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

January 2021
ASSA 2021 Virtual Annual Meeting

South Korea's Case

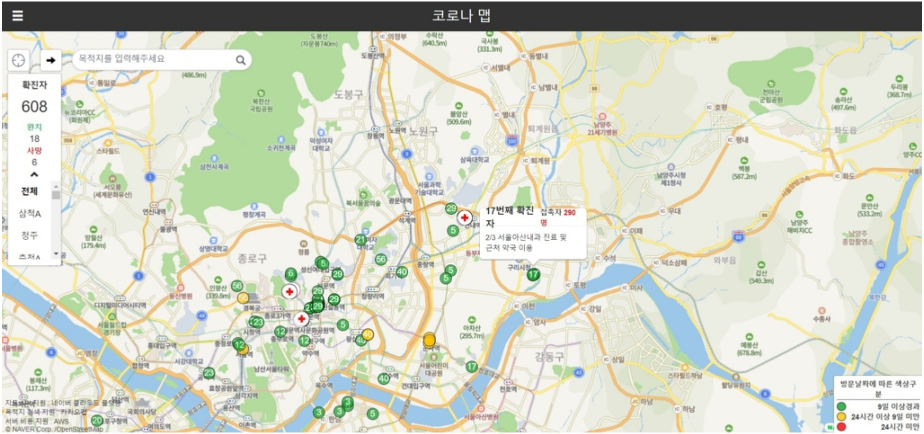
- Disclosure of detailed information of confirmed cases.
  - Text messages, official websites, mobile apps.
- Targeted social distancing: avoid places where transmission risk is high
- Self-selection into changing commuting: own cost-benefit analysis, exploit heterogeneity in the benefits and costs of social distancing.
- Reduce the transmission of virus and the costs of social isolation.

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: Mobile App - February 24, 2020



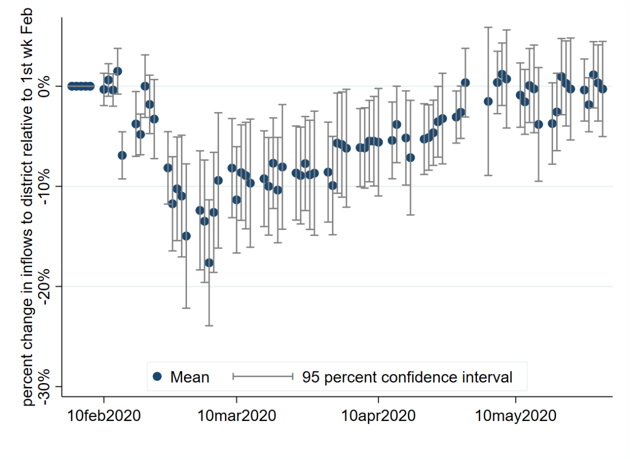
This Paper

- This paper: quantify the effect of public disclosure on the transmission of the virus and economic losses in Seoul.
    - Use detailed mobile phone 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.
- Findings:
    - change in commuting patterns due to public disclosure lowers the number of cases and deaths
    - economic cost of lockdown is almost four times higher compared to the disclosure scenario

Data

- Mobile Phone Data
  - Korean largest telecommunication company, SK Telecom.
  - data on daily bilateral commuting flows across Seoul's districts from January 2020 to May 2020.
  - A person's movement is included when she stays in the origin district for more than two hours, commutes to another district and stays in that district for more than two hours.
  - The data splits users by the gender and by age group.

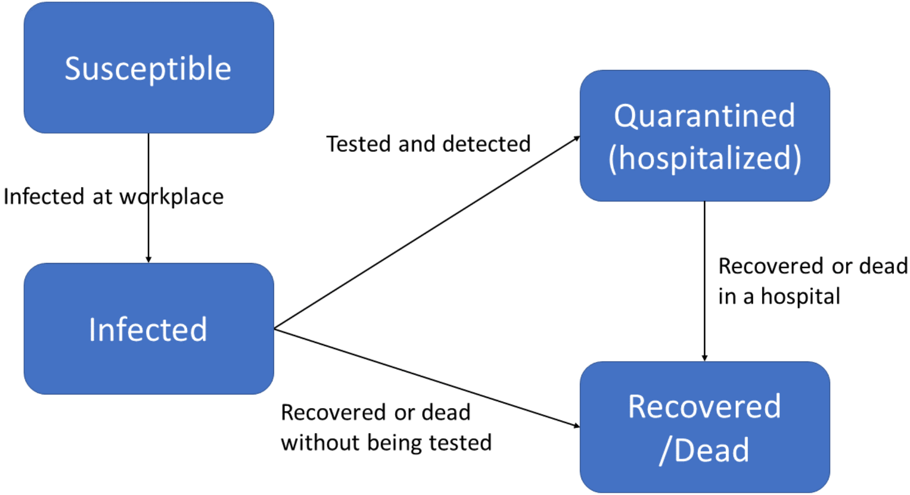
Change in Weekday Inflows into Districts in Seoul



- Traffic declines in districts with a larger number of cases and visits.

Regression

Susceptible, Infected, Quarantined, Recovered



Susceptible, Infected, Quarantined, Recovered ▸ Full SIR Model

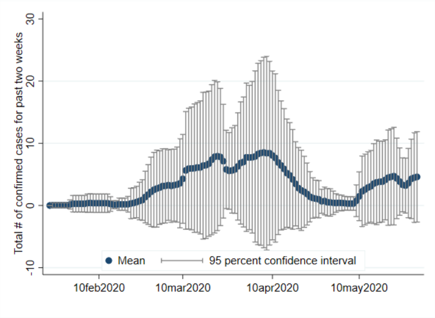
$$\Delta I_i^a(t) = \beta \sum_{j \neq \text{home}} \left[\underbrace{\frac{\sum_s \sum_a \pi_{sj}^a(t) I_s^a(t)}{\sum_s \sum_a \pi_{sj}^a(t) N_s(t)}}_{\text{Share infected in } j} \times \underbrace{\pi_{ij}^a(t) S_i^a(t)}_{\#\text{ of Susceptible from } i \text{ in } j} \right] - \gamma I_i^a(t) - d_I I_i^a(t)$$

- $\pi_{ij}^a(t)$: 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.

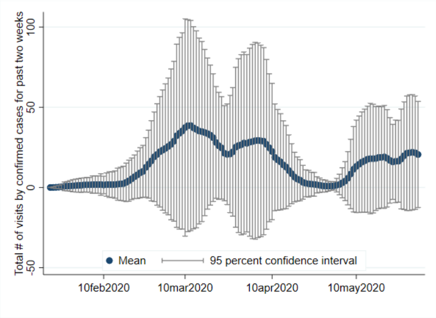
Spatial Model ▸ Full Spatial Model

- Quantitative model of internal city structure.
  - Allow for heterogeneity across age groups (young and old).
  - Weeks are divided into weekdays and weekends.
  - Districts differ in productivity (weekdays) or amenities (weekends)
  - Workers can choose to work from home.
- Distance: $\ln d_{ij}^{a}(t) = \kappa\tau_{ij} + \delta^{a}\ln C_j(t) + \xi^{a}\ln V_j(t) + \zeta^{a}(t)$
  - τ_{ij}: travel distance between i and j
  - $C_j(t)$: the number of *residents* of j confirmed as COVID patients in the two weeks prior to time t
  - $V_j(t)$: the number of *visits* by confirmed COVID patients to neighborhood j in the two weeks prior to t
  - $\zeta^{a}(t)$: the change in commuting costs that is independent of destination-specific information.
- Individual heterogeneity + local information $\implies$ Self-selection

Cases and Visits in Each District



(a) Cases for Past Two Weeks



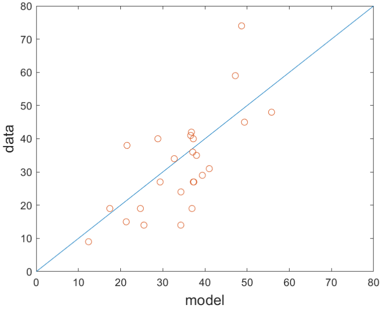
(b) Visits for Past Two Weeks

Calibration of COVID-19-specific Parameters

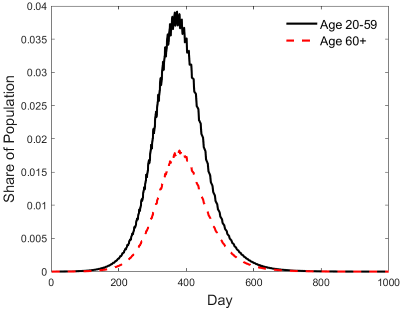
Frechet Parameters

Parameter	Value (young, old)	Definition
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.
δ^a	0.00209, 0.00247	Elasticity of commuting to local confirmed cases by age.
ξ^a	0.00138, 0.00096	Elasticity of commuting to local visits by infected by age.
Internally Calibrated		
β	0.1524	Transmission rate (target: total cases by May 31st).
d_I	0.0163	Daily detection rate (target: fraction of undetected infections)

Predicted Spread of Disease

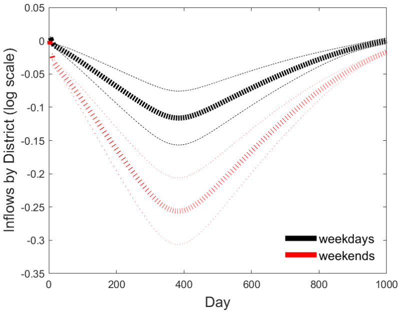


(a) Data vs Model by May 31st

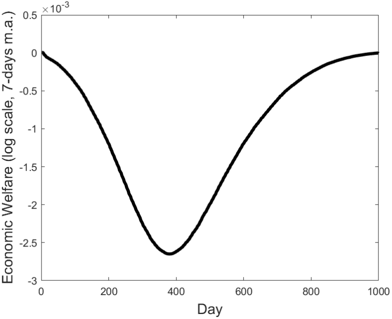


(b) Share of Infected People

Inflows by District and Economic Welfare



(a) Inflows by District



(b) Economic Welfare

Disclosure Policy: Cases ▸ Sensitivity

	Full Disclosure (Korea case)	No Disclosure
Total Cases	780,907	840,709
Total Death	18,743	20,744
age 20-59	6,255	6,687
age 60+	12,489	14,057
Welfare Loss per day (%)	0.14	0.07
age 20-59	0.13	0.07
age 60+	0.16	0.08

Disclosure Policy and Lockdown: Cases and Welfare

	Full Disclosure (Korea case)	22% Lockdown Days 280 to 380
Total Cases	780,907	780,692
Total Death	18,743	20,488
age 20-59	6,255	6,106
age 60+	12,489	14,381
Welfare Loss per day (%)	0.14	0.50
age 20-59	0.13	0.64
age 60+	0.16	0.07

- Disclosure: same cases and 73% lower economic welfare losses.

Conclusion

- Information disclosure:
  - Targeted social distancing.
  - Self-selection.
- Reduce the spread of the virus while minimizing costs of isolation.

Commuting Flow Equation Estimation

Back

	ln Commuting Flows (November 2019)		Δ ln Commuting Flows (relative to week 1, Feb 2020)	
τ_{ij}	-0.1413 (0.0028)	-0.1666 (0.0034)	–	–
$\ln C_j(t)$	–	–	-0.0087 (0.0049)	-0.0103 (0.0043)
$\ln C_j(t) \times$ **weekend**	–	–	-0.0016 (0.0009)	-0.0019 (0.0008)
$\ln V_j(t)$	–	–	-0.0058 (0.0031)	-0.0040 (0.0026)
$\ln V_j(t) \times$ **weekend**	–	–	-0.0010 (0.0005)	-0.0007 (0.0005)
weekend	–	–	-0.1360 (0.0539)	-0.1008 (0.0502)
Period	Nov 2019	Nov 2019	Jan-May 2020	Jan-May 2020
Age Group	All	All	Under 60	Above 60
Days	Weekdays	Weekends	All	All
Fixed Effects	-	-	Time	Time
Cluster	-	-	Two-way (bootstrapped)	Two-way (bootstrapped)
Observations	625	625	95,000	95,000
R-squared	0.8603	0.8405	-	-
Root MSE	-	-	0.2375	0.2275

Commuting Flow Equation Estimation ◂ Back

$$\Delta \ln \pi^a_{ij}(t) = \delta^a \varepsilon^{wd} \ln C_j(t) + \delta^a(\varepsilon^{wn} - \varepsilon^{wd}) \ln C_j(t) \times \textbf{weekend} + \xi^a \varepsilon^{wd} \ln V_j(t) +$$
$$+ \delta^a(\varepsilon^{wn} - \varepsilon^{wd}) \ln V_j(t) \times \textbf{weekend} + \varphi^a \times \textbf{weekend} + \theta^a_i + \lambda^a_j + \zeta^a(t)$$

where $\zeta^a(t)$ are the date fixed effects.
The dependent variable is the daily *change* in the commuting flows relative to the first week of February 2020 computed from SK Telecom's data and **weekend** is an indicator variable for a day that falls on a weekend.

Susceptible, Infected, Quarantined, Recovered ◂ Back

$$
\begin{aligned}
\Delta S_i^a(t) &= -\beta \sum_{j \neq \text{home}} \left[\frac{\sum_s \sum_a \pi_{sj}^a(t) I_s^a(t)}{\sum_s \sum_a \pi_{sj}^a(t) N_s^a(t)} \times \pi_{ij}^a(t) S_i^a(t) \right] \\
\Delta I_i^a(t) &= \beta \sum_{j \neq \text{home}} \left[\frac{\sum_s \sum_a \pi_{sj}^a(t) I_s^a(t)}{\sum_s \sum_a \pi_{sj}^a(t) N_s^a(t)} \times \pi_{ij}^a(t) S_i^a(t) \right] - \gamma I_i^a(t) - d_I I_i^a(t) \\
\Delta Q_i^a(t) &= d_I I_i^a(t) - \rho^a Q_i^a(t) \\
\Delta R_i^a(t) &= \gamma I_i^a(t) + \rho^a Q_i^a(t) \\
\Delta N_i^a(t) &= N_i^a(t-1) - \Delta Q_i^a(t)
\end{aligned}
$$

- $\pi_{ij}^a(\mathrm{t})$: 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: Setup Back

- We assume individuals make commuting choices every day and we distinguish between weekdays and weekends.
- Utility of a worker of age a that lives in i and works in j during the weekdays:

$$U_{ij}^{a}(t) = z_{j}^{a,wd} / d_{ij}^{a}(t) \tag{1}$$

where $z_j^{a,wd}$ is idiosyncratic *productivity* from working in j during the weekday and $d_{ij}^{a}(t)$ is the cost of commuting from i to j.

- Utility of a worker of age a that lives in i and works in j during the weekends:

$$U_{ij}^{a}(t) = z_{j}^{a,wn} / d_{ij}^{a}(t) \tag{2}$$

where $z_j^{a,wn}$ denotes idiosyncratic *preferences* from leisure in neighborhood j during the weekends.

Distance and Discrete Choice

Back

- Distance: $\ln d_{ij}^{a}(t) = \kappa \tau_{ij} + \delta^{a} \ln C_j(t) + \xi^{a} \ln V_j(t) + \zeta^{a}(t)$
  - τ_{ij}: travel distance between i and j
  - $C_j(t)$: the number of *residents* of j confirmed as COVID patients in the two weeks prior to time t
  - $V_j(t)$: the number of *visits* by confirmed COVID patients to neighborhood j in the two weeks prior to t
  - $\zeta^{a}(t)$: the change in commuting costs that is independent of destination-specific information.
- Idiosyncratic component of productivity/utility ($z_{jo}^{a,k}$) is drawn from an independent Fréchet distribution:

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

Commuting Probabilities

Back

- The probability that a resident of neighborhood i chooses to work in j during the weekday is:

$$\pi_{ij}^{a}(t = weekday) = \frac{E_j^{a,wd} d_{ij}^{a}(t)^{-\varepsilon^{wd}}}{\sum_s E_s^{a,wd} d_{is}^{a}(t)^{-\varepsilon^{wd}}}$$

- Similarly, the probability she travels to neighborhood j during the weekend is:

$$\pi_{ij}^{a}(t = weekend) = \frac{E_j^{a,wn} d_{ij}^{a}(t)^{-\varepsilon^{wn}}}{\sum_s E_s^{a,wn} d_{is}^{a}(t)^{-\varepsilon^{wn}}}$$

Welfare ◂ Back

- Expected utility of an individual living in neighborhood i is

$$\mathbb{E}[U_i^a(t = weekday)] = \Gamma\left(1 - 1/\varepsilon^{wd}\right)\left(\sum_s E_s^{a,wd} d_{is}^a(t)^{-\varepsilon^{wd}}\right)^{1/\varepsilon^{wd}}$$

during the weekday and

$$\mathbb{E}[U_i^a(t = weekend)] = \Gamma(1 - 1/\varepsilon^{wn})\left(\sum_s E_s^{a,wn} d_{is}^a(t)^{-\varepsilon^{wn}}\right)^{1/\varepsilon^{wn}}$$

during the weekend where $\Gamma(\cdot)$ is a gamma function.

Gravity ◂ Back

- 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 Back

- 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 $E_i^{a,k}$ Back

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

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

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

Sensitivity to Transmission and Detection Rates ◂ Back

	20% lower β=0.1219		20% higher β=0.1829	
	No disclosure	Full disclosure	No disclosure	Full disclosure
Total # of Cases	81,314	58,384	1,143,903	1,090,291
Welfare Loss per day(%)	0.04	0.06	0.05	0.11
	Frac. of undetected=0.8		Frac. of undetected=0.95	
	$\beta = 0.1682, d_I = 0.0357$		$\beta = 0.1515, d_I = 0.0076$	
	No disclosure	Full disclosure	No disclosure	Full disclosure
Total # of Cases	907,202	776,173	565,072	538,609
Welfare Loss per day(%)	0.08	0.15	0.05	0.11