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 2021
CEMLA-FRBNY-ECB
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.

<table>
<thead>
<tr>
<th>Date</th>
<th>Activity Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 24</td>
<td>Return trip from Wuhan without symptoms.</td>
</tr>
<tr>
<td>January 26</td>
<td>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*.</td>
</tr>
<tr>
<td>January 27</td>
<td>Restaurant* and two supermarkets* in the afternoon.</td>
</tr>
<tr>
<td>January 28</td>
<td>Hair salon* in Seongbuk district, supermarket* and restaurant* in Jungnang district by bus, wedding shop* in Gangnam district by subway, home by subway.</td>
</tr>
<tr>
<td>January 29</td>
<td>Tested at a hospital in Jungnang district.</td>
</tr>
<tr>
<td>January 30</td>
<td>Confirmed and hospitalized.</td>
</tr>
</tbody>
</table>

Note: The * denotes establishments whose exact names have been disclosed.
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.
Traffic declines in districts with a larger number of cases and visits.
Susceptible, Infected, Quarantined, Recovered

- Susceptible
  - Infected at workplace → Infected
- Quarantined (hospitalized)
- Recovered or dead in a hospital
- Recovered or dead without being tested → Recovered/Dead
- Tested and detected
Susceptible, Infected, Quarantined, Recovered

\[
\Delta I^a_i(t) = \beta \sum_{j \neq \text{home}} \left[ \frac{\sum_s \sum_a \pi_{sj}^a(t) I_s^a(t)}{\sum_s \sum_a \pi_{s}^a(t) N_s(t)} \times \frac{\pi^a_{ij}(t) S_i^a(t)}{\# \text{ of Susceptible from } i \text{ in } j} \right] - \gamma I^a_i(t) - d I^a_i(t)
\]

- \( \pi^a_{ij}(t) \): people of age group \( a \) living in \( i \)'s probability of working in \( j \) at time \( t \).
- \( \beta \): transmission rate.
- \( \gamma \): daily recovery rate.
- \( d_i \): daily rate at which infectious individuals are detected.
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: \( \ln d_{ij}(t) = \kappa \tau_{ij} + \delta^a \ln C_j(t) + \xi^a \ln V_j(t) + \zeta^a(t) \)
  - \( \tau_{ij} \): 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 \( \Rightarrow \) 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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value (young, old)</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Externally Calibrated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1/18</td>
<td>Daily rate at which active cases recover.</td>
</tr>
<tr>
<td>$\tau^a$</td>
<td>1/8.5, 1/10.2</td>
<td>Mean duration of hospitalization.</td>
</tr>
<tr>
<td>$\psi^a$</td>
<td>0.21%, 2.73%</td>
<td>Case fatality rate.</td>
</tr>
<tr>
<td>$\delta^a$</td>
<td>0.00209, 0.00247</td>
<td>Elasticity of commuting to local confirmed cases by age.</td>
</tr>
<tr>
<td>$\xi^a$</td>
<td>0.00138, 0.00096</td>
<td>Elasticity of commuting to local visits by infected by age.</td>
</tr>
<tr>
<td>Internally Calibrated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.1524</td>
<td>Transmission rate (target: total cases by May 31st).</td>
</tr>
<tr>
<td>$d_I$</td>
<td>0.0163</td>
<td>Daily detection rate (target: fraction of undetected infections).</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th>Full Disclosure (Korea case)</th>
<th>No Disclosure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Cases</strong></td>
<td>780,907</td>
<td>840,709</td>
</tr>
<tr>
<td><strong>Total Death</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age 20-59</td>
<td>18,743</td>
<td>20,744</td>
</tr>
<tr>
<td>age 60+</td>
<td>6,255</td>
<td>6,687</td>
</tr>
<tr>
<td>12,489</td>
<td>14,057</td>
<td></td>
</tr>
<tr>
<td><strong>Welfare Loss per day (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age 20-59</td>
<td>0.14</td>
<td>0.07</td>
</tr>
<tr>
<td>age 60+</td>
<td>0.13</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>0.16</td>
<td>0.08</td>
</tr>
</tbody>
</table>
Disclosure Policy and Lockdown: Cases and Welfare

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

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.

- Information disclosure not a panacea by itself: combined with other measures useful complement.
### Commuting Flow Equation Estimation

<table>
<thead>
<tr>
<th></th>
<th>In Commuting Flows (November 2019)</th>
<th>Δ In Commuting Flows (relative to week 1, Feb 2020)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_{ij} )</td>
<td>-0.1413 (0.0028)</td>
<td>-0.1666 (0.0034)</td>
</tr>
<tr>
<td>ln( C_j(t) )</td>
<td>-</td>
<td>-0.0087 (0.0049)</td>
</tr>
<tr>
<td>ln( C_j(t) ) × weekend</td>
<td>-</td>
<td>-0.0016 (0.0009)</td>
</tr>
<tr>
<td>ln( V_j(t) )</td>
<td>-</td>
<td>-0.0058 (0.0031)</td>
</tr>
<tr>
<td>ln( V_j(t) ) × weekend</td>
<td>-</td>
<td>-0.0010 (0.0005)</td>
</tr>
<tr>
<td>weekend</td>
<td>-</td>
<td>-0.1360 (0.0539)</td>
</tr>
</tbody>
</table>

#### Period
- Nov 2019
- Jan-May 2020

#### Age Group
- All
- Under 60
- Above 60

#### Days
- Weekdays
- Weekends

#### Fixed Effects
- Time
- Two-way

#### Cluster
- Two-way (bootstrapped)

#### Observations
- 625
- 95,000

#### R-squared
- 0.8603
- 0.8405

#### Root MSE
- 0.2375
- 0.2275
\[
\Delta \ln \pi_{ij}^a(t) = \delta^a \varepsilon^{wd} \ln C_j(t) + \delta^a (\varepsilon^{wn} - \varepsilon^{wd}) \ln C_j(t) \times \text{weekend} + \xi^a \varepsilon^{wd} \ln V_j(t) + \\
+ \delta^a (\varepsilon^{wn} - \varepsilon^{wd}) \ln V_j(t) \times \text{weekend} + \varphi^a \times \text{weekend} + \theta_i^a + \lambda_j^a + \zeta^a(t)
\]

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 \text{weekend} is an indicator variable for a day that falls on a weekend.
Susceptible, Infected, Quarantined, Recovered

\[
\begin{align*}
\Delta S_i^a(t) &= -\beta \sum_{j \neq \text{home}} \left[ \frac{\sum_s \sum_a \pi^a_{sj}(t) l_s^a(t)}{\sum_s \sum_a \pi^a_{sj}(t) N_s^a(t)} \times \pi^a_{ij}(t) S_i^a(t) \right] \\
\Delta I_i^a(t) &= \beta \sum_{j \neq \text{home}} \left[ \frac{\sum_s \sum_a \pi^a_{sj}(t) l_s^a(t)}{\sum_s \sum_a \pi^a_{sj}(t) N_s^a(t)} \times \pi^a_{ij}(t) S_i^a(t) \right] - \gamma I_i^a(t) - d_l I_i^a(t) \\
\Delta Q_i^a(t) &= d_l I_i^a(t) - \rho^a Q_i^a(t) \\
\Delta R_i^a(t) &= \gamma I_i^a(t) + \rho^a Q_i^a(t) \\
\Delta N_i^a(t) &= N_i^a(t-1) - \Delta Q_i^a(t)
\end{align*}
\]

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

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

where $z_{j,wd}^a$ 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:

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

where $z_{j,wn}^a$ denotes idiosyncratic preferences from leisure in neighborhood $j$ during the weekends.
Distance: \( \ln d_{ij}^a(t) = \kappa \tau_{ij} + \delta^a \ln C_j(t) + \xi^a \ln V_j(t) + \zeta^a(t) \)

- \( \tau_{ij} \): 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:

\[
F_{a,wd}(z_{jo}^{a,wd}) = e^{E_j^{a,wd}(z_{jo}^{a,wd}) \epsilon_{wd}}, \quad E_j^{a,wd} > 0, \epsilon_{wd} > 1
\]

\[
F_{a,wn}(z_{jo}^{a,wn}) = e^{E_j^{a,wn}(z_{jo}^{a,wn}) \epsilon_{wn}}, \quad E_j^{a,wn} > 0, \epsilon_{wn} > 1
\]
The probability that a resident of neighborhood $i$ chooses to work in $j$ during the weekday is:

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

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

$$\pi_{ij}^a(t = \text{weekend}) = \frac{E_j^{a,wn} d_{ij}^a(t) - \epsilon^{wn}}{\sum_s E_s^{a,wn} d_{is}^a(t) - \epsilon^{wn}}$$
Expected utility of an individual living in neighborhood $i$ is

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

during the weekday and

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

during the weekend where $\Gamma(\cdot)$ is a gamma function.
From the commuting probabilities, before the outbreak of the virus:

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

- $\pi_{ij}^k$: commuting probabilities from cell phone data.
- $\tau_{ij}$: 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.

- $\nu^{wd} = 0.1413$. $\nu^{wn} = 0.1666$. 
The coefficient of variation in wages within a region is:

\[
\frac{\text{Variance}}{\text{Mean}^2} = \frac{\Gamma\left(1 - \frac{2}{\varepsilon}\right)}{\Gamma\left(1 - \frac{1}{\varepsilon}\right)^2} - 1
\]

where \(\Gamma\) is a Gamma function.

- \(\varepsilon^{wd} = 4.1642\).
- \(\kappa = \nu^{wd} \times \varepsilon^{wd} = 0.0339\).
- \(\varepsilon^{wn} = \nu^{wn}/\kappa = 4.9144\)
We estimate $E_j^{a,wd}$ and $E_j^{a,wn}$ using the following conditions:

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

\[
\mathbb{E} \left[ H_{Mj}^{a,wn} - \sum_{i=1}^{S} e_{ij} \frac{E_j^{a,wn}}{e^{\nu wn \tau_{ij}}} \sum_{s=1}^{S} e_{is} \frac{E_s^{a,wn}}{e^{\nu wn \tau_{is}}} H_{Ri}^{a} \right] = 0
\]
### Sensitivity to Transmission and Detection Rates

<table>
<thead>
<tr>
<th>Scenario</th>
<th>No disclosure</th>
<th>Full disclosure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>20% lower $\beta = 0.1219$</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total # of Cases</td>
<td>81,314</td>
<td>58,384</td>
</tr>
<tr>
<td>Welfare Loss per day(%)</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>Frac. of undetected</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta = 0.1682, d_I = 0.0357$</td>
<td>No disclosure</td>
<td>Full disclosure</td>
</tr>
<tr>
<td>Total # of Cases</td>
<td>907,202</td>
<td>776,173</td>
</tr>
<tr>
<td>Welfare Loss per day(%)</td>
<td>0.08</td>
<td>0.15</td>
</tr>
<tr>
<td><strong>20% higher $\beta = 0.1829$</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total # of Cases</td>
<td>1,143,903</td>
<td>1,090,291</td>
</tr>
<tr>
<td>Welfare Loss per day(%)</td>
<td>0.05</td>
<td>0.11</td>
</tr>
<tr>
<td>Frac. of undetected</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta = 0.1515, d_I = 0.0076$</td>
<td>No disclosure</td>
<td>Full disclosure</td>
</tr>
<tr>
<td>Total # of Cases</td>
<td>565,072</td>
<td>538,609</td>
</tr>
<tr>
<td>Welfare Loss per day(%)</td>
<td>0.05</td>
<td>0.11</td>
</tr>
</tbody>
</table>