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

David Argente[†]
Pennsylvania State University

Chang-Tai Hsieh[‡]
University of Chicago

Munseob Lee[§]
University of California San Diego

July 2020

Abstract

South Korea publicly disclosed detailed location information of individuals that tested positive for COVID-19. We quantify the effect of public disclosure on the transmission of the virus and economic losses in Seoul. We use detailed foot-traffic data from South Korea's largest mobile phone company to document the change in the flows of people across neighborhoods in Seoul in response to information on the locations of positive cases. We analyze the effect of the change in commuting flows in a SIR meta-population model where the virus spreads due to these flows. We endogenize these flows in a model of urban neighborhoods where individuals commute across neighborhoods to access jobs and leisure opportunities. Relative to a scenario where no information is disclosed, the change in commuting patterns due to public disclosure lowers the number of cases by 200 thousand and the number of deaths by 6.5 thousand in Seoul over two years. Compared to a city-wide lock-down that results in the same number of cases over two years, the economic cost is 40% lower with full disclosure.

*We want to thank Fernando Alvarez, Jingting Fan, David Lagakos, Marc-Andreas Muendler, Valerie A. Ramey, and Nick Tsivanidis for their comments and suggestions and seminar participants at UC San Diego, FGV, and Insper. Results in this article are calculated based on proprietary data provided by the SK Telecom Co., Ltd. We thank Geovision at the SK Telecom for their assistance with the data.

[†]Email: dargente@psu.edu. Address: 403 Kern Building, University Park, PA 16801.

[‡]Email: chsieh@chicagobooth.edu. Address: 5807 South Woodlawn Ave., Chicago, IL 60637.

[§]Email: munseoblee@ucsd.edu. Address: 9500 Gilman Drive #0519, La Jolla, CA 92093-0519.

1 Introduction

In the current battle against COVID-19, South Korea stands out. The number of COVID-19 cases and deaths in South Korea is among the lowest in the world. South Korea also never implemented the strict social isolation policies followed by China, Europe, and many US states. Instead of locking down its population, South Korea chose to combat the pandemic by testing, tracing the contacts of infected individuals, and quarantining. The limitation of this approach is that, even when done on a massive scale as Korea did, it is likely that many infected individuals, particularly those that are asymptomatic, are missed and can continue to spread the disease.

However, a key innovation in South Korea is that in addition to the yeoman's work of testing and contact tracing, it also publicly disclosed detailed information on the individuals that had tested positive for COVID-19. South Koreans received text messages whenever new cases were discovered in their neighborhood, as well as information and timelines of infected people's travel. If the location of infected individuals that have not yet been detected is correlated with the location of people who tested positive, then disclosure can help people "target" their social distancing. Instead of reducing social activity across the board, individuals can avoid neighborhoods where the probability of becoming infected is high. The endogenous reduction in social interaction in high risk neighborhoods can help reduce the transmission rate. At the same time, individuals can continue to travel to places where the transmission risk is low, so social interactions in low risk neighborhoods do not have to be affected.

Second, compared to a lock-down strategy, South Korea's strategy allows individuals to self select into social distancing. The selection can be based on a person's health risk from exposure or from their economic losses from social isolation. Individuals with a high health risk from commuting to a neighborhood with many detected cases can change their commuting pattern, while individuals with low health risk can make a different choice. Individuals that can easily substitute between working in the office and working at home can do that, while others where the substitution is costly can continue to commute to work. In contrast, a lockdown applies to all individuals and does not discriminate between individuals with different cost/benefit ratios for social isolation.

In this paper, we measure the welfare effect of the disclosure of public information of COVID-19 cases in Seoul, where one fifth of South Korea's population lives. We combine detailed foot-traffic data in Seoul from South Korea's largest mobile phone company with publicly disclosed information on the location of individuals who had tested positive. The data indicates that density of foot-traffic data declines in neighborhoods with a larger number

of infected patients. This pattern is particularly pronounced during the weekends and among those over the age of 60.

We then filter this data through the lenses of SIR model augmented in two ways. First, we include several sub-populations (neighborhoods) where the epidemic transmits across neighborhoods due to the flow of people across these neighborhoods. Second, we assume that the flow of people across neighborhoods is the outcome of individuals maximizing their utility by choosing a neighborhood to work and to enjoy their leisure time. The flow of people across neighborhoods generates economic gains from the optimal match of people with the place of work and leisure, but it is precisely this flow that transmit the virus across neighborhoods. We use the model to simulate the effect of the change in commuting patterns in response to the disclosure of public information on two key outcomes. First, we simulate how the change in commuting patterns lowers the transmission of the virus across neighborhoods. Second, we measure the economic losses from the fact that some individuals have changed their place of work or leisure. Our estimate is that over the next two years, South Korea's current strategy will lead to a cumulative 602 thousand cases in Seoul, 12.5 thousand deaths, and welfare losses that average 0.6 percent.

We then consider three alternative scenarios. The first scenario is no disclosure of information, which we model as a scenario in which there is no change in commuting patterns. We estimate that, compared to the case of complete disclosure, non-disclosure results in 200 thousand more infected people and 6.5 thousand additional deaths in Seoul (over the next two years). The welfare losses (not including the value of lost lives), on the other hand, are lower compared to full disclosure. Over the next two years, the loss in economic welfare are 0.32 percentage points lower for those under 60 and 0.49 percentage points lower for those older than 60.

A second scenario is a lockdown, which we model as a uniform reduction in the movement of the population across districts in Seoul and in which a certain fraction of the population is forced to stay at home. To ease the comparison, we calibrate the length and intensity of the lockdown in Seoul so that it results in the total number of cases under the full disclosure of information case. The lockdown would have to last for approximately 100 days and approximately 34% of the population would have to stay at home in order to have the same number of confirmed cases as in the full disclosure case. The welfare losses for those under 60 years of age, on the other hand, are almost twice as large as the full disclosure case. This is because the "optimal" commuting patterns are severely disrupted for a large number of people in the lockdown.

The last scenario is a partial shutdown. We simulate a partial shutdown as a policy that stops people from commuting to the first four districts that had a confirmed case in

Seoul and that prohibits the residents of those districts to commute outside them. We show that shutdowns are very costly policies in terms of welfare and not very effective at delaying the spread of the disease. Seoul is a city that is very inter-connected with vast amount of commuting across districts. Even if the shutdown is implemented very rapidly after the first few cases are detected, undetected cases commute across districts; this interaction, albeit short-lasting, is enough to spread the disease in the entire city over time.

We use a standard SIR epidemiology model where movements of the population transmit the virus across space. Such models have been used in epidemiology to analyze disease outbreaks and to forecast the geographical spread of epidemics.[1] We extend the epidemic meta-population models in two ways. First, in addition to residential districts, we allow for an additional partition and model the SIR dynamics for individuals of different ages as in Acemoglu et al. (2020). This allows us to consider not only different SIR parameters and contact rates across groups but also to study their differential response to policies, even those not specifically targeted towards particular groups. Second, we allow the contact rates across groups to be endogenous outcomes of a commuting choice model. The quantitative model of internal city structure draws heavily from Ahlfeldt et al. (2015), Monte et al. (2018), and Tsivanidis (2019). The model includes stochastic shocks to commuting decisions, which yield a gravity equation for commuting flows, which in addition to distance includes the disclosed information of the recent confirmed cases of the disease as well as the visits of infected people to different districts.

The rest of the paper is organized as follows. We first describe the background of the pandemic in South Korea, particularly in Seoul. The next section describes the foot traffic data we use to estimate the model. We then describe the commuting patterns of Seoul during the pandemic and in response to the disclosed cases. Section 5 presents the spatial SIR model that allows us to consider the geographic spread of the disease. Section 6 presents the spatial model with commuting choices and Section 7 its calibration. In Section 8 we perform several counter-factual exercises such as comparing our benchmark economy with one with no disclosure and one with a lockdown.

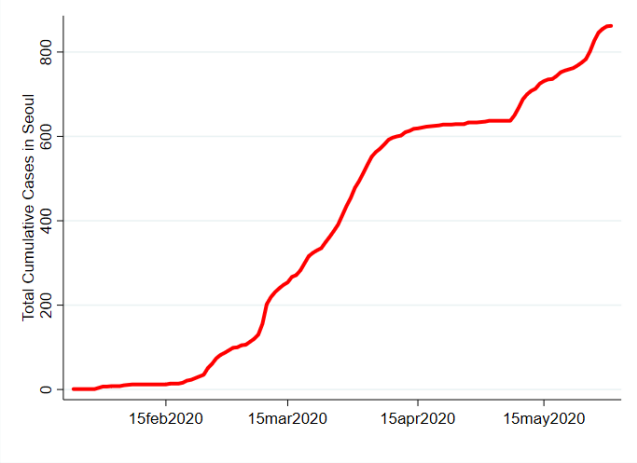
2 Background

South Korea was once considered one of the epicenters of the COVID-19 outbreak. However, the country has been remarkably successful in containing the pandemic. The first confirmed case of COVID-19 in South Korea was detected on January 19, 2020. The first case in Seoul

[1] See, for example, Danon et al. (2009), Keeling et al. (2010), Brockmann and Helbing (2013), Hufnagel et al. (2004) and Wesolowski et al. (2012).

was found on January 24, 2020. Figure 1 shows the total number of cases detecte in Seoul. The number of cases increased rapidly between the end of February and the beginning of April, they stagnated until mid May before increasing again at the end of May.

Figure 1: Total Cumulative Cases in Seoul



Notes: The figure shows the cumulative total cases in Seoul from January to May 2020.

As of May 31 of 2020, South Korea had 11,468 confirmed cases of COVID-19 and 270 deaths. The city of Seoul, a city of almost 10 million inhabitants, had only 861 confirmed cases and 4 deaths. These numbers are remarkably low in comparison to cities of similar size. Importantly, Seoul is one of the most densely populated cities in the world. The strategy of the country has been based on massive testing and extensive contact tracing. In addition, local governments disclose in real time (via text messages) information on individuals that had tested positive for COVID-19. The Seoul Metropolitan Government also developed a dedicated website and a mobile app to enable residents to access real-time information. The mobile app also allowed individuals to report on their symptoms.

When a person tests positive, their district sends out an alert to people living nearby about the movements of the infected individual before being diagnosed. A typical alert can contain the infected person's age and gender, and a detailed log of their movements based on contact tracing combined with data from cell phone and credit card records. The alerts also included detailed statistics on the status of COVID-19 cases (e.g. confirmed, suspected, being tested, self-quarantined, under surveillance). In some districts, public information includes which rooms of a building the person was in and whether or not they wore a mask. The publicly available database also provides information on "clean zones" - places that have been disinfected following visits by confirmed patients. Table 1 shows an example of public

disclosure. This is the second case in Seoul confirmed on January 30. This information has been available on official website and has been circulated by text alerts for people living in certain districts.

Table 1: Example of Public Disclosure

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.

Notes: The table shows example of public disclosure for the second case in Seoul confirmed on January 30. The * denotes establishments whose exact names have been disclosed.

3 Data

3.1 Mobile Phone Data

We obtain two proprietary data sets from the Korean telecommunication company SK Telecom which has the largest market share in Korean mobile phone market.[2] The first data set is the hourly foot traffic data covering 25 districts in Seoul and the second is the daily bilateral flows across the 25 districts in Seoul.

The hourly foot traffic data reports the total number of subscribers present in Seoul's 25 districts each hour. The presence of subscribers in a given districts is based on any of their cellular phone activity including calls, texts, and the internet connection. The location of a

[2]In January 2020, its market share is 41.9%, followed by KT with 26.4% and LG U+ with 20.6% (source: Korea Communications Commission).

user is inferred by cellular tower triangulation. The data splits users by the gender and by age group. We use the data from January 2020 to March 2020 for empirical documentation. Table 2 shows the structure of the data. The table shows that on Sunday January 5, 2020, there were 42,570 females between 20 and 29 years of age in Gangnam district at 3 am. The number increased to 63,330 at 3 pm, due to people visiting Gangnam district during the day.

Table 2: Sample Observations - Foot Traffic Data (Seoul)

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

Notes: The table shows sample observations in the foot traffic data. The data reports the number of people in each district every hour for six age groups (20, 30, 40, 50, 60, 70) and for each gender.

The daily bilateral flow data reports the total number of people who moved from a district to another district within Seoul. To be included in the data, a person should stay at origin district for more than two hours, and then stay at different destination district for more than two hours. If a person moves multiple times during the day, the data records only the main movement. Similar to the the hourly foot traffic data, the data set contains information by gender and by age group. We use the data from November 2019 to May 2020 to document empirical patterns and calibrate parameters in the model.

3.2 Seoul Survey

We also use the micro-data of the 2018 Seoul Survey. The survey is conducted on a yearly basis and contains commuting patterns as well as labor information of Seoul's residents. The survey targets 20,000 households within the city of Seoul. All household members 15 years or older in the targeted households participate. The total number of survey participants every year is approximately around 40,000 and 50,000. In order for the sample to be representative, the survey follows a spatially stratified sampling strategy considering the household size distribution.

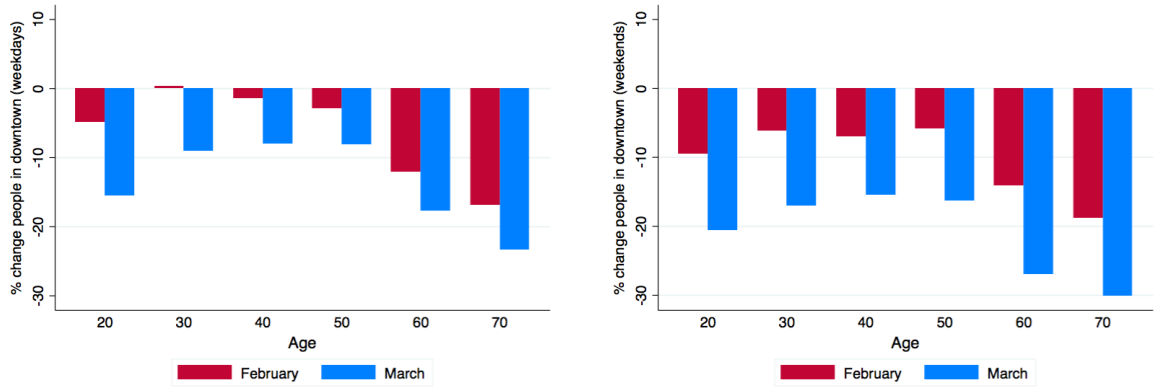
We obtain the wage distribution for different demographic groups across districts. Importantly, the survey contains information on the fraction of people that do not commute during weekdays and also the wages that they receive, which allows us to infer the size of the home sector before the outbreak of the virus.

4 Empirical Patterns

Figure 2 shows the percent change in the average number of people in downtown (at 3pm) in February and March relative to the month of January.[3] Recall that the first case in Korea was identified on January 20th and that there were only seven confirmed cases by the end of the same month. The figure shows that there were significant changes in the commuting patterns of people before and after the virus outbreak. Not surprisingly, the population that is less than 20 years old decreased their commuting to downtown given that school activities were suspended. More remarkably, the figure shows that the individuals older than 60 years old, those that are more vulnerable to the virus, significantly decrease their commuting to downtown. During weekends, individuals of all ages commuted less to downtown during the month of February and March; the most vulnerable groups decreased their commuting more substantially.

[3]Downtown Seoul includes the following districts: Jung-gu, Jongno-gu, Gangnam-gu, Seocho-gu, and Yeongdeungpo-gu.

Figure 2: Percent Change People in Downtown

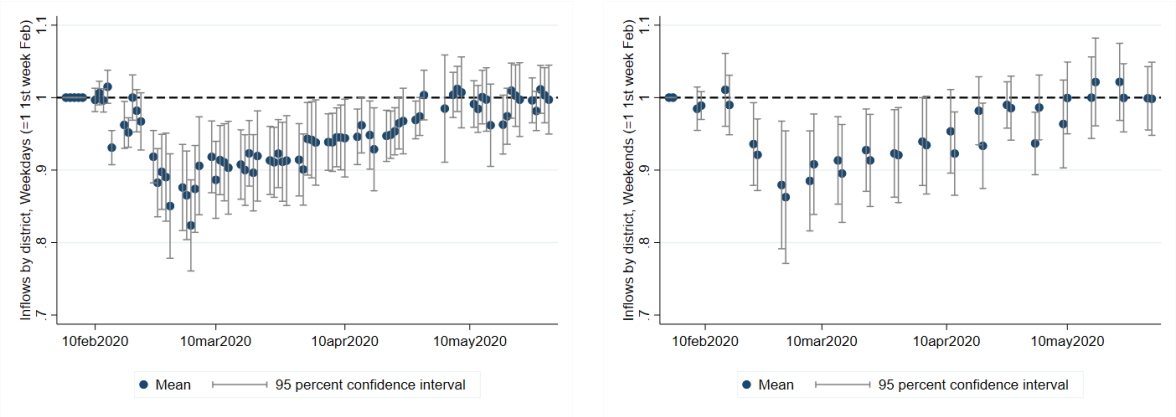


(a) Weekdays (b) Weekends

Notes: The figure shows the change in the average number of people in downtown in February and March relative to the month of January. The total number of people is approximated by the total people in the downtown districts at 3pm. Panel (a) shows the percent change during the weekdays and Panel (b) show the changes during the weekends. The downtown districts are: Jung-gu, Jongno-gu, Gangnam-gu, Seocho-gu, and Yeongdeungpo-gu.

Figure 3 shows the change in the inflow of people to the districts during weekdays and weekends. It shows mean of changes in inflows in 25 districts compared to the first week of February, along with 95 percent confidence interval. On average, inflows decreased to around 90 percent in March and April compared to the first week of February. During weekends, the variance of the changes in inflows is larger, suggesting that people reallocated their commuting destinations more compared to weekdays. In May, the average inflow returned to the levels seen before the outbreak of the virus.

Figure 3: Change in Inflows by Districts



(a) Weekdays (b) Weekends

Notes: The figure shows the change in the inflow of people at the district level. Panel (a) shows the change in the inflows during the weekdays. It shows the mean of the changes in inflows in the 25 districts compared to the first week of February, along with the 95 percent confidence interval. Panel (b) show the change in the inflows during the weekends.

We explore the response of people to the overall spread of disease and to the availability of information on the total number of confirmed cases and visits in each district. First, we quantify change in commuting flows with respect to the overall spread of disease implementing the following specification:

$$\ln\pi_{ijt} = \alpha + \iota \ln C_t + \nu\tau_{ij} + \gamma_i + \theta_j + e_{ijt} \tag{1}$$

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

Table 3 shows the results. During weekdays, the commuting probability of people under 60 deceases by 0.025 percent when the total cases increase by 1 percent. The response is larger for people above 60 both during weekdays and weekends. This change in commuting flow is not related to the detailed information disclosed by the local authorities of Seoul since the daily total number of cases is also publicly known in most countries.

Table 3: Commuting Response to Overall Spread of Disease

	(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

Notes: The table shows the results of estimating equation (1). The dependent variable is the bilateral commuting probability from district i to district j on day t. The independent variable is the total number of cases in Seoul on day t. Column (1) includes weekdays for the young (under 60). Column (2) includes weekdays for the old (above 60). Column (3) includes weekends for the young (under 60). Column (4) includes weekends for the old (above 60).

Thus, in order to quantify the change in commuting flow that results for the disclosure of the total number of cases and visits at the local level, we use the following specification:

$$\ln \pi_{ijt} = \alpha + \varphi_1 \ln C_{jt} + \varphi_2 \ln V_{jt} + \nu \tau_{ij} + \gamma_i + \theta_j + e_{ijt} \tag{2}$$

where $C_{j,t}$ is the number of cases active in district j and $V_{j,t}$ is the number of visits by the detected cases in district j. Public disclosure is captured by the total number of confirmed cases in each district. We consider only the cases that are still active; this is, infected people that have not yet recovered and that acquired the virus less than 14 from day t. Similarly, we only include the total visits of the last 14 days to capture only the visits of the active cases, which captures the additional information provided by the travel logs of infected individuals.

The results are shown in Table 4. Column (1) shows the results for people under 60 during weekdays. The commuting probability to a district decreases by 0.018 percent when the total number of cases in that district increase by 1 percent and by 0.021 percent when total visits in that district increase by 1 percent. Importantly, Column (2) shows that people above 60 years of age respond more drastically to the information available on the total number of cases and visits to their district. These patterns are more noticeable during weekends as shown in Columns (3)-(4). Given the differences in the commuting response across days of the week as well as across people of different age groups, we allow for heterogeneity along these dimensions in the model developed in Section 6.

Table 4: Commuting Response to Public Disclosure

	(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

Notes: The table shows the results of estimating equation (2). The dependent variable is the total number of people present at a given district j on day t. The independent variables are the cumulative number of confirmed cases on a given day in district j (net of those that have recovered) and total number of visits to district j by infected people. Columns (1)-(2) include only weekdays and Columns (3)-(4) only weekends. Columns (1)-(3) include only people under the age of 60 and Columns (2)-(4) only those above 60 years of age.

5 Modeling Spread of Disease

The typical approach in the epidemiology literature is to study the dynamics of the pandemic, for infected and recovered, as functions of some exogenously chosen diffusion parameters β, which are in turn related to various policies (e.g. the partial lockdown of schools, universities, offices, and other measures of diffusion mitigation), and where the diffusion parameters are stratified by age and other individual covariates. In the model, individuals are classified into three types: susceptible (S; at risk of contracting the disease), infectious (I; capable of transmitting the disease), and recovered (R; those who recover or die from the disease). In this model, the reproduction number R, which measures the transmissibility of a virus, represents the average number of new infections generated by each person is defined as $R = \frac{\beta}{\gamma}$. The basic SIR model focuses on the epidemic propagation in a single population.

We consider a simple SIR meta-population model that allows for the spatial dissemination of the epidemic across regions and takes into account the movements of the population from one district to another (Keeling et al., 2010; Keeling and Rohani, 2011). The dynamics are discrete and deterministic. We consider individuals of different ages, young and old, and residing in different locations. No additional substructure of the population is considered

(e.g. schools or workplaces), as our aim is to introduce a rather simple epidemic model to test the adequacy of different commuting sources for the simulation of dissemination within a city. Subpopulations are coupled by directed weighted links representing the commuting fluxes between two locations, thus defining the meta-population structure of the model. No other type of movement is considered.

Mobility is described in terms of recurrent daily movements between place of residence and workplace so that the infection dynamics can be separated into two components, each of them occurring at each location. π_{ij}^a is the probability of people whose age is a and who reside in i of choosing working location j. The commuting probabilities, π_{ij}^a, determine the contact rate between groups and will be endogenous outcomes of the commuting model developed in the next section. The epidemic dynamics are described by the following equations:

$$
\begin{aligned}
\Delta S_i^a(t) &= -\beta \sum_j \left[\frac{\sum_{a,s} \pi_{sj}^a(t) I_s^a(t)}{\sum_{a,s} \pi_{sj}^a(t) N_s^a(t)} \times \pi_{ij}^a(t) S_i^a(t) \right] \\
\Delta I_i^a(t) &= \beta \sum_j \left[\frac{\sum_{a,s} \pi_{sj}^a(t) I_s^a(t)}{\sum_{a,s} \pi_{sj}^a(t) N_s(t)} \times \pi_{ij}^a(t) S_i^a(t) \right] - \gamma I_i^a(t) - d_I I_i^a(t) \\
\Delta Q_i^a(t) &= d_I I_i^a(t) - \tau^a Q_i^a(t) \\
\Delta R_i^a(t) &= \gamma I_i^a(t) + \tau^a Q_i^a(t)
\end{aligned} \tag{3}
$$

As in Acemoglu et al. (2020), individuals are partitioned into groups; we partitioned individuals based on their age and their district of residence. Susceptible individuals of age a residing in district i may become infected by coming into contact with infected individuals at their working location j. The first equation describes the change in the number of susceptible individuals residing in district i, after taking into account their commuting patterns determined by the age-specific commuting probabilities π_{ij}^a. Given the empirical evidence presented in the previous section, below we allow the commuting probabilities to also differ across weekends and weekdays.

Notice that the equation aggregates across all districts j to which susceptible individuals of age a residing in district i commute. The "matching technology" is the same as in the basic SIR model, which in turn is similar to the quadratic matching technology in Diamond and Maskin (1979, 1981), where the number of matches between two groups is the product of the size of the two groups. The matching technology can be made more general, but we believe this is a good starting point given the geographic context of our model. The parameter β is the number of susceptible agents per unit of time to whom an infected agent can transmit the virus after contact. All susceptible agents that get the virus become infected. A fraction

of infected individuals recover or die at rate γ. We assume that recovered agents that did not die are immune thereafter. We assume non-zero fatality rate, $\psi > 0$.

Infected individuals can also be detected at rate d_I and removed to quarantine. In this case, the average time they spend in isolation is $1/\tau^a$ and after that they either recover or die. We allow τ^a to be age-specific since older individuals are more likely to require critical care.

6 Spatial Model

We develop a simple model to guide our empirical analysis. The model includes commuting choice decisions as in Ahlfeldt et al. (2015). Each district i in Seoul is populated by an exogenous measure of H^a_{Ri} workers of age a. For simplicity, we abstract from residential location choice. A worker decides where to work after observing the realization for her idiosyncratic utility, and picks the district to commute during weekdays and weekends within the city that maximizes her utility. During weekdays, workers produce a final traded good and consume it. The price of the final traded good is chosen as the numeraire ($p = 1$). During weekends, people make commuting decisions for leisure. Districts differ in their average productivity (during weekdays) or amenities (during weekends), the number of infected individuals that reside there, and their access to the transport network, which determines travel times between any two districts in the city.

6.1 Workers

Workers are risk neutral and have preferences that are linear in a consumption index: $U^{a,k}_{ijo} = C^a_{ijo}$, where $C^{a,k}_{ijo}$ denotes the consumption index for worker o of age a residing in district i and commuting to district j. Weeks are divided into weekdays ($k = wd$) and weekends ($k = wn$). This consumption index depends on the consumption of the traded good ($c^{a,k}_{ijo}$); the disutility from commuting from residence district i to workplace district j ($d^a_{ij} \geq 1$); and an idiosyncratic shock that is specific to individual workers and varies with the worker's district of employment ($z^{a,k}_{jo}$). This idiosyncratic shock captures the idea that individual workers can have idiosyncratic reasons for commuting to different parts of the city. Source of heterogeneity is productivity during weekdays and preference during weekends.

The aggregate consumption index is simply:

$$C^{a,wd}_{ijo} = \frac{c^{a,wd}_{ijo}}{d^a_{ij}}, \quad C^{a,wn}_{ijo} = \frac{z^{a,wn}_{jo}}{d^a_{ij}} \tag{4}$$

where the iceberg commuting cost $d_{ij}^a = e^{\kappa\tau_{ij}+\delta^a \ln C_j+\xi^a \ln V_j+\alpha^a \ln C}$ increases with the travel time (τ_{ij}) between district i and j. Travel time is measured in distance. The commuting costs, $d_{ij}^a \in [1,\infty)$, also increase with the number of detected cases C_j that are active in the district where people commute, with the number of visits by detected cases V_j, and with the total number of detected cases in Seoul C. The parameter κ control the impact of distance on the commuting costs. The parameters δ^a, ξ^a, and α^a are the sensitivity of the commuting costs to the disclosed information on the detected cases and visits. Given our empirical evidence, all parameters are allowed to vary by age.

We model the heterogeneity in the utility that workers derive from commuting to different parts of the city following McFadden (1974) and Eaton and Kortum (2002). For each worker o of age a living in district i and commuting to district j, the idiosyncratic component of utility ($z_{ijo}^{a,k}$) is drawn from an independent Fréchet distribution:

$$F^{a,k}(z_{jo}^{a,k}) = e^{E_j^{a,k}(z_{jo}^{a,k})^{\epsilon^k}}, \quad E_j^{a,k} > 0, \epsilon^k > 1 \tag{5}$$

where the scale parameter $E_j^{a,k}$ determines the average productivity (during weekdays) or utility (during weekends) derived from commuting to district j; and the shape parameter ϵ^k controls the dispersion of idiosyncratic utility.

The indirect utility from residing in district i and commuting to district j during weekdays can be expressed in terms of the wage per efficiency unit paid at this workplace (w_j^a), commuting costs (d_{ij}^a) and the idiosyncratic productivity ($z_{jo}^{a,wd}$):

$$u_{ijo}^{a,wd} = \frac{w_j^a z_{jo}^{a,wd}}{d_{ij}^a} \tag{6}$$

where we have used utility maximization and price for traded goods is normalized to 1 as numeraire.[4]

During weekends, the indirect utility from residing in district i and commuting to district j can be expressed in terms of commuting costs (d_{ij}^a) and the idiosyncratic utility ($z_{jo}^{a,wn}$):

$$u_{ijo}^{a,wn} = \frac{z_{jo}^{a,wk}}{d_{ij}^a} \tag{7}$$

Since indirect utility is a monotonic function of the idiosyncratic shock ($z_{jo}^{a,k}$), which has a Fréchet distribution, the indirect utility for workers living in district i and commuting to district j also has a Fréchet distribution. Each worker chooses the commute that offers her

[4]Commuting costs are proportional to wages to capture changes over time in the opportunity cost of travel time.

the maximum utility, where the maximum of Fréchet distributed random variables is itself Fréchet distributed. Workers can choose to stay at the home sector. The home sector has its own $E^{a,k}_{\text{home}}$ and w^a_{home}. The scale parameters $E^{a,wd}_{\text{home}}$ and $E^{a,wn}_{\text{home}}$ can be interpreted as average productivity at home and average amenity at home, respectively. We assume that $d^a_{i,\text{home}} = 1$ for all living districts and all age groups; this is, no commuting cost is incurred either by physical distance or by risk of being infected if a worker stays at home.

Weekdays: After observing her realizations for idiosyncratic utility for the employment districts, each worker chooses where to work to maximize her utility, taking as given factor prices and the location decisions of other workers. Therefore, workers, given their residence, sort across employment districts depending on their idiosyncratic productivity and the characteristics of these locations. Other things equal, workers are more likely to work in district j, the higher its wage w^a_j, the higher its average idiosyncratic productivity as determined by $E^{a,wd}_j$, and the lower its commuting costs d^a_{ij} from residential locations. Conditional on living in district i, the probability that a worker commutes for work to district j is

$$\pi^{\text{a,wd}}_{ij|i} = \frac{E^{a,wd}_j (w^a_j / d^a_{ij})^{\epsilon^{wd}}}{\sum_s E^{a,wd}_s (w^a_s / d^a_{is})^{\epsilon^{wd}}} \tag{8}$$

The commuting market clearing condition that equates the measure of workers employed in district j ($H^{a,wd}_{Mj}$) with the measure of workers choosing to commute to district j:

$$H^{\text{a,wd}}_{Mj} = \sum_i \pi^{\text{a,wd}}_{ij|i} H^a_{Ri} = \sum_i \frac{E^{a,wd}_j (w^a_j / d^a_{ij})^{\epsilon^{wd}}}{\sum_s E^{a,wd}_s (w^a_s / d^a_{is})^{\epsilon^{wd}}} H^a_{Ri} \tag{9}$$

Due to the disease, certain fraction of people are quarantined, and they cannot work. We assume that all others, susceptible, infected, and recovered people, can work. Residence employment (H^a_{Ri}) is defined as:

$$H^a_{Ri} \equiv N^a_i - Q^a_i - D^a_i. \tag{10}$$

where N^a_i is population of age a, Q^a_i is the number of quarantined people and D^a_i is total death in district i.

Expected worker income conditional on living in district i is equal to the wages in all possible employment locations weighted by the probabilities of commuting to those locations

conditional on living in i:

$$\mathbb{E}[w_j^a|i] = \sum_i \pi_{ij|i}^{\text{a,wd}} w_j^a$$

Weekends: During the weekends, the destination of each worker is not tied to their work (wages) and only tied to their idiosyncratic preference for each location. Thus, after observing her realizations for idiosyncratic utility, each worker chooses where to visit to maximize her utility, taking as given the location decisions of other workers. Other things equal, workers are more likely to visit district j, the higher its average idiosyncratic utility as determined by E_j^a, and the lower its commuting costs d_{ij}^a from residential locations. Conditional on living in district i, the probability that a worker visits district j on the weekend is

$$\pi_{ij|i}^{\text{a,wn}} = \frac{E_j^{a,wn} \left(d_{ij}^a\right)^{-\epsilon^{wn}}}{\sum_s E_s^{a,wn} \left(d_{ij}^a\right)^{-\epsilon^{wn}}} \tag{11}$$

The commuting market clearing condition that equates the measure of workers visiting district j (H_{Mj}^a) with the measure of workers choosing to visit to district j:

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

6.2 Production

Production of the final good occurs under conditions of perfect competition. We assume that the production technology for the traded good is age-specific

$$Y_j^a = A_j^a H_{Mj}^{\text{a,wd}} \, \mathbb{E}[z_j^{a,wd}|\text{People choose j}] \tag{12}$$

As a result, wage per efficiency unit is the same as total factor productivity of the district; $w_j^a = A_j^a$. The aggregate production in the economy is defined as:

$$Y \equiv \sum_{a,j} Y_j^a = \sum_{a,j} A_j^a H_{Mj}^{\text{a,wd}} \, \mathbb{E}[z_j^{a,wd}|\text{People choose j}] \tag{13}$$

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

6.3 Equilibrium

In this section, we characterize the properties of the model. Given the model's parameters $\beta, \boldsymbol{\epsilon}, \boldsymbol{\kappa}, \boldsymbol{\delta}, \boldsymbol{\xi}, \boldsymbol{\alpha}$ and vectors of exogenous location characteristics $\boldsymbol{A}, \boldsymbol{E}, \boldsymbol{\tau}$, the general equilibrium of the model is referenced by the vectors $\boldsymbol{w}, \boldsymbol{\pi}$ and total city population in each city H_{Ri}. These components of the equilibrium vector are determined by the following system of three equations: the commuting choice probability, profit maximization, and zero profits.

Assuming exogenous, finite, and strictly positive location characteristics, $\boldsymbol{E}$, $\tau_{ij} \in (0, \infty) \times (0, \infty)$, and exogenous, finite, and non-negative final goods productivity, $\boldsymbol{A}$, there exists a unique general equilibrium vector $\boldsymbol{w}, \boldsymbol{\pi}$.

There is also a unique mapping from the observed variables to unobserved location characteristics. These unobserved location characteristics include production and residential fundamentals and several other unobserved variables. Let the transformed wages for each employment location be denoted as $\omega_j^{a,wd} = \left(w_j^a\right)^{\epsilon^{wd}} E_j^{a,wd}$ and $\omega_j^{a,wn} = E_j^{a,wn}$, which capture the wage and the Fréchet scale parameter for that location because these both affect the relative attractiveness of a commuting location to workers.

Given known values for the parameters $\beta, \boldsymbol{\epsilon}, \boldsymbol{\kappa}, \boldsymbol{\delta}, \boldsymbol{\xi}, \boldsymbol{\alpha}$ and the observed data $\boldsymbol{w}, \boldsymbol{H_M}, \boldsymbol{\tau}$, there exist unique vectors of the unobserved location characteristics $\boldsymbol{E}$ that are consistent with the data being an equilibrium of the model. The economics underlying this identification result are as follows. Given our assumptions on the production side, the data of observed wages pin down the vector of productivity. Given observed workplace, and our measures of travel times, worker commuting probabilities can be used to solve for unique transformed wages consistent with commuting market clearing.

7 Calibration and Simulation of the Model

7.1 Gravity

From the commuting probabilities, one of the model's key predictions is a semi-log gravity equation for commuting flows from residence i to commuting destination j. Before the outbreak of the virus the gravity equation can be written as:

$$ln\pi_{ij}^k = -\nu^k \tau_{ij} + \theta_i + \theta_j + e_{ij}^k \tag{14}$$

where θ_i and θ_j are residence and commuting destination fixed effects. Commuting costs are $d_{ij}^a = e^{\kappa \tau_{ij}}$ for all age groups and travel times τ_{ij} are measured in kilometers. The parameter $\nu^k = \epsilon^k \kappa$ is the semi-elasticity of commuting flows with respect to travel distance and is a

combination of the commuting cost parameter κ and the commuting heterogeneity parameter ϵ^k. We augment the gravity equation with a stochastic error that captures measurement error in travel distances e_{ij}^k.

We estimate equation (14) using a linear fixed effects estimator, with the bilateral commuting flow across 25 districts in Seoul in November 2019. We find that the semi-elasticity of commuting with respect to travel time is 0.1413 during weekdays and 0.1666 during weekends, which are statistically significant at the one percent level. These estimates implies that each additional kilometer of travel reduces the flow of commuters by around 14.1 percent during weekdays and 16.6 percent during weekends. From the regression $\mathbb{R}^2$, this gravity equation specification explains around 84 percent of the variation in bilateral commuting patterns. Taken together, these results suggest that the gravity equation predicted by the model before the outbreak of the virus provides a good approximation to observed commuting behavior.

By the properties of Frechet distribution (as in Hsieh et al. (2019) and Lee (2016)), the coefficient of variation in wages within a region is a function of the shape parameter during weekdays, ϵ^{wd}:

$$\frac{\text{Variance}}{\text{Mean}^2} = \frac{\Gamma(1-\frac{2}{\epsilon^{wd}})}{\Gamma(1-\frac{1}{\epsilon^{wd}})^2} - 1 \tag{15}$$

where Γ is a Gamma function. We calculate the coefficient of variation in wages within a region from the Seoul Survey in 2018. Estimated ϵ^{wd} is 4.1642.

From the semi-elasticity of commuting flows, $\nu^k = \epsilon^k \kappa$, we estimate $\kappa = \nu^{wd} \times \epsilon^{wd} = 0.0339$. Then, the shape parameter during weekends, $\epsilon^{wn} = \nu^{wn}/\kappa$, is estimated to be 4.9144, which is higher than the shape parameter during weekdays. Therefore, it predicts that variance of productivity across districts during weekdays is larger than variance of utility across districts during weekends. This is consistent with the empirical findings that people changed their commuting behavior more during weekends. Furthermore, these values of ϵ for commuting decisions are broadly in line with the range of estimates for the Fréchet shape parameter for international trade flows (the range of estimates in Eaton and Kortum (2002) is from 3.6 to 12.8) and for commuting flows (estimates in Ahlfeldt et al. (2015), Monte et al. (2018), Tsivanidis (2019) are 6.8, 3.3 and 2.7-3.2, respectively).

As in Ahlfeldt et al. (2015), we define adjusted wages as $\tilde{w}_j^{a,wd} = w_j^a (E_j^{a,wd})^{1/\epsilon^{wd}}$ and $\tilde{w}_j^{a,wn} = (E_j^{a,wn})^{1/\epsilon^{wn}}$, which capture the wage and the Fréchet scale parameter for that location since both affect the relative attractiveness of a commuting location to workers. For simplicity, we introduce two age groups: the young (age 20-59) and the old (age 60+). Using our estimate for ν^{wd} and ν^{wn}, the model's labor market clearing condition can be solved for a transformation of wages $\omega_j^{a,wd} = (\tilde{w}_j{}^{a,wd})^{\epsilon^{wd}} = E_j^{a,wd}(w_j^a)^{\epsilon^{wd}}$ and $\omega_j^{a,wn} = (\tilde{w}_j{}^{a,wn})^{\epsilon^{wn}} = E_j^{a,wn}$

in each location using observed workplace employment (H^a_{Mj}), residence employment (H^a_{Ri}), and bilateral travel times (τ_{ij}):

$$H^{a,k}_{Mj} = \sum_{i=1}^{S} \frac{\omega_j^{a,k} / e^{\nu^k \tau_{ij}}}{\sum_{s=1}^{S} \omega_s^{a,k} / e^{\nu^k \tau_{ij}}} H^a_{Ri} \tag{16}$$

where weeks are divided into weekdays (wd) and weekends (wn). As a normalization, we assume that geometric average of the location parameter $E^{a,k}$ is one.

We separately calculate the share of home sector during weekdays and weekends. Using the Seoul Survey in 2018, we find that 17% of the young (age 20-59) and 61% of the old (age 60+) are not commuting but earning positive income. We use those numbers for the home sector shares during weekdays. In order to infer the home sector shares during weekends, we use the Time Use Survey and calculate the share of people staying at home during weekends. We assume that people who spends less than 5 hours on outside activities are at the home sector; the inferred shares are 46 percent for the young (age 20-59) and 67 percent for the old (age 60+).

7.2 Calibration of COVID-19-specific Parameters

There is considerable uncertainty about infection, recovery, and mortality rates of the virus. For this reason, we calibrate γ and τ borrowing the most common estimates used in the recent literature and confirming they are consistent with the data from the World Health Organization (WHO) compiled by the Johns Hopkins University Center for Systems Science and Engineering (JHU CCSE). The parameter γ governing the rate (per day) at which infected people either recover or die is considered a fixed parameter of the disease and is set to γ=1/18 reflecting an estimated duration of illness of 18 days as in Atkeson (2020), but also consistent with the fraction of infected agents that recovered or died according to the WHO.

The estimates for τ are borrowed from Ferguson et al. (2020), who calculated that the average duration of stay in hospital is of 8 days if critical care is not required, and 16 days (with 10 days in ICU) if critical care is required. Ferguson et al. (2020) also estimates that 6.3% of those between the ages of 40-49 require critical care whereas 27% of those between the ages of 60-69 require it. Using these estimates, we set τ=1/8.5 for the young and τ=1/10.2 for the old so that in the model they spend 8.5 days and 10.2 days quarantined respectively. Importantly, in South Korea all the confirmed cases have been hospitalized. Lastly, although we focus mainly on comparing the total number of cases in each of the counter-factual exercises, we set the fatality rate to be 0.21 % for the young and 2.73% for the old. We

obtain these estimates from the Korean Centers for Disease Control & Prevention, which estimate a fatality rate those fatality rates for the groups between 40-49 years of age and 60-69 years of age.

7.3 Elasticity Parameters

We calibrate the elasticity parameters in the gravity equation, α^a, δ^a and ξ^a, using the structure of the model and the data compiled and made public by the local government of Seoul on the total number of cases, including detailed information of the confirmed cases, such as their demographic characteristics as well as their travel logs during the days before they were quarantined.

Using this information, the observed commuting probabilities during the pandemic period, and our estimated values for ϵ^{wd} and ϵ^{wn}, we estimate equation 8 to find the elasticities α^a, δ^a and ξ^a. Notice that in equation 8, these parameters are scaled by ϵ^{wd} on the weekdays and ϵ^{wn} on the weekends. Thus, we infer them using a constrained regression approach. Specifically, we estimate:

$$\begin{aligned}\ln\, \pi^a_{ijt} = \varphi^a_1 \tau_{ij} + \varphi^a_2 \ln\, C_{jt} + \varphi^a_3 \ln\, C_{jt} \times \mathbb{1}\left\{k = \text{weekend}\right\} \\ + \varphi^a_4\, \ln V_{jt} + \varphi^a_5\, \ln V_{jt} \times \mathbb{1}\left\{k = \text{weekend}\right\} + \theta^a_i + \lambda^a_j + \nu^a_{ijt}\end{aligned}$$

subject to $\varphi^a_3 = \frac{(\epsilon^{wn} - \epsilon^{wd})}{\epsilon^{wd}} \varphi^a_2$ and $\varphi^a_5 = \frac{(\epsilon^{wn} - \epsilon^{wd})}{\epsilon^{wd}} \varphi^a_4$. These restrictions are such that the elasticities are consistent with our previously estimated values for ϵ^{wd} and ϵ^{wn}. This is because $\varphi^a_2 = \epsilon^{wd}\delta^a$ is the elasticity of commuting flows with respect to the local cases during the weekdays and $\varphi^a_2 + \varphi^a_3 = \epsilon^{wn}\delta^a$ is the elasticity during the weekends. Similarly, $\varphi^a_4 = \epsilon^{wd}\xi^a$ is the elasticity of commuting flows with respect to the number of visits during the weekdays and $\varphi^a_4 + \varphi^a_5 = \epsilon^{wn}\xi^a$ is the elasticity during the weekends. We follow a similar procedure to estimate the elasticity of commuting flows to the total cases in Seoul:

$$\ln\, \pi^a_{ijt} = \iota^a_1 \tau_{ij} + \iota^a_2 \ln\, C_t + \iota^a_3 \ln\, C_t \times \mathbb{1}\left\{k = \text{weekend}\right\} + \theta^a_i + \lambda^a_j + \nu^a_{ijt}$$

subject to $\iota^a_3 = \frac{(\epsilon^{wn} - \epsilon^{wd})}{\epsilon^{wd}} \iota^a_2$ where $\iota^a_2 = \epsilon^{wd}\alpha^a$ and $\iota^a_2 + \iota^a_3 = \epsilon^{wn}\alpha^a$. The calibrated parameters are shown in Table 5.

7.4 Internally Calibrated Parameters

The rest of the parameters, β and d_I, are calibrated internally using two moments from the data. First, we target total number of detected cases in Seoul (861) until May 31st, which

captures the overall spread of the disease. Second, we target the fraction of undetected infections from the estimates in Stock et al. (2020), who use results from Iceland's two testing programs and estimate that the fraction of undetected infections range from 88.7% to 93.6%. We target a fraction of 90% undetected infections, which are also consistent with the estimates for the US in Hortaçsu et al. (2020). We show sensitivity analysis of the internally calibrated parameters in Section 8.4. The calibrated parameters are shown in Table 5.

Table 5: Parameter Values

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.
α^a	0.00798, 0.01171	Elasticity of commuting to total confirmed cases by age.
δ^a	0.00466, 0.00621	Elasticity of commuting to local confirmed cases by age.
ξ^a	0.00772, 0.01046	Elasticity of commuting to local visits by infected by age.
Internally Calibrated		
β	0.15513	Transmission rate (target: total cases by May 31st).
d_I	0.01632	Daily detection rate (target: fraction of undetected infections).

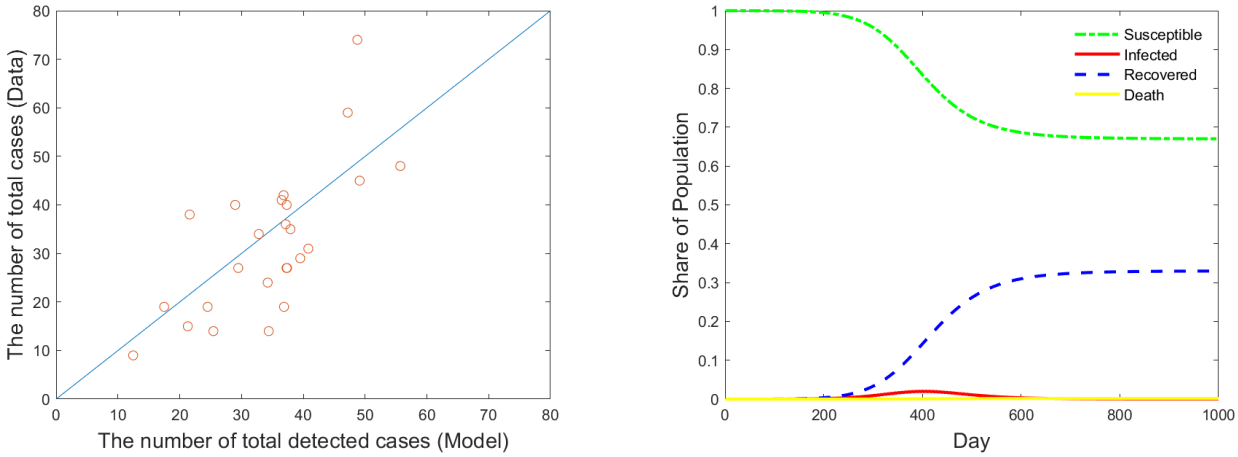
7.5 Simulation

Our initial conditions are the first 4 cases confirmed in the city of Seoul placed at the districts where the people infected reside. All of these cases can be traced back to people that at some point visited Wuhan, China. Since 90% of the cases are undetected, our initial conditions are a total of 4 cases in quarantine and 36 cases undetected. The infected people that are undetected follow the predicted commuting patterns of the model. The first day of the simulation is January 30th, 2020. In the data, two distinct places are disclosed on average. Following this, in the model, we calculate places visited by quarantined people by top destinations during weekdays and weekends.

We simulate the model to evaluate its performance relative to the actual data of confirmed cases in Seoul. Panel (a) of Figure 4 shows the performance of the model predicting the data of confirmed cases in each of the districts of Seoul. Each dot in the figure represents the

confirmed cases in each district by May 31st, 2020. The figure shows that the model is able to replicate well the geographical spread of the disease, as most dots are close to the 45 degree line. Panel (b) of Figure 4 shows the SIR dynamics that will take place in the upcoming months as predicted by the model. The figure shows that according to our model over the next two years approximately 30 percent of the population of Seoul will be infected. This can be seen by the fraction of people that eventually recover or die from the disease. The peak of infection will take place approximately 13-14 months after the outbreak of the virus.

Figure 4: Predicted Spread of Disease



(a) Data vs Model (b) SIR Dynamics

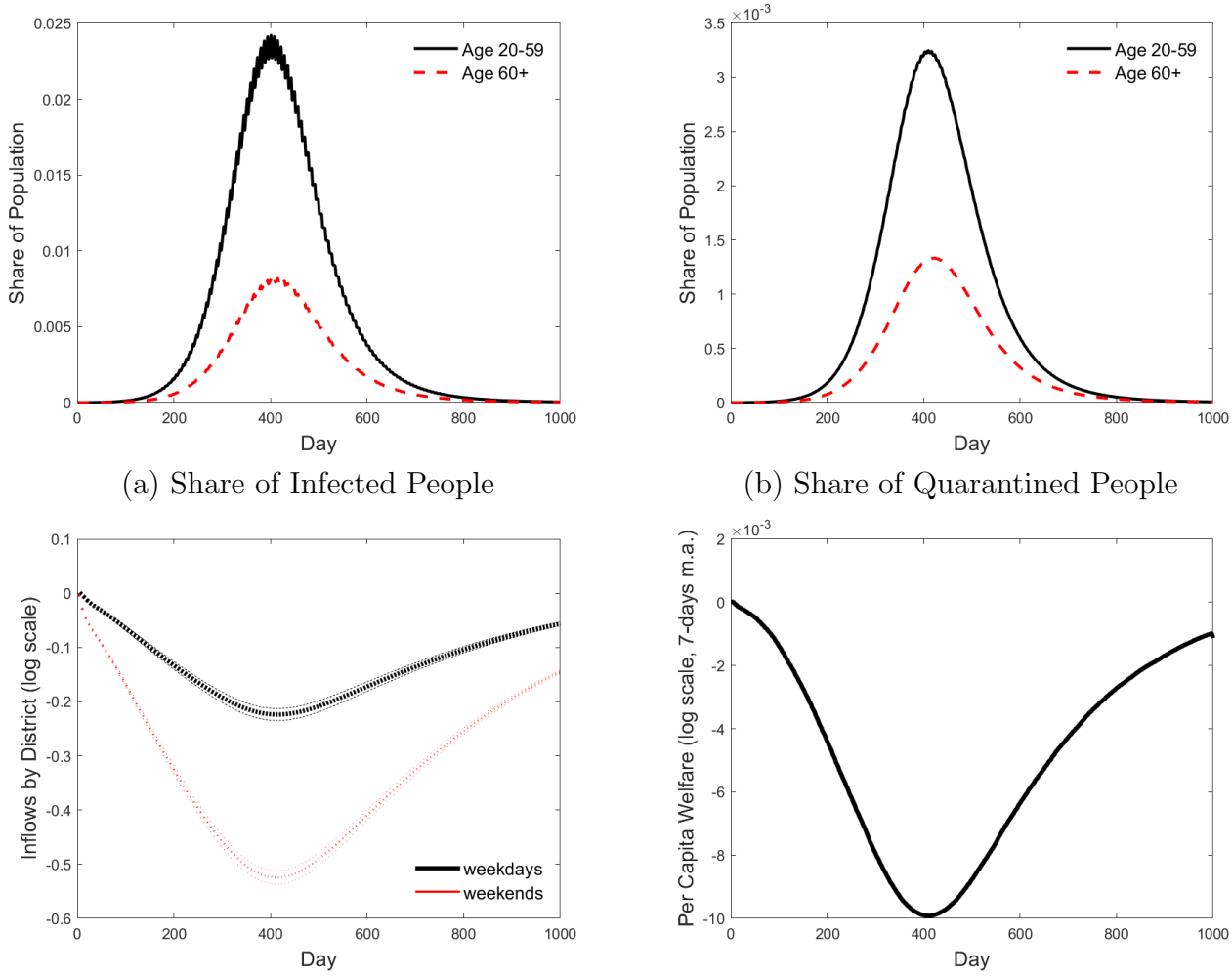
Notes: Panel (a) shows the total number of confirmed cases in the data (y-axis) and in the model (x-axis). Each dot is a district of Seoul. The blue line indicates the 45 degree line. Panel (b) shows the epidemic dynamics of the city of Seoul. They are simulated using a SIR compartmental mode across the districts of Seoul. The figure shows the fraction of the population that is susceptible (green line), infected (red line), quarantined (black line), recovered (blue line), or death (yellow line).

Given the policies so far implemented in the city of Seoul, the peak of infection will involve less than 2% of the population at the same time. Figure 5 shows the share of infected and quarantined population by age groups. Panel (a) shows that the share of people infected is, at its peak, approximately 2.4% of the young and 0.8% of the old. Panel (b) shows that at the peak around 0.035% of the young and 0.015% of the old are quarantined. Relative to the people below 60 years of age, there are less people above 60 years of age that are infected or quarantined at a given point in time. This is because, on average, they are more likely to stay at the home sector and they respond more drastically to the information related to the number of confirmed cases and visits in each district.

Panel (c) of Figure 5 shows the change in inflows by district during weekdays and week-

ends. As disease spreads, inflows to each district declines. At the peak of disease, average inflows are 20 percent lower during weekdays and 50 percent lower during weekends. Consistent with empirical evidences, workers change commuting patterns more during weekends in the model. Panel (d) of Figure 5 shows the change in per capita welfare, aggregating weekdays and weekends by taking 7-days moving average. The welfare declines because workers realize the cost of getting infected and change their commuting behavior. At the peak of disease, per capita welfare declines by 1 percent.

Figure 5: The Infected, The Quarantined, Inflows by District, and Per Capita Welfare



(a) Share of Infected People

(b) Share of Quarantined People

(c) Inflows by District

(d) Per Capita Welfare

Notes: Panel (a) shows the share of infected population for the young (age 20-59) and the old (age 60+). Panel (b) shows the share of quarantined population for the young (age 20-59) and the old (age 60+). Panel (c) shows the change in inflows by district during weekdays and weekends, along with 95 percent confidence intervals. Panel (d) shows the change in per capita welfare as a 7-day moving average.

8 Counter-factual Exercises

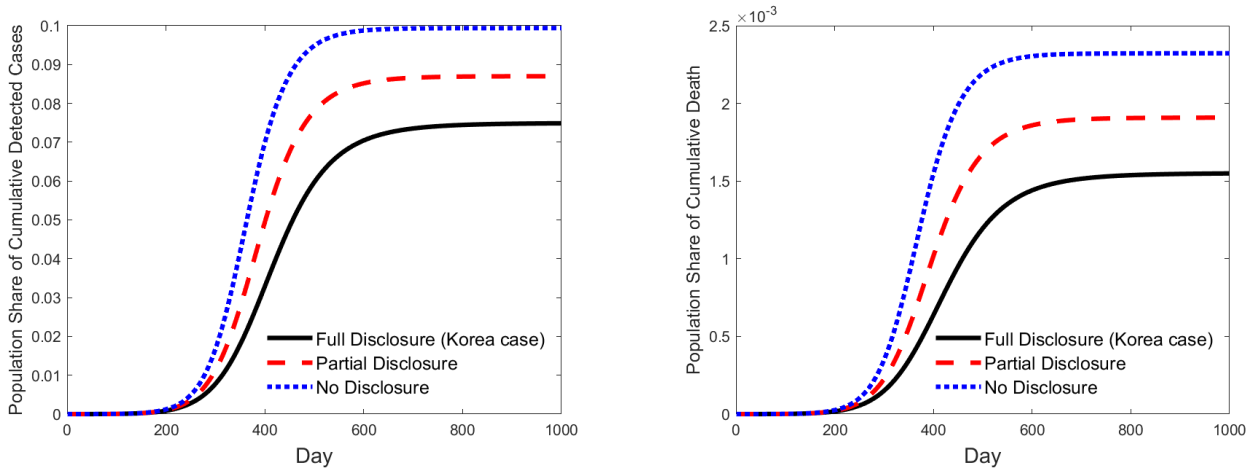
In this section, we evaluate the effectiveness of different non-pharmaceutical mitigation strategies: (i) information disclosure, (ii) lockdown, and (iii) shutdown. First, to quantify the effectiveness of the disclosure policy, we simulate the model without allowing commuting to respond to information. This can be interpreted as the case without information disclosure or, alternatively, the case where people cannot change their commuting behavior despite of information. We believe the first interpretation is closer to the real world given the observed changes in commuting behavior from the foot traffic data. Second, we quantify the effectiveness of a lockdown policy and a shutdown policy, relative to the information disclosure case, simulating the model implementing the geographic restrictions of such policies and assuming there is no information disclosure in these cases.

8.1 Information Disclosure

The government of Seoul disclosed to different types of information: (i) the total number of confirmed cases in each district, C_j, and (ii) total number of visits by confirmed cases to each of the districts, V_j. We first examine a partial disclosure case by shutting down each channel and chaining the results. The case without information disclosure is that in which $\delta^a = \xi^a = 0$. In this case, the gravity equation depends on the physical distance from origin to destination and semi-elasticity with respect to total cases in Seoul, governed by κ and α respectively.

Figure 6 shows the share of the population that is infected and detected from day 0 to 800 in Seoul. Under partial and no information disclosure, we find more detected cases. Under full information disclosure (the current case of South Korea), the city reaches heard immunity at a point where a substantially lower share of the population is infected. The difference between the full information disclosure and no disclosure scenarios is also significant when comparing the total number of deaths shown in Panel (b). The scenario with no disclosure of information yields approximately 50 percent more the number of deaths compared to the full disclosure scenario. This is because individuals above 60 years of age, those that are more vulnerable to the virus, are sensitive to the information disclosed and altered their commuting patterns significantly in response.

Figure 6: Disclosure Policy: Detected Cases and Deaths



(a) Cumulative Detected Cases (b) Cumulative Death

Notes: Panel (a) shows a population share of cumulative detected cases under three scenarios: (i) full disclosure (South Korea case), (ii) partial disclosure (no travel route information) and (iii) no disclosure. Panel (b) the share of he population death under the same three scenarios.

Table 6 reports the total number of detected cases, the total number of death, and the welfare losses over two years. For the young, the total number of death increases by 15 percent under partial disclosure, and 30 percent under no disclosure. For the old, the total number of death increases by 30 percent under partial disclosure, and 66 percent under no disclosure.

The public health benefits from information disclosure come at the cost of welfare losses. We calculate the average daily welfare losses relative to the period before the outbreak of the virus. The average daily welfare loss for the young (old) decreases by 0.14 (0.28) percentage points under partial disclosure and 0.33 (0.49) percentage points under no disclosure. In this case, the welfare losses are attenuated due to the fact that workers are able to choose their second or third best location when they maximize their utility even if their preferred commuting choice is disrupted by the information obtained about the confirmed cases.

Table 6: Disclosure Policy: Cases, Death and Welfare

	Full Disclosure	Partial Disclosure	No Disclosure
Total # of Cases	602,999	707,475	810,158
Total # of Death	12,435	15,505	18,928
age 20-59	5,003	5,796	6,540
age 60+	7,432	9,709	12,388
Welfare Loss per day	0.57	0.32	0.14
age 20-59	0.55	0.31	0.13
age 60+	0.67	0.39	0.18

Notes: The table reports the total number of detected cases, the total number of death, and the welfare losses over two years in the city of Seoul under full disclosure (original model), partial disclosure, and no disclosure. The welfare losses are shown in percent.

8.2 Lockdown

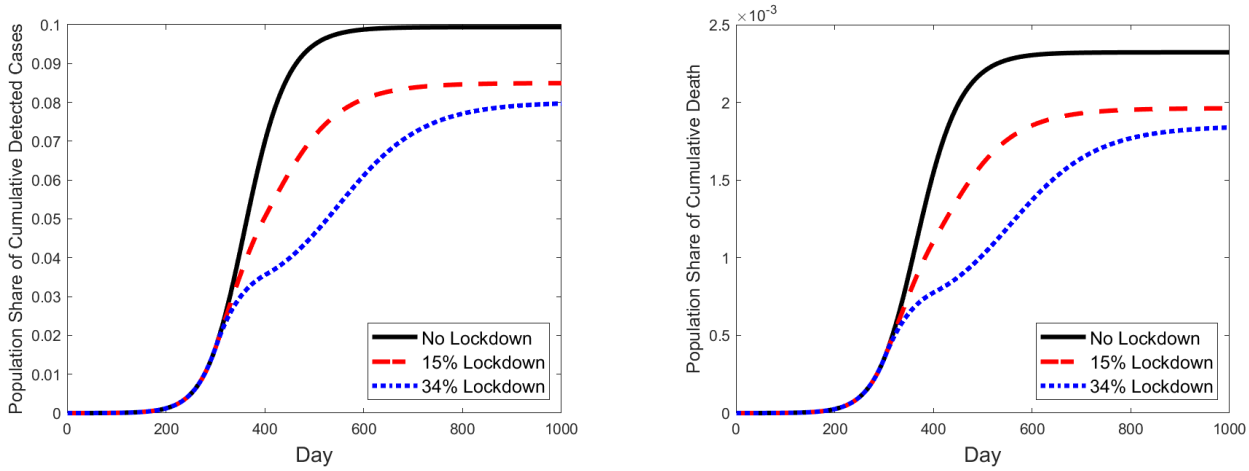
We impose a lockdown policy assuming that in this case no information about the confirmed cases is disclosed. Under the lockdown policy, similar to that so far implemented in many countries including the United States, a certain fraction of the population is required to stay at home. In the model, this is implemented by randomly choosing a certain fraction of people and forcing them to work from home during weekdays and to stay at home during weekends. We ignore for now the possibility that the mandated lockdown is only partially effective (e.g. people ignoring the government's order). Naturally, the disease does not spread at home; at home, workers who are susceptible cannot be infected and workers who are infected but not detected do not spread the disease anymore. We assume that the lockdown policy is implemented from day 300 to 400, when the spread of disease is the fastest when no information is disclosed.

Figure 7 shows the total number of detected cases from day 0 to day 800 when a lockdown policy is implemented from day 300 to 400.[5] We implement lockdowns in which 15% and 34% of the population is mandated to stay at home. The lockdown policy slows down the spread of the disease right after it is implemented. Once the economy is reopened, the disease

[5] An lockdown implemented earlier substantially delays infections and herd immunity, but it has virtually no impact on that total number of infected once herd immunity is achieved. In contrast, a later intervention substantially reduces the long-run level of infections. There is less overshooting of herd immunity because the later intervention slows the pandemic's momentum.

spreads further but the city reaches the point where the virus does not spread anymore after infecting a lower fraction of the population compared to the case without a lockdown.

Figure 7: Lockdown Policy: Detected Cases and Deaths



(a) Cumulative Detected Cases (b) Cumulative Death

Notes: Panel (a) shows the share of the population that is infected and detected under three scenarios: (i) no lockdown, (ii) 15% lockdown and (iii) 34% lockdown. For all scenarios, we assume no public information disclosure. Panel (b) shows the share of the population that dies in each of the three scenarios.

Table 7 reports the total number of detected cases, the total number of deaths, and the welfare losses over two years under the different lockdown policies. Under 34% lockdown, the total number of cases is comparable to the model with full disclosure. In this scenario, the total number of deaths is higher, especially among the old. The welfare losses under a lockdown are substantially higher relative to the full information disclosure scenario. The lockdown misallocates workers and mitigation efforts. This is because, under the lockdown policy, workers who do not like working from home are mandated to do so. On the other hand, under information disclosure, workers who enjoy working from home select themselves to do so after seeing more confirmed cases and visits at their preferred districts. Similar circumstances occur with people with low health risks or that have recovered from the disease as they are mandated to stay home under the lockdown whereas, under full information, those that have higher health risks are those that choose to stay at home.

Table 7: Lockdown Policy: Cases, Death and Welfare

Lockdown day 300-400 (100 days)	No Lockdown	15% Lockdown	34% Lockdown
Total # of Cases	810,158	686,495	604,223
Total # of Death	18,928	15,822	13,797
age 20-59	6,540	5,558	4,894
age 60+	12,388	10,264	8,903
Welfare Loss per day	0.14	0.48	0.94
age 20-59	0.13	0.56	1.15
age 60+	0.18	0.23	0.28

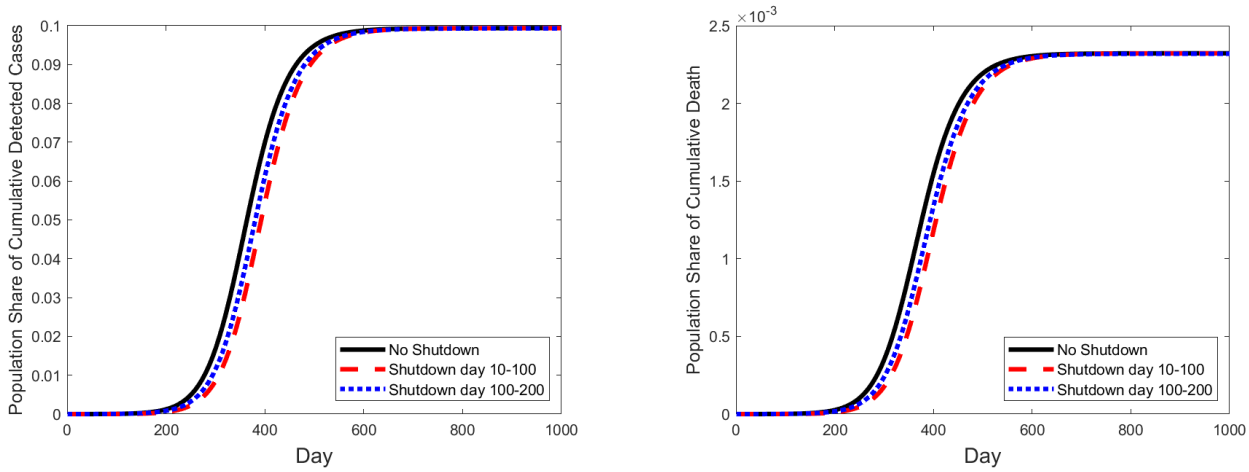
Notes: The table reports the total number of detected cases, the total number of death, and the welfare losses over two years in the city of Seoul under no lockdown, 15% lockdown, an 34% lockdown. Note that the first column is the same as the last column in table 6. The welfare losses are shown in percent.

8.3 Shutdown

Lastly, we impose a shutdown policy in the model. The initial 4 confirmed cases in January 2020 where found in four distinct districts. Under a shutdown policy, people living in those districts cannot commute outside. They could work from home or commute to the same district. People living in other districts cannot commute to those four districts. As a result, even if the policy is implemented only in four districts, the commuting choices of all workers are limited.

Figure 8 shows the total number of detected cases under different shutdown policies. We study the effect of two shutdown policies. The first is implemented for 90 days, from day 10 to day 100. The second is implemented for 100 days from days 100 to 200. An early shutdown delays the spread of the disease but it has virtually no impact on total share of the population infected in the long run for two reasons. First, Seoul is a city that is very inter-connected with vast amount of commuting across districts. And, second, undetected cases commute across districts; this interaction, even if short-lasting, is enough to spread the disease in the entire city over time. A shutdown policy implemented later is not as effective either; it does not have a substantial impact reducing the total number of cases in the city since by the time this policy is implemented, the disease has already spread across all districts.

Figure 8: Shutdown Policy: Detected Cases and Deaths



(a) Cumulative Detected Cases (b) Cumulative Death

Notes: Panel (a) shows a population share of cumulative detected cases under three scenarios: (i) no shutdown, (ii) shutdown from day 10 to 100 and (iii) shutdown from day 200 to 300. For all scenarios, we assume no public information disclosure. Panel (b) shows the cumulative share of the population that dies under each scenario.

Table 8 reports the total number of detected cases, the total number of deaths, and the welfare losses over two years under different shutdown policies. Shutdowns as mitigation strategies are very costly. Relative to the case with no information disclosure, under the early shutdown scenario, the welfare losses increase by 0.63 percentage points for the young and 0.28 percentage points for the old. However, total number of deaths over two years is almost the same. A similar picture emerges when we examine a shutdown policy implemented later. This policy is less effective than a lockdown policy, yielding the same welfare losses, at decreasing the total number of cases due to Seoul's level of interconnectedness and the large presence of undetected cases at the time the policy is implemented.

Table 8: Shutdown Policy: Cases, Death and Welfare

	No Shutdown	Shutdown 10-100	Shutdown 100-200
Total # of Cases	810,158	809,823	809,304
Total # of Death	18,928	18,918	18,902
age 20-59	6,540	6,537	6,533
age 60+	12,388	12,381	12,369
Welfare Loss per day	0.14	0.69	0.74
age 20-59	0.13	0.76	0.83
age 60+	0.18	0.46	0.49

Notes: The table reports the total number of detected cases, the total number of death, and the welfare losses over two years in the city of Seoul under no shutdown, shutdown from day 10 to 100, and shutdown from day 100 to 200. For all scenarios, we assume no public information disclosure. Note that the first column is the same as the last column in table 6. The welfare losses are shown in percent.

8.4 Sensitivity Analysis

We check the robustness of our results by checking their sensitivity to adjusting the two internally calibrated parameters of our model, the transmission rate (β) and the daily detection rate (d_I). First, we run a simulation and a counter-factual without disclosure both increasing and decreasing β by 20 percent. Table 10 shows that the results under 20% lower and 20% higher transmission rate, β. With lower β, the model delivers much lower number of cases, deaths, and welfare losses. Importantly, the disclosure of information is more effective at reducing the number of deaths, which reduce by 65%. On the other hand, higher β yields higher number of cases, deaths, and welfare losses. Information disclosure is still effective, but less so, reducing the number of death by 28%.

Table 9: Sensitivity Analysis: Lower and Higher β

	20% lower β=0.1241		20% higher β=0.1862	
	Full disclosure	No disclosure	Full disclosure	No disclosure
Total # of Cases	33,317	91,743	913,008	1,108,376
Total # of Death	690	1,989	19,441	27,176
age 20-59	273	742	7,528	8,841
age 60+	417	1,247	11,913	18,335
Welfare Loss per day	0.21	0.08	0.51	0.11
age 20-59	0.20	0.08	0.48	0.10
age 60+	0.26	0.10	0.58	0.14

Notes: The table reports the total number of detected cases, the total number of death, and the welfare losses over two years in the city of Seoul under full disclosure and no disclosure with lower and higher β. The welfare losses are shown in percent.

Second, we re-calibrate the model by targeting a lower and a higher fraction of undetected infections. Note that in the benchmark calibration, we target 90% of undetected infections to be consistent with the estimates in Stock et al. (2020). Under 80% undetected infections, β (0.17144) and d_I (0.0357) are calibrated to be higher than in our baseline calibration (0.15513 and 0.01632) so that they also jointly match the total number of cases by May 31st. With a much higher daily detection rate, there is more information disclosed on the total number of cases and visits. As a result, workers change their commuting behavior more and the number of deaths declines significantly. Thus, information disclosure is more effective in this case. On the other hand, when we target 95% of undetected infections, β (0.15219) and d_I (0.0077) are calibrated to be lower than benchmark calibration. With lower detection rate, information disclosure is less effective because there is less information disclosed on the local cases and visits. This result emphasizes that information disclosure could be more effective when it is combined with massive testing and contact tracing to increase the number of detected cases.

Table 10: Sensitivity Analysis: Lower and Higher Fraction of Undetected Infections

	Frac. of undetected infection=0.8 β=0.17144, d_I=0.0357		Frac. of undetected infection=0.95 β=0.15219, d_I=0.0077	
	Full disclosure	No disclosure	Full disclosure	No disclosure
Total # of Cases	430,112	867,881	444,449	539,707
Total # of Death	4,943	11,265	17,818	24,439
age 20-59	2,075	4,104	6,821	8,050
age 60+	2,868	7,161	10,996	16,389
Welfare Loss per day	0.54	0.16	0.48	0.12
age 20-59	0.51	0.15	0.46	0.11
age 60+	0.63	0.20	0.56	0.14

Notes: The table reports the total number of detected cases, the total number of death, and the welfare losses over two years in the city of Seoul under full disclosure and no disclosure with lower and higher fraction of undetected infections. β and d_I are re-calibrated to match this new target moment, along with the number of total cases by May 31st (861). The welfare losses are shown in percent.

9 Conclusion

This paper uses an SIR model with multiple sub-populations and an economic model of commuting choice between the sub-populations to measure the effect of the disclosure of information about COVID-19 cases in Seoul. We use the model to calibrate the effect of the change in commuting patterns after the public disclosure of information on the transmission of the virus and the economic losses due to the change in commuting patterns. We find that compared to a scenario without disclosure, public disclosure reduces the number of COVID-19 cases by 200 thousand and deaths by 6.5 thousand in Seoul over 2 years. And compared to a lockdown that results in about the same number of cases as the full disclosure strategy, the latter results in economic losses that are 40% percent lower.

To be clear, disclosure of public information is infringes upon to the privacy of the affected individuals. We do not attempt to measure the cost of the loss of privacy, but whenever such measures are available, they can be weighted against the benefits of public disclosure we provide here. Also, in our analysis the community (in Seoul) reaches herd immunity within the next two years. That is, we assume no vaccine will be available in the next two years.

The analysis will obviously be different, and the tradeoffs between the different scenarios we model in the paper, will be different as well if a vaccine is available within the next two years.

The broader point is that, in the absence of a vaccine, targeted social distancing may be a much more effective way to reduce the transmission of the disease while minimizing the economic cost of social isolation. We view the public dissemination of information in Korea as one way to accomplish that. We are hopeful that there could be other, perhaps more effective ways, to target social distancing to get the maximum benefit for the least cost.

References

Acemoglu, Daron, Victor Chernozhukov, Iván Werning, and Michael D. Whinston, "A Multi-Risk SIR Model with Optimally Targeted Lockdown," *NBER Working Paper No. 27102*, 2020.

Ahlfeldt, Gabriel M, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf, "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 2015, *83* (6), 2127–2189.

Atkeson, Andrew, "What Will Be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios," *NBER Working Paper No. 26867*, 2020.

Brockmann, Dirk and Dirk Helbing, "The hidden geometry of complex, network-driven contagion phenomena," *Science*, 2013, *342* (6164), 1337–1342.

Danon, Leon, Thomas House, and Matt J Keeling, "The role of routine versus random movements on the spread of disease in Great Britain," *Epidemics*, 2009, *1* (4), 250–258.

Diamond, Peter A. and Eric Maskin, "An Equilibrium Analysis of Search and Breach of Contracts, I: Steady States," *Bell Journal of Economics*, 1979, *10*, 282–316.

_ and _ , "An equilibrium analysis of search and breach of contract II. A non-steady state example," *Journal of Economic Theory*, 1981, *25* (2), 165 – 195.

Eaton, Jonathan and Samuel Kortum, "Technology, geography, and trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Ferguson, Neil, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, ZULMA Cucunuba Perez, and Gina Cuomo-Dannenburg, "Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand," 2020.

Hortaçsu, Ali, Jiarui Liu, and Timothy Schwieg, "Estimating the Fraction of Unreported Infections in Epidemics with a Known Epicenter: an Application to COVID-19," Technical Report 2020.

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow, "The allocation of talent and us economic growth," *Econometrica*, 2019, *87* (5), 1439–1474.

Hufnagel, Lars, Dirk Brockmann, and Theo Geisel, "Forecast and control of epidemics in a globalized world," *Proceedings of the National Academy of Sciences*, 2004, *101* (42), 15124–15129.

Keeling, Matt J and Pejman Rohani, *Modeling infectious diseases in humans and animals*, Princeton University Press, 2011.

_ , Leon Danon, Matthew C Vernon, and Thomas A House, "Individual identity and movement networks for disease metapopulations," *Proceedings of the National Academy of Sciences*, 2010, *107* (19), 8866–8870.

Lee, Munseob, "Allocation of female talent and cross-country productivity differences," 2016.

McFadden, Daniel, "The measurement of urban travel demand," *Journal of Public Economics*, 1974, *3* (4), 303–328.

Monte, Ferdinando, Stephen J Redding, and Esteban Rossi-Hansberg, "Commuting, migration, and local employment elasticities," *American Economic Review*, 2018, *108* (12), 3855–90.

Stock, James H, Karl M Aspelund, Michael Droste, and Christopher D Walker, "Estimates of the undetected rate among the sars-cov-2 infected using testing data from Iceland," *MedRxiv*, 2020.

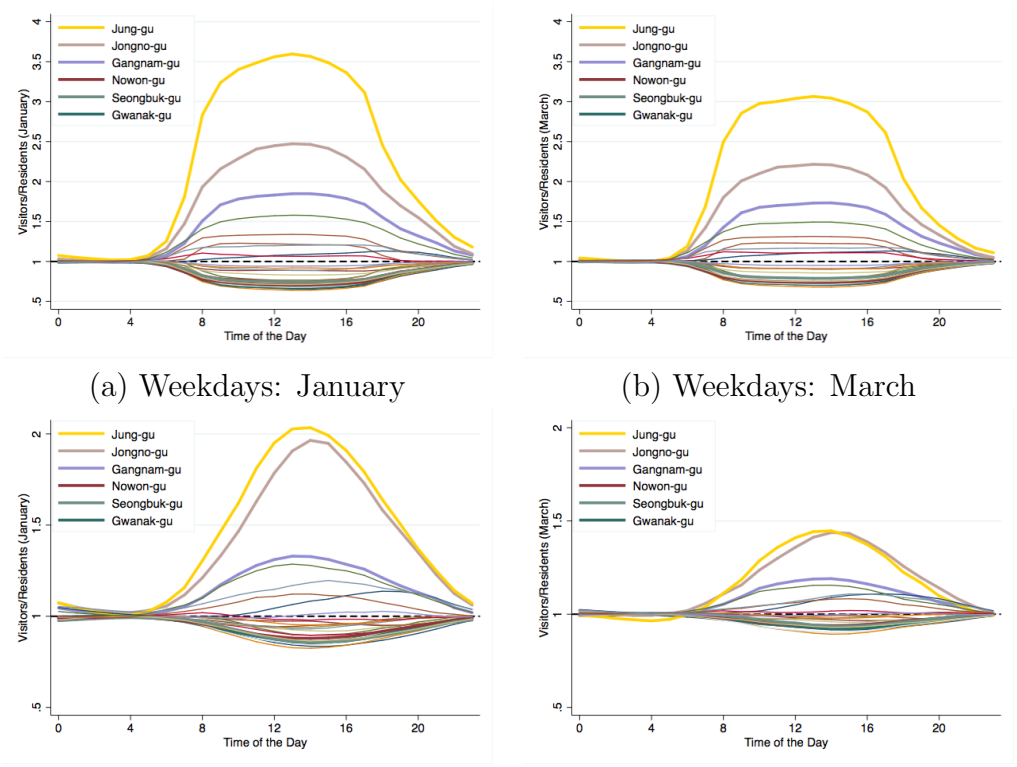
Tsivanidis, Nick, "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio," *UC Berkeley, mimeo*, 2019.

Wesolowski, Amy, Nathan Eagle, Andrew J Tatem, David L Smith, Abdisalan M Noor, Robert W Snow, and Caroline O Buckee, "Quantifying the impact of human mobility on malaria," *Science*, 2012, *338* (6104), 267–270.

APPENDIX

A Additional Empirical Findings

Figure A1: Share of Visitors over Residents by District



(a) Weekdays: January

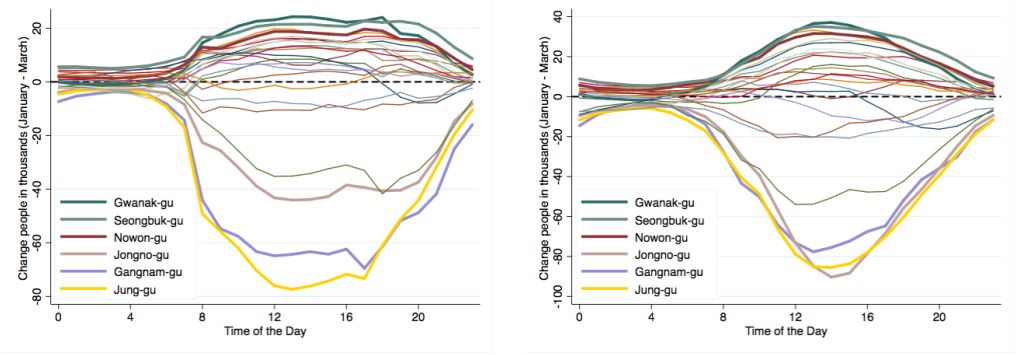
(b) Weekdays: March

(c) Weekends: January

(d) Weekends: March

Notes: The figure shows the ratio of visitors over residents in each district every hour of the day. Each line represents one of the 25 districts of Seoul. Panel (a) shows the the ration of visitors over residents for the weekdays during the month of January and Panel (b) for the weekdays during the month of March. Panel (c) shows the same ratio for the weekends of January and Panel (d) for the weekends of March.

Figure A2: Change in the Number of People by District (January - March)



(c) Weekdays (d) Weekends

Notes: The figure shows the average hourly change in the number of people in each district from the month of January to the month of March. Each line represents one of the 25 districts of Seoul. The graph is plotted in thousands of people. Panel (a) shows the patterns during the weekdays and Panel (b) the same patterns but for the weekends.