NBER WORKING PAPER SERIES

THE COST OF PRIVACY: WELFARE EFFECT OF THE DISCLOSURE OF COVID-19 CASES

David O. Argente
Chang-Tai Hsieh
Munseob Lee

Working Paper 27220
http://www.nber.org/papers/w27220

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2020

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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Cost of Privacy: Welfare Effect of the Disclosure of COVID-19 Cases
David O. Argente, Chang-Tai Hsieh, and Munseob Lee
NBER Working Paper No. 27220
May 2020
JEL No. E0,I0

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 400 thousand and the number of deaths by 13 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 50% lower with full disclosure.

David O. Argente
Pennsylvania State University
Department of Economics
403 Kern Building
University Park
State College, PA 16801
dargente@psu.edu

Chang-Tai Hsieh
Booth School of Business
University of Chicago
5807 S Woodlawn Ave
Chicago, IL 60637
and NBER
chsieh@chicagoBooth.edu

Munseob Lee
University of California at San Diego
9500 Gilman Drive, #0519
La Jolla, CA 92093-0519
munseoblee@ucsd.edu

A data appendix is available at http://www.nber.org/data-appendix/w27220

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 925 thousand cases in Seoul, 17 thousand deaths, and economic losses that average 1.2% of GDP.

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 400 thousand more infected people and 13 thousand additional deaths in Seoul (over the next two years). The economic 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 0.33 percentage points lower for those under 60 and 0.22 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 at least 40% 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. As of May 13 of 2020, South Korea had 10,962 confirmed cases of COVID-19 and 259 deaths. The city of

[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).

Seoul, a city of almost 10 million inhabitants, had only 707 confirmed cases and 3 deaths. These numbers are remarkably low in comparison to cities of similar size. What's more, 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 discloses 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 Foot Traffic Data

We use the foot traffic data in Seoul, provided by Korean telecommunication company SK Telecom, which has the largest market share in Korean mobile phone market.[2] The 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 user is inferred by cellular tower triangulation. The data splits users by the gender and by age group. We use the data from January 1st to February 29th. 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.

[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).

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.

3.2 Seoul Survey

We 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 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.

From the survey we obtain, for each of the 25 districts in Seoul, bilateral commuting patterns. A unique feature of the survey is that it asks the working location of each respondent. Based on that information, we can restore representative origin-destination commuting patterns at the district level. We also obtain the wage distribution for different demographic groups across districts as well as estimates of the size of the home sector before the outbreak of the virus. Importantly, the survey contains information on the fraction of people that do not commute during weekdays and also the wages that they receive, two important inputs to the model develop in the next sections.

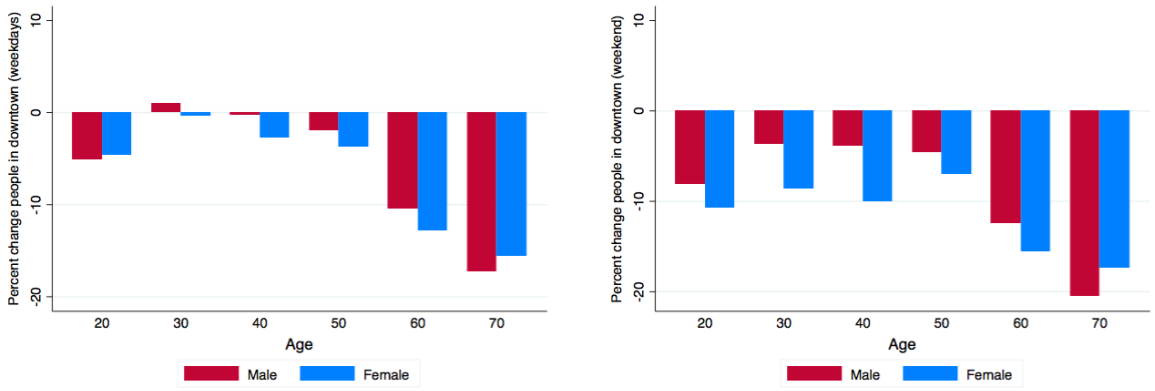
4 Empirical Patterns

Figure 1 shows the percent change in the average number of people in downtown (at 3pm) between the months of January and February.[3] Recall that the first case in Korea was iden-

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

tified 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 50 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; the most vulnerable groups decreased their commuting more substantially.

Figure 1: Percent Change People in Downtown

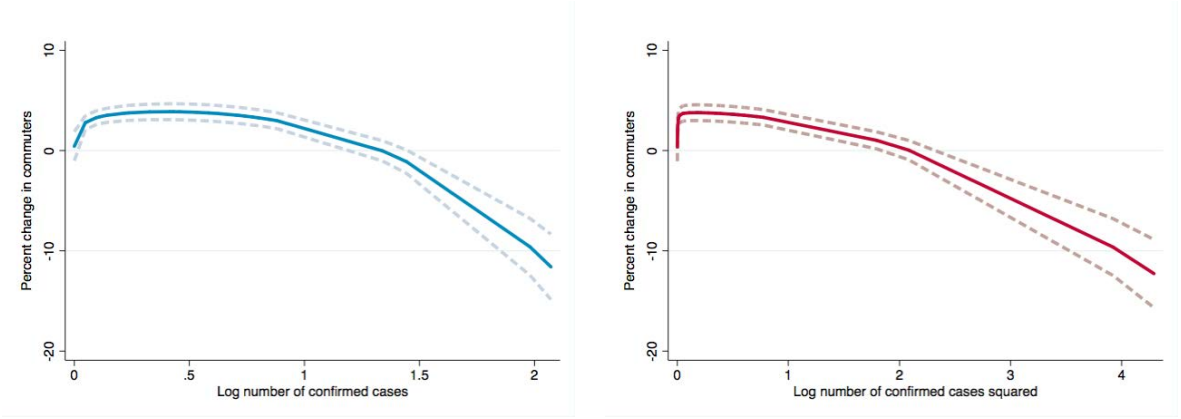


(a) Weekdays (b) Weekends

Notes: The figure shows the change in the average number of people in downtown in February relative to the previous month. 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.

This can also be seen if we focus on the commuting patterns as a function of the number of confirmed cases disclosed by the government in each district. We consider only the cases that are active, those that have not recovered or died. Panel (a) of Figure 2 shows that from January to February, people were more likely to commute to the districts with less confirmed cases and less likely to commute to the districts with more cases. In fact, Panel (b) shows the relationship is not linear; commuting patterns become more sensitive to the number of cases disclosed as they increase. Importantly, these changes took place without a government-imposed lockdown.

Figure 2: Commuting Changes and Confirmed Cases



(a) Log Cases (b) Log Cases Squared

Notes: The figure shows on the y-axis of both panels the change in the average number of people in each district in February relative to the previous month. The total number of people is approximated by the total people in each district at 3pm. The x-axis of Panel (a) shows the natural logarithm of the number of confirmed cases in each district, after considering only those that are active. The x-axis of Panel (b) shows the natural logarithm of the number of confirmed cases squared in each district, after considering only those that are active. The figure shows the relationship between these two variables using a fractional-polynomial prediction with 95% confidence intervals, where the weights are the total number of users in each demographic group.

We explore the response of people to the information of the total number of confirmed cases in their districts quantitatively implementing the following specification:

$$\ln H_{M,i,j,t} = \alpha + \iota_1 \ln C_{j,t} + \iota_2 \left(\ln C_{j,t}\right)^2 + \varphi_1 \ln V_{j,t} + \varphi_2 \left(\ln V_{j,t}\right)^2 + \theta_{i,j} + \gamma_t + e_{i,j,t} \qquad (1)$$

where $H_{M,i,j,t}$ is the number of people with characteristics i (i.e. age $\times$ gender), present in district j, on day t. Since we focus on the commuting patterns of people, we count the number of people present in each district at 3pm in the afternoon, when they are likely to be at their work location during weekdays or at a given place for leisure during weekends.[4] Public disclosure is captured by the total number of confirmed cases in each district, $\ln C_{j,t}$. 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. To capture the additional information provided by the travel logs of infected individuals, we also include in the regression the number of visits they made to each district, $\ln V_{j,t}$. We only include the total visits of the last 14 days to capture only the visits of the active cases. The specification includes district

[4]Our results are qualitatively very similar if we focus on other times.

$\times$ age $\times$ gender effects, $\theta_{i,j}$, and time effects, γ_t.

The results can be found in Table 3. Column (1) shows that people visit less the districts with higher number of confirmed cases. Column (2)-(3) show that they also are less likely to visit districts that were visited by infected individuals. The columns show that the people of Seoul are responding to the disclosed information about the local cases in their districts. Importantly, Column (4) shows that the elasticity of the number of commuters to a given district increases as the number of cases and the number of visits increase. This is relevant since, as the pandemic progresses, the number of commuters to each district is likely to decrease substantially. Columns (5) and (6) show that similar patterns occur both during weekdays and during weekends.

Table 3: Commuting Response to Public Disclosure

	(1)	(2)	(3)	(4)	(5)	(6)
ln Cases	-0.011***		-0.012***	0.010***	0.011***	0.008
	(0.002)		(0.002)	(0.003)	(0.003)	(0.007)
$(\ln \text{Cases})^2$				-0.015***	-0.013***	-0.020***
				(0.002)	(0.002)	(0.004)
ln Visits		-0.004***	-0.003***	0.005***	0.001	0.016***
		(0.001)	(0.001)	(0.002)	(0.002)	(0.005)
$(\ln \text{Visits})^2$				-0.003***	-0.003***	-0.002*
				(0.000)	(0.000)	(0.001)
Observations	9,600	8,640	8,640	8,640	6,048	2,592
R-squared	0.994	0.994	0.994	0.994	0.996	0.988
District-Age-Gender	Y	Y	Y	Y	Y	Y
Time	Y	Y	Y	Y	Y	Y
Days	All	All	All	All	Weekdays	Weekends

Notes: The table shows the results of estimating equation (1). 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)-(4) include all the days of the week, Column (5) only weekdays and Column (6) only weekends.

Table 4 splits individuals by their age group: those under 60 years of age and those above 60 years of age. We do so to quantify, more formally, the evidence presented in Figure 1 and to understand whether the elasticity of commuting to the public information of the cases is larger for the more fragile groups. The table also splits the days of the week into weekdays

and weekends given that individuals between 20 and 60 years of age are more likely to have work responsibilities during the weekdays. The table shows similar patterns as those depicted in Figure 2 and in Table 3; commuting decreases as the information of the number of cases or visits by infected people increase. Importantly, Column (2) shows that people above 60 years of age respond more drastically to the information on the total number of cases and less to the visits to their district by infected people. These patterns are more noticeable during weekends as shown in Columns (3)-(4). Given the differences in the commuting response by people of different age groups, who also differ in how vulnerable they are to the virus, as well as across days of the week, we allow for heterogeneity along these dimensions in the model developed in Section 6.[5]

Table 4: Commuting Response to Public Disclosure by Age Group

	(1)	(2)	(3)	(4)
ln Cases	0.009**	0.016***	0.004	0.016
	(0.004)	(0.004)	(0.008)	(0.012)
$(\ln \text{Cases})^2$	-0.011***	-0.017***	-0.018***	-0.023***
	(0.003)	(0.003)	(0.005)	(0.008)
ln Visits	0.002	-0.001	0.020***	0.008
	(0.002)	(0.002)	(0.005)	(0.008)
$(\ln \text{Visits})^2$	-0.003***	-0.001**	-0.003**	0.000
	(0.000)	(0.001)	(0.001)	(0.002)
Observations	4,032	2,016	1,728	864
R-squared	0.996	0.995	0.985	0.982
District-Age-Gender	Y	Y	Y	Y
Time	Y	Y	Y	Y
Days	Weekdays	Weekdays	Weeekends	Weeekends
Group	Under 60	Above 60	Under 60	Above 60

Notes: The table shows the results of estimating equation (1). 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] In Section A.1, we show that similar patterns emerge if we consider a commuting measures that counts the number of visitors to a given district net of the residents in the district.

5 Modeling Spread of Disease

The typical approach in the epidemiology literature is to study the dynamics of the pandemic, for infected, deaths, 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:

$$
\begin{aligned}
\dot{S}_i^a(t) &= -\beta \sum_j \left[\frac{\sum_{a,s} \pi_{sj}^a(t) I_s^a(t)}{\sum_{a,s} \pi_{sj}^a(t) N_s^a(t)} \times \pi_{ij}^a(t) S_i^a(t) \right] \\
\dot{I}_i^a(t) &= \beta \sum_j \left[\frac{\sum_{a,s} \pi_{sj}^a(t) I_s^a(t)}{\sum_{a,s} \pi_{sj}^a(t) N_s(t)} \times \pi_{ij}^a(t) S_i^a(t) \right] - \gamma I_i^a(t) - d_I I_i^a(t) \\
\dot{Q}_i^a(t) &= d_I I_i^a(t) - \tau^a Q_i^a(t) \\
\dot{R}_i^a(t) &= \gamma I_i^a(t) + \tau^a Q_i^a(t)
\end{aligned} \tag{2}
$$

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.

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_{Ri}^a 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 employment district within the city that maximizes her utility. The price of the final traded good is chosen as the numeraire ($p = 1$). Districts differ in terms of their final goods productivity, number of infected individuals and 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: $U_{ijo}^a = C_{ijo}^a$, where C_{ijo}^a denotes the consumption index for worker o of age a residing in district i

and working in district j. This consumption index depends on the consumption of the traded good (c^a_{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 and residence (z^a_{jo}). This idiosyncratic shock captures the idea that individual workers can have idiosyncratic reasons for working in different parts of the city. The aggregate consumption index is simply:

$$C^a_{ijo} = \frac{z^a_{jo} c^a_{ijo}}{d^a_{ij}} \tag{3}$$

where the iceberg commuting cost $d^a_{ij} = e^{\kappa \tau_{ij} + \delta^a (\ln C_j)^2 + \xi^a (\ln V_j)^2}$ increases with the travel time (τ_{ij}) between district i and j. Travel time is measured in distance. The commuting costs, $d^a_{ij} \in [1, \infty)$, also increase with the total number of detected cases C_j that are active in the district where workers work, and with the total number of visits by detected cases V_j. Given our empirical evidence, we allow C_j and V_j to affect the commuting costs non-linearly. The parameter κ control the impact of distance on the commuting costs and the parameters δ^a and ξ^a the sensitivity of the commuting costs to the disclosed information on the detected cases, given our empirical evidence, both parameters are allowed to vary by age.

We model the heterogeneity in the utility that workers derive from living and working in 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^a_{ijo}) is drawn from an independent Fréchet distribution:

$$F(z^a_{jo}) = e^{E^a_j (z^a_{jo})^{\epsilon}}, \quad E_j > 0, \epsilon > 1$$

where the scale parameter E^a_j determines the average utility derived from working in district j; and the shape parameter $\epsilon > 1$ controls the dispersion of idiosyncratic utility.

The indirect utility from residing in district i and working in district j can be expressed in terms of the wage paid at this workplace (w^a_j), commuting costs (d^a_{ij}) and the idiosyncratic shock (z^a_{jo}):

$$u^a_{ijo} = \frac{z^a_{jo} w^a_j}{d^a_{ij}} \tag{4}$$

where we have used utility maximization and price for traded goods is normalized to 1 as a numeraire.[6]

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

Since indirect utility is a monotonic function of the idiosyncratic shock (z^a_{jo}), which has a Fréchet distribution, it follows that indirect utility for workers living in district i and working in district j also has a Fréchet distribution. Each worker chooses the commute that offers her the maximum utility, where the maximum of Fréchet distributed random variables is itself Fréchet distributed. Workers can choose to work at the home sector. The home sector has its own E^a_{home} and w^a_{home}. We assume that $d^a_{i,\text{home}} = e^{\kappa\tau_{i,\text{home}}+\delta^a(\ln C_{\text{home}})^2+\xi^a(\ln V_{\text{home}})^2} = e^0 = 1$ for 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 preferences 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 utility as determined by E^a_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

$$\pi^{\text{a,weekdays}}_{ij|i} = \frac{E^a_j(w^a_j/d^a_{ij})^\epsilon}{\sum_s E^a_s(w^a_s/d^a_{is})^\epsilon} \tag{5}$$

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

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

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:

$$H^a_{Ri} \equiv N^a_i - Q^a_i - D^a_i. \tag{6}$$

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:

$$\mathbb{E}[w_j^a|i] = \sum_i \frac{E_j^a (w_j^a/d_{ij}^a)^\epsilon}{\sum_s E_s^a (w_s/d_{is}^a)^\epsilon} w_j^a$$

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 factor prices and 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

$$\pi_{ij|i}^{\text{a,weekends}} = \frac{E_j^a \left(d_{ij}^a\right)^{-\epsilon}}{\sum_s E_s^a \left(d_{ij}^a\right)^{-\epsilon}} \tag{7}$$

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:

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

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

$$Y_j^a = A_j^a H_{Mj}^{\text{a,weekdays}} \tag{8}$$

As a result, $w_j^a = A_j^a$. The aggregate production in the economy is defined as:

$$Y \equiv \sum_{a,j} Y_j^a = \sum_{a,j} A_j^a H_{Mj}^{\text{a,weekdays}} \tag{9}$$

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 $\beta, \boldsymbol{\epsilon}, \boldsymbol{\kappa}, \boldsymbol{\delta}, \boldsymbol{\xi}$ and vectors of exogenous location characteristics $\boldsymbol{A}, \boldsymbol{E}, \boldsymbol{\tau}$, the general equilibrium of the model is referenced by the vectors $\boldsymbol{w}, \boldsymbol{\pi_M}, \boldsymbol{\tau}$ and total city population in each city H_i. These components of the equilibrium vector are determined by the following system of three equations: the workplace choice probability, profit maximization, and zero profits.

PROPOSITION 1. 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 $\boldsymbol{w}, \boldsymbol{\pi_M}$.

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 wage for each employment location be denoted as $\omega_j^a = \left(w_j^a\right)^{\epsilon} E_j^a$, which captures the wage and the Fréchet scale parameter for that location because these both affect the relative attractiveness of an employment location to workers.

PROPOSITION 2. Given known values for the parameters $\beta, \boldsymbol{\epsilon}, \boldsymbol{\kappa}, \boldsymbol{\delta}, \boldsymbol{\xi}$ 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 workplace j. Before the outbreak of the virus the gravity equation can be written as:

$$ln\pi_{ij} = -\nu\tau_{ij} + \theta_i + \theta_j + e_{ij} \tag{10}$$

where θ_i and θ_j are residence and workplace 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 $\nu = \epsilon\kappa$ 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. We augment the gravity equation with a stochastic error that captures measurement error in travel distances e_{ij}.

We estimate equation (10) using a linear fixed effects estimator. We find that the semi-elasticity of commuting with respect to travel time is 0.1765 which is statistically significant at the one percent level. This estimate implies that each additional kilometer of travel reduces the flow of commuters by around 1.7 percent. From the regression $\mathbb{R}^2$, this gravity equation specification explains around 70 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.

As in Ahlfeldt et al. (2015), we define adjusted wages as $\tilde{w}_j^a = w_j^a (E_j^a)^{1/\epsilon}$, which captures the wage and the Fréchet scale parameter for that location because both affect the relative attractiveness of an employment 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, the model's labor market clearing condition can be solved for a transformation of wages $\omega_j^a = (\tilde{w_j^a})^\epsilon = E_j^a (w_j^a)^\epsilon$ in each location using observed workplace employment (H_{Mj}^a), residence employment (H_{Ri}^a), and bilateral travel times (τ_{ij}):

$$H_{Mj}^a = \sum_{i=1}^{S} \frac{\omega_j^a / e^{\nu \tau_{ij}}}{\sum_{s=1}^{S} \omega_s^a / e^{\nu \tau_{ij}}} H_{Ri}^a \tag{11}$$

From the labor market clearing condition, transformed wages (ω_j^a) are determined independently of ϵ from workplace employment and travel times. Therefore, ϵ merely determines the monotonic transformation that maps transformed wages into adjusted wages and only scales the dispersion of log adjusted wages relative to the dispersion of log transformed wages: $\sigma^2_{\ln \tilde{w}_i^a} = \left(\frac{1}{\epsilon}\right)^2 \sigma^2_{\ln \omega_i^a}$.

Following Ahlfeldt et al. (2015), we jointly estimate mean parameters E_j^a and shape parameter ϵ using the following moment conditions:

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

$$\mathbb{E}\left[\left(\frac{1}{\epsilon}\right)^2 \sigma^2_{\ln \omega_i^a} - \sigma^2_{\ln w_i^a} \right] = 0$$

The first moment conditions requires that the total number of workers, belonging to

each age-group, corresponds to that in the data before the outbreak of the virus. The second moment condition requires that the variance of log transformed wages is equal to the variance of log wages in the data obtained from the Seoul Survey. The shape parameters ϵ are estimated to be 3.5427. 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).[7]

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] The estimated value for $\kappa = \nu/\epsilon$ is 0.0498.

7.3 Internally Calibrated Parameters

The rest of the parameters are calibrated internally by targeting 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. The data includes their contacts with previous confirmed cases, the district where they reside, and the districts that they visited before being quarantined.

We calibrate 10 parameters, β, d_I, $\delta^{a,\text{weekdays}}$, $\delta^{a,\text{weekends}}$, $\xi^{a,\text{weekdays}}$, and $\xi^{a,\text{weekends}}$ (for each age group), jointly using 10 moments from the data. First, we target total number of detected cases in Seoul (530) until April 6th, which captures the overall spread of the disease. Second, we target the fraction of 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). Lastly, we target the responsiveness of the commuting decisions in Seoul to information on the total cases at the district level and the total visits by infected individuals presented in Table 4. Specifically, we estimate the following regression in the data and in the model:

$$\log H^a_{Mj,t} = \alpha + \iota^a_1 \ln C_{j,t} + \iota^a_2 \left(\ln C_{j,t}\right)^2 + \varphi^a_1 \ln V_{j,t} + \varphi^a_2 \left(\ln V_{j,t}\right)^2 + \theta_j + \gamma_t + e^a_{j,t} \tag{12}$$

where $H_{Mj,t}$ denotes the total number of workers commuting to district j, $C_{j,t}$ denotes the total number of confirmed cases in district j at time t and $V_{j,t}$ is the total number of visits by infected individuals to district j prior to being quarantined. The regression includes both district effects, θ_j, and time effects, γ_t. We estimate equation (12) in our model separately for weekend and weekends and choose the parameters the minimize the distance between the coefficients in the data and those estimated in the data. The calibrated parameters are shown in Table 5.

Table 5: Parameter Values

Parameter	Value (young, old)	Definition/Target Moments
Externally Calibrated		
γ	1/18	Daily rate at which active cases recover.
τ^a	1/8.5, 1/10.2	Mean duration of hospitalization.
ψ^a	0.21%, 2.73%	Case fatality rate.
Internally Calibrated		
β	0.1758	Total # of detected case in two months.
d_I	0.0193	The fraction of undetected infections.
$\delta^{a,\text{weekdays}}$	0.0043, 0.0053	Semi-elasticity to confirmed cases (weekdays).
$\delta^{a,\text{weekends}}$	0.0040, 0.0064	Semi-elasticity to confirmed cases (weekends).
$\xi^{a,\text{weekdays}}$	0.00098, 0.000315	Semi-elasticity to visits by infected (weekdays).
$\xi^{a,\text{weekends}}$	0.00094, 0.000027	Semi-elasticity to visits by infected (weekends).

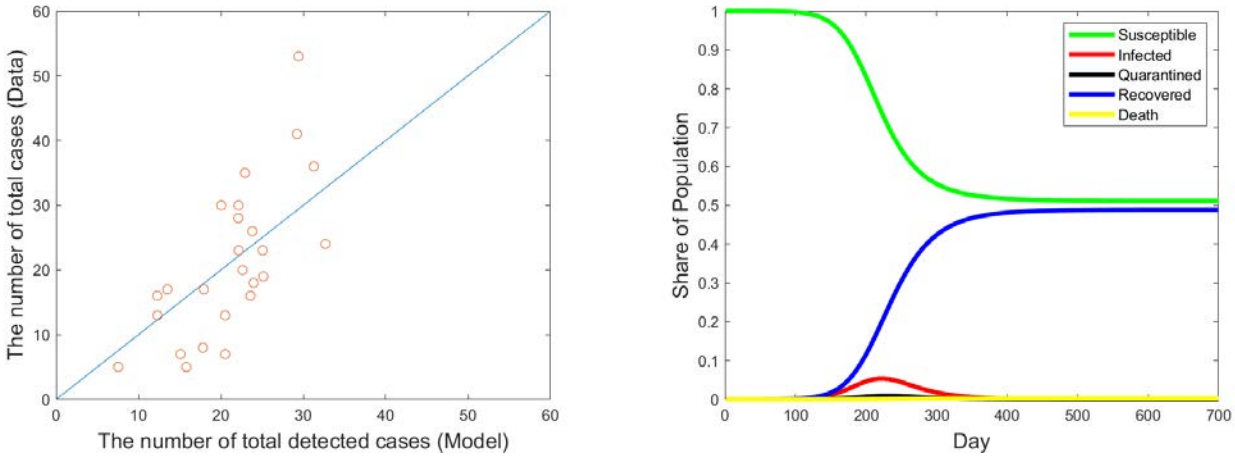
7.4 Simulation

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.

We simulate the model to evaluate its performance relative to the actual data of confirmed cases in Seoul. Panel (a) of Figure 3 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 April 6th, 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 3 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 half 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 8 months after the outbreak of the virus. Given the policies so far implemented in the city of Seoul, the peak of infection will involve less than 5% of the population at the same time. The fraction of confirmed and subsequently quarantined

cases is even lower as shown by the black line in the figure.

Figure 3: Predicted Spread of Disease

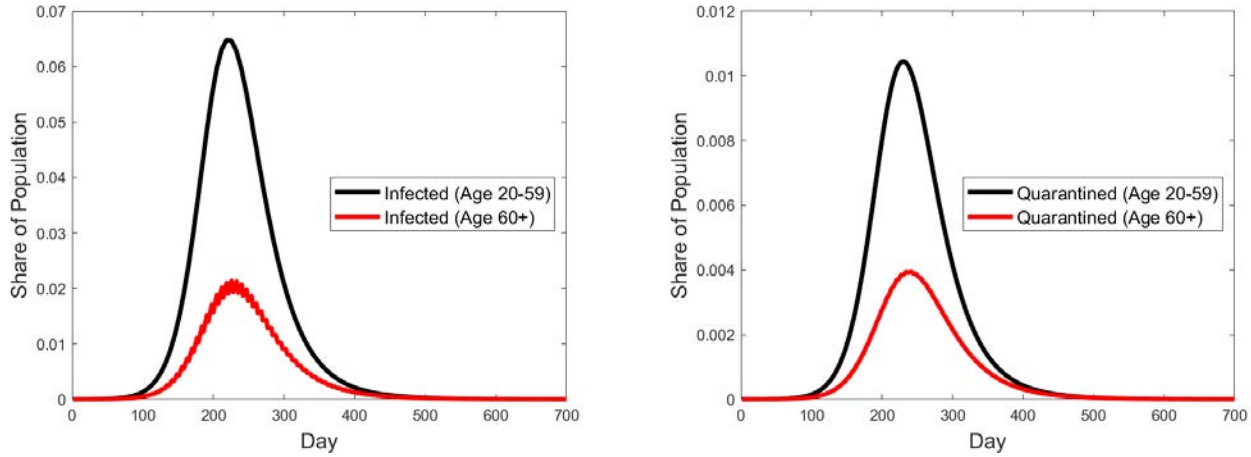


(a) Data vs Model (b) SIR Dynamics

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).

Figure 4 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 6% of the young and 2% of the old. Panel (b) shows that at the peak around 1% of the young and 0.4% 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 work at the home sector and they respond more drastically to the information related to the number of confirmed cases in each district.

Figure 4: Infected and Quarantined Population by Age Groups



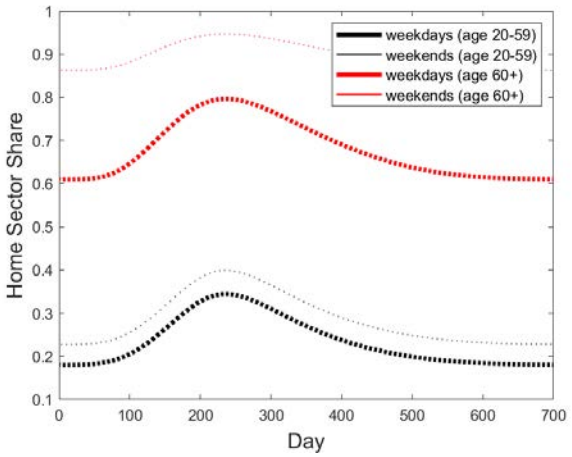
(a) Share of Infected People (b) Share of Quarantined People

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+).

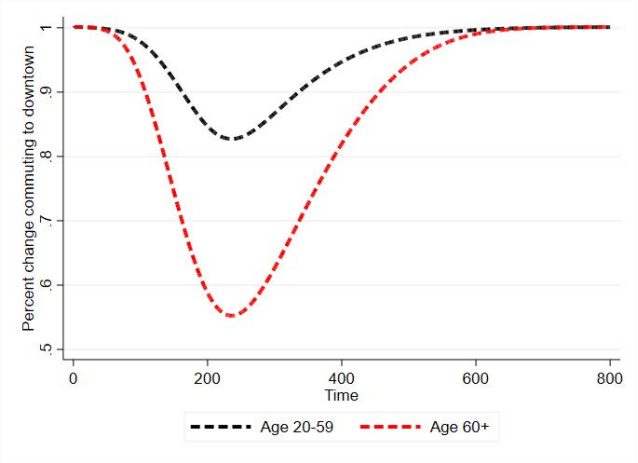
Figure 5 shows the commuting patterns of the people in Seoul according to the model. Panel (a) shows the fraction of the people that choose the home sector. From the data, we know that, before the outbreak of the virus, during weekdays approximately 18% of the young and 61% of the old stay at home during weekdays. Using the same amenity level across destinations, we predict that the home sector shares are approximately 33% for the young and 80% for the old during the weekends. The figure shows that as the virus spreads in the city, at the peak of the infections, approximately 32% of the young and 79% of the old will choose to stay at home during the weekdays.

Panel (b) of Figure 5 shows the change in commuting to downtown. We first compute the fraction of the total population that commutes to each of the districts that are part of downtown Seoul (i.e. Jung-gu, Jongno-gu, Gangnam-gu, Seocho-gu, and Yeongdeungpo-gu). We then normalize this fraction to one in each of the districts and average these shares across districts. The figure shows a large decrease in the fraction of people commuting to downtown at the infectious peak. Consistent with the foot traffic data, commuting is more sensitive to the disclosure of information during the weekends. The figure shows that commuting decreases approximately 18% for the young and 44% (age 20-59) for the old (age 60+) to downtown when the number of infected cases peaks. Importantly both panels show that the commuting patterns take a long time to recover; as long as there are confirmed cases of the virus, the commuting patterns of the city will be affected.

Figure 5: Home Sector Share and Commuting to Downtown



(a) Home Sector Share



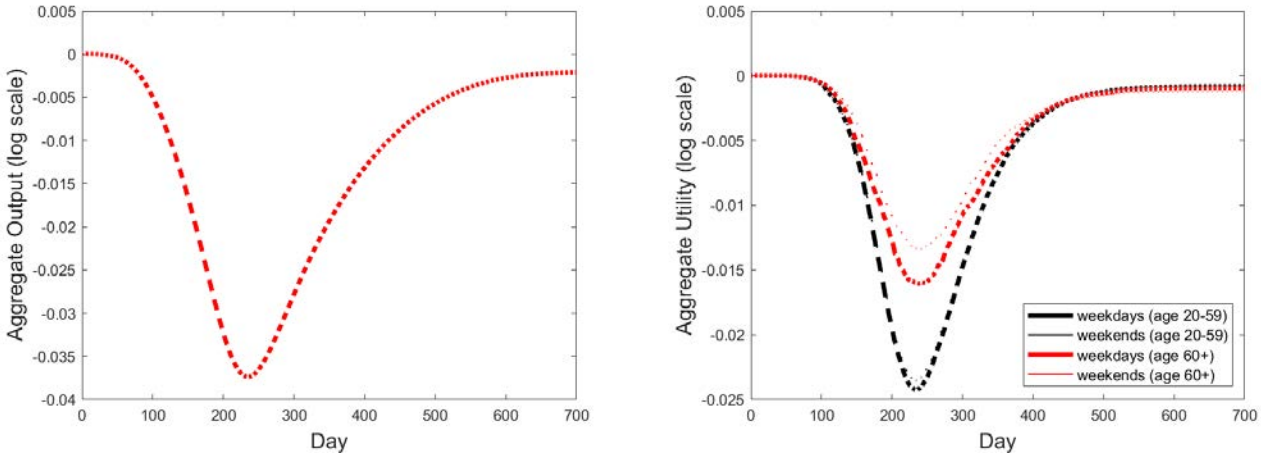
(b) Commuting to Downtown

Notes: The figure shows home sector share and commuting to downtown over time. Panel (a) shows the share of people that work choose to work in the home sector. The black line is the share for the young (age 20-59) and the red line is the share for the old (age 60+) Panel (b) shows the change in commuting to downtown.

Panel (a) of Figure 6 shows the predicted path of total output for the city of Seoul. The panel shows that according to the model output is predicted to fall substantially, up to 3.8% at the peak of infections. This is the results of both, people adjusting their commuting patterns and moving away from their preferred working locations, and also due to the assumption that quarantined people do not work (e.g. hospitalized). Similar forces affect the path of welfare depicted in Panel (b) of Figure 6. Importantly, even if in the model production only takes place during the weekdays, people adjust their commuting choices significantly during the weekends to avoid being exposed to the disease.

Both output and welfare take more than 2 years to recover and, in the current calibration, do not go back to the same levels observed before the outbreak of the virus. This is because the dead people are not contributing anymore to the output of the economy so that the total labor force is reduced and also because, as long as there are confirmed cases in the city, the patterns of commuting will be disrupted.

Figure 6: Aggregate Output and Welfare



(a) Aggregate Output (b) Aggregate Welfare

Notes: Panel (a) shows the level of output over time in the city of Seoul. Aggregate output is the sum of the output of each district. Panel (b) shows the level of welfare over time in the city of Seoul with public disclosure. Aggregate welfare is the sum of the utility of each worker. Black line represents the young (age 20-59) and red line represents the old (age 60+).

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.

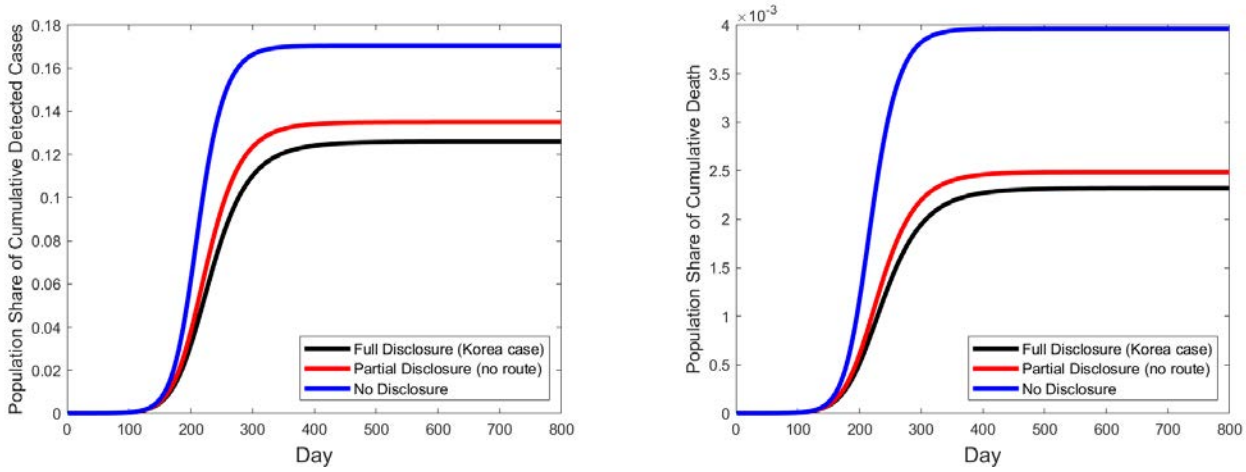
8.1 Information Disclosure

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 setting $\xi^a = 0$.

This assumes that detailed travel route information by infected individuals is not disclosed. Workers can still respond to the total number of confirmed cases in their districts. The case without information disclosure is that in which $\delta^a = \xi^a = 0$. In this case, the gravity equation depends only on the physical distance from origin to destination as it is commonly assumed in the literature.

Figure 7 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 twice 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 7: 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, the average daily output loss, and the welfare losses over two years. For the young, the total number of death increases by 7 percent under partial disclosure, and 27 percent under no disclosure. For the old, the total number of death increases by 7 percent under partial disclosure, and

98 percent under no disclosure.

The public health benefits from information disclosure come at the cost of output and welfare losses. We calculate the average daily output losses and average daily welfare losses relative to the period before the outbreak of the virus. Under full disclosure, the average daily output loss is 1.21 percent over two years and the average daily welfare loss is 0.56 percent for the young (age 20-59) and 0.41 for the old (age 60+). Average daily output loss decreases by 0.15 percentage points (from 1.21 percent to 1.06 percent) under partial disclosure and 0.77 percentage points under no disclosure. The average daily welfare loss for the young (old) decreases by 0.17 (0.11) percentage points under partial disclosure and 0.33 (0.22) percentage points under no disclosure. Under information disclosure, the welfare losses are substantially smaller than the output losses; 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, Output and Welfare

	Original Model	Partial Disclosure	No Disclosure
Total # of Cases	971,924	1,041,433	1,314,189
Total # of Death (age 20-59)	7,083	7,591	9,048
Total # of Death (age 60+)	10,819	11,581	21,518
Output Loss per day*	1.21	1.06	0.44
Welfare Loss per day* (age 20-59)	0.56	0.39	0.23
Welfare Loss per day* (age 60+)	0.41	0.30	0.19

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

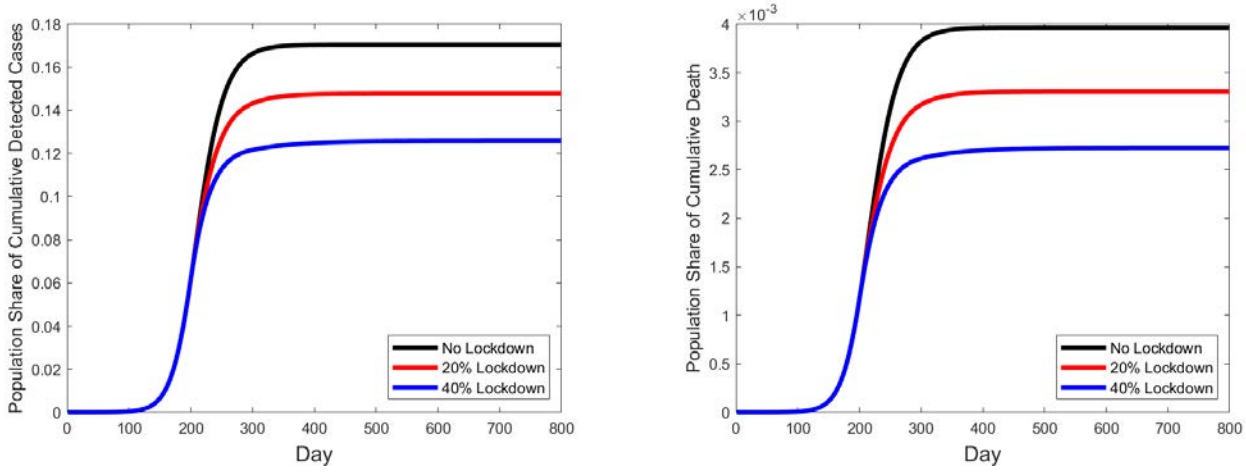
8.2 Lockdown

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 populations 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 200 to 300, when the spread of disease is the fastest when no information is disclosed.

Figure 8 shows the total number of detected cases from day 0 to day 800 when a lockdown policy is implemented from day 200 to 300.[8] We implement lockdowns in which 20% and 40% 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 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 8: 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) 20% lockdown and (iii) 40% 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, the daily average output losses and the welfare losses over two years under the different lockdown policies. Under 40% lockdown, the total number of cases is comparable to the model with

[8] 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.

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, Output and Welfare

	No Lockdown	20% Lockdown	40% Lockdown
Total # of Cases	1,314,189	1,139,916	971,077
Total # of Death (age 20-59)	9,048	7,932	6,815
Total # of Death (age 60+)	21,518	17,572	14,216
Output Loss per day*	0.44	1.15	1.63
Welfare Loss per day* (age 20-59)	0.23	0.88	1.68
Welfare Loss per day* (age 60+)	0.19	0.23	0.27

Notes: The table reports the total number of detected cases, the total number of death, the average daily output losses, and the welfare losses over two years in the city of Seoul under no lockdown, 20% lockdown, an 40% lockdown. Note that the first column is the same as the last column in table 6. The rows with * are shown in percent.

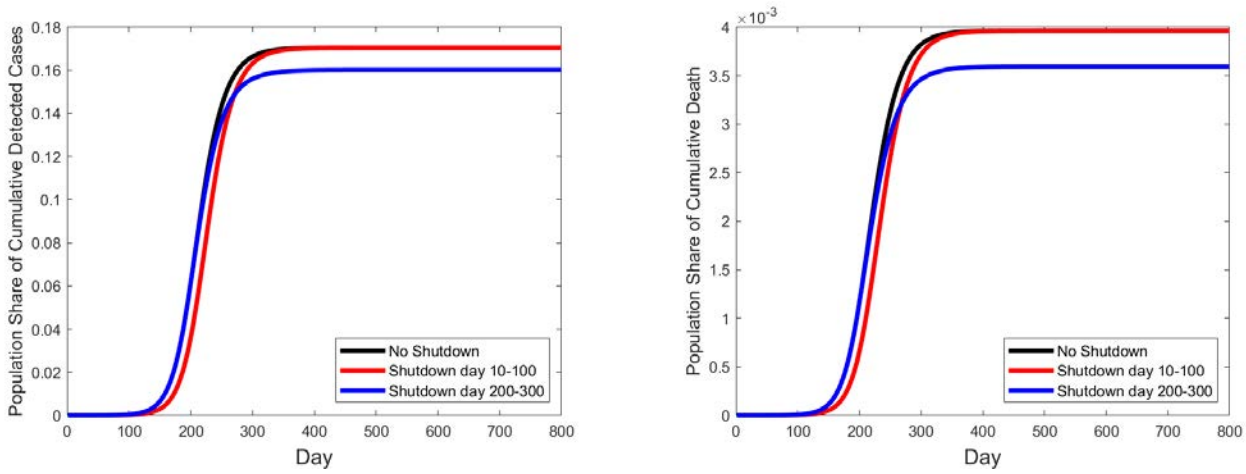
8.3 Shutdown

Lastly, we impose a shutdown policy in the model. The initial four 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 9 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 200 to 300. 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 9: 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, the average daily output losses 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 output loss increases by 0.1 percentage points and the welfare losses increase by 0.77 percentage points for the young and 0.35 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, Output and Welfare

	No Shutdown	Shutdown day 10-100	Shutdown day 200-300
Total # of Cases	1,314,189	1,314,097	1,235,337
Total # of Death (age 20-59)	9,048	9,048	8,588
Total # of Death (age 60+)	21,518	21,515	19,145
Output Loss per day*	0.44	0.54	0.52
Welfare Loss per day* (age 20-59)	0.23	1.00	1.07
Welfare Loss per day* (age 60+)	0.19	0.54	0.56

Notes: The table reports the total number of detected cases, the total number of death, the average daily output losses, 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 200 to 300. 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 rows with * are shown in percent.

9 Conclusion

This paper uses a 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 400 thousand and deaths by 13 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 50 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.

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.

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.