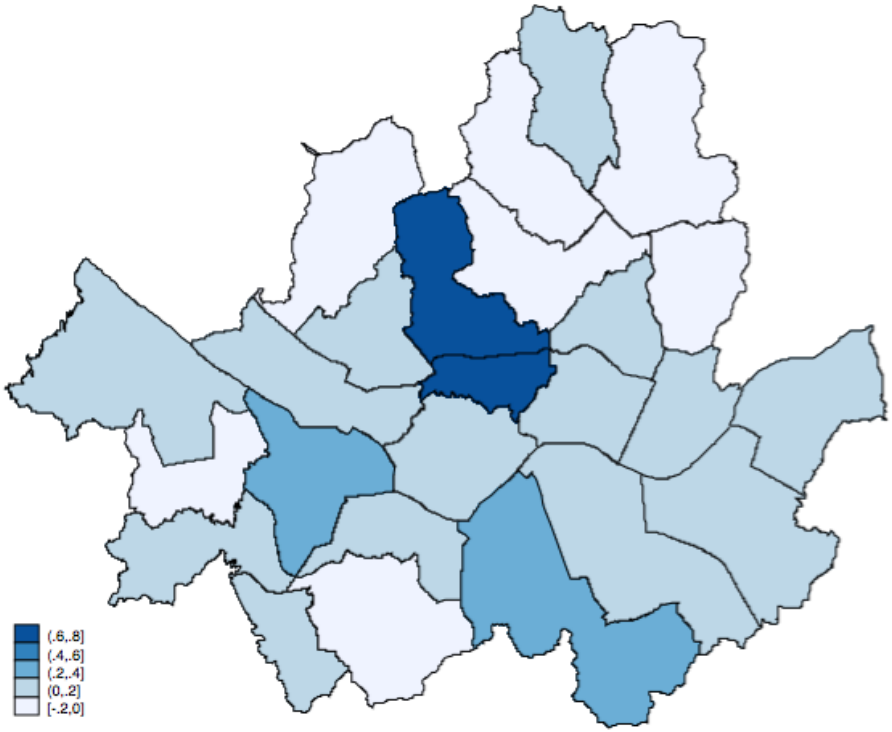
Foot Traffic Data

Back

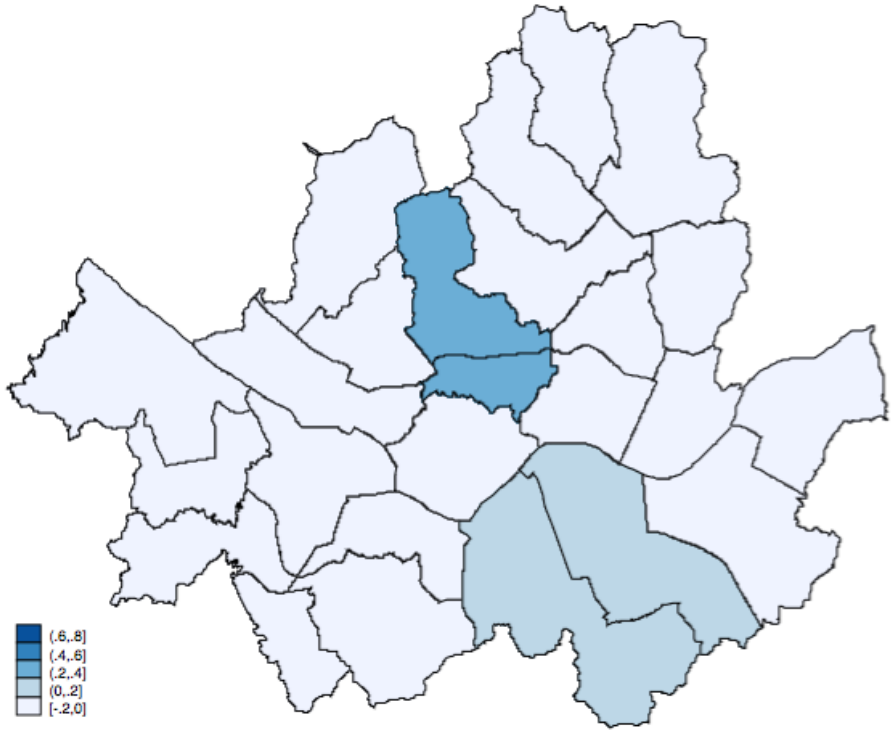
Date	Time	Age group	Gender	City	District	Number of people
January 5	3 am	20-30	Female	Seoul	Gangnam-gu	42,570
January 5	3 am	20-30	Male	Seoul	Gangnam-gu	41,590
⋮						
January 5	3 pm	20-30	Female	Seoul	Gangnam-gu	63,330
January 5	3 pm	20-30	Male	Seoul	Gangnam-gu	54,520

% Change Visitors over Residents (Weekend) - Age 60+

Back



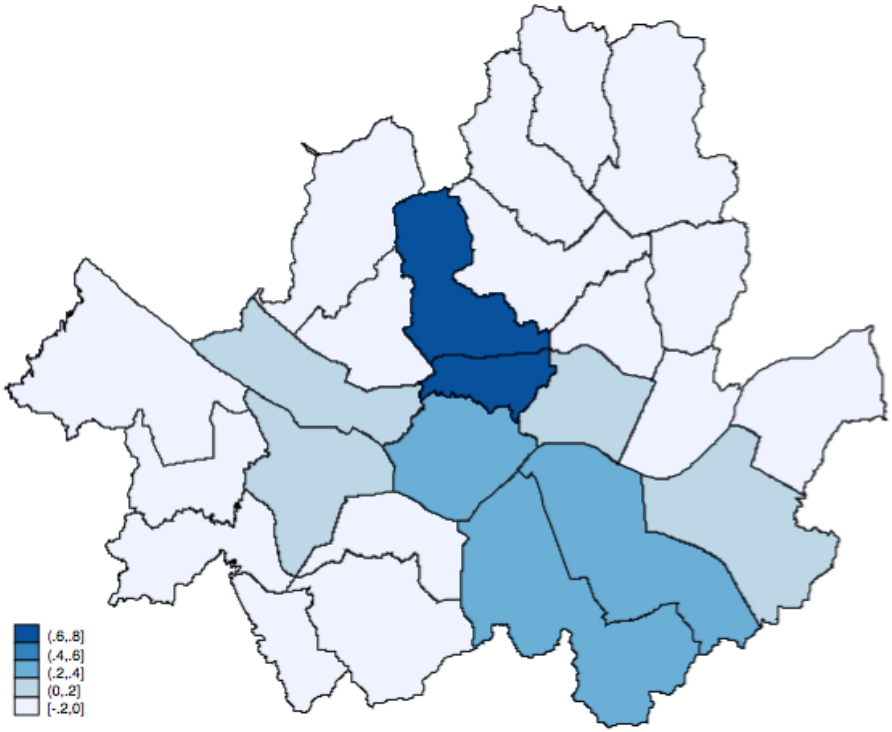
(a) January



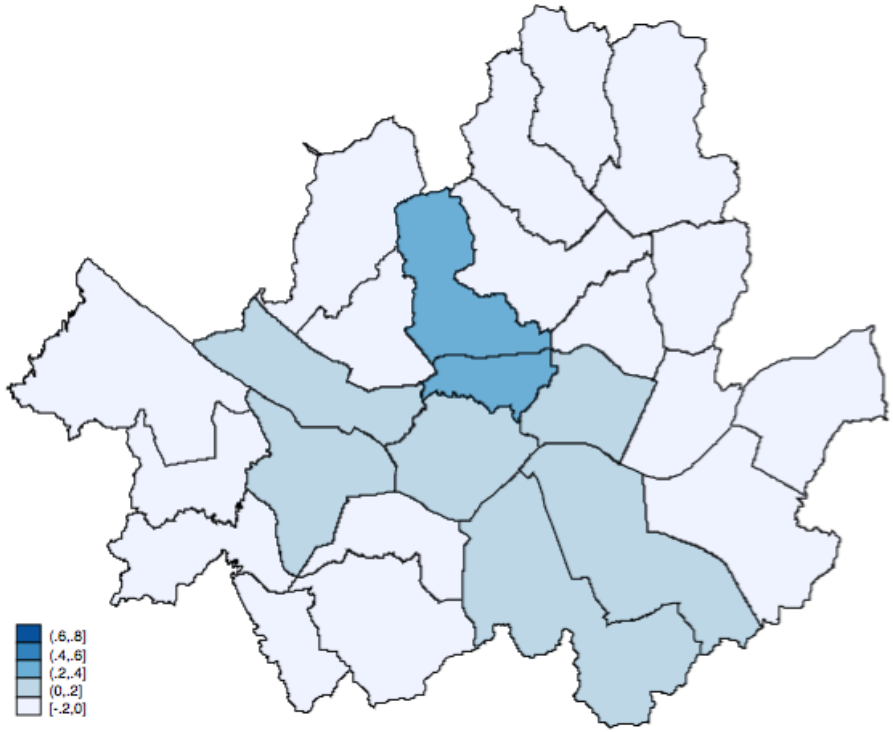
(b) March

% Change Visitors over Residents (Weekend) - Age 20-59

Back



(a) January



(b) March

Susceptible, Infected, Quarantined, Recovered

Back



- Floor plan of the 11th floor of building X, site of a coronavirus disease outbreak, Seoul, South Korea, 2020. (Park et al. 2020)