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

David Argente, Chang-Tai Hsieh, and Munseob Lee*

Abstract—South Korea publicly disclosed detailed location information of individuals who tested positive for COVID-19. We quantify the effect of public disclosure on the transmission of the virus and economic losses in Seoul. The change in commuting patterns due to public disclosure lowers the number of cases by 60,000 and the number of deaths by 2,000 in Seoul over two years. Compared to a city-wide lockdown that results in the same number of cases over two years as the disclosure scenario, the economic cost of such a lockdown is almost four times higher.

I. Introduction

ON January 30, 2020, residents of Jungnang district in northeast Seoul received the following message on their cell phone about the second person in South Korea who had tested positive for COVID: "Korean male, born 1987, living in Jungnang district. Confirmed on January 30, Hospitalized in Seoul Medical Center." The text went on to disclose the whereabouts of the individual in the past few days:

- January 24: Return trip from Wuhan without symptoms.
- January 26: Merchandise store* at Seongdong by subway at 12 pm; massage spa* by subway in afternoon; two convenience stores* and two supermarkets*.[1]
- January 27: Restaurant* and two supermarkets* in afternoon.
- January 28: Hair salon* in Seongbuk; supermarket* and restaurant* in Jungnang by bus; wedding shop in Gangnam by subway; home by subway.
- January 29: Tested at hospital in Jungnang.

Over the following weeks, as more people were tested for COVID, South Koreans received similar texts for every relevant patient who tested positive. This information was widely disseminated on websites and incorporated into mobile phone applications.

This paper uses South Korea's experience to show that public disclosure can be an important tool to combat the spread of a virus. Since a testing regime is likely to catch only a fraction of infected individuals, disclosure of locations of people who have tested positive can help noninfected people avoid places where they are more likely to be in contact with infected people who have not yet been detected. For example, on March 30, local authorities in Seoul disclosed that a patient visited a coffee shop in Mapo district on March 28. Nobody visited the coffee shop on May 30.[2] If the person who visited the coffee shop on March 28 infected others in the vicinity, this behavior by the public can help reduce the probability that an undetected infected person spreads the virus.

Figure 1 shows that the response of the public to the disclosure of the visit by the infected person to the coffee shop in Mapo district may hold more generally. The figure uses daily data on commuting flows between neighborhoods in Seoul from South Korea's largest mobile phone company.[3] It shows the mean and confidence bands of the change in the number of commuters from other districts of Seoul, relative to the first week for February, in each of the 25 districts in Seoul during the weekdays. We highlight two facts. First, commuter inflows in the average district in Seoul decreased by about 5% in March and April and recovered by the end of May. Second, there is substantial heterogeneity across districts in the change in commuting inflows. At the peak of the pandemic in early March, inflows fell by 3% in some districts to about 12% in other districts. In late May when inflows in the average district were back to "normal," inflows into some districts appear to be permanently lower, while in other districts, the reverse is true. We show that commuting inflows fell by more in districts where a large number of COVID patients lived or visited compared to districts with few COVID patients.

We then quantify the effect of the change in commuting flows in Seoul shown in figure 1 on the transmission of the virus. We use a susceptible-infected-recovered (SIR) model where the virus spreads as susceptible people in a neighborhood commute and come into contact with infected people from other neighborhoods. We show that the change in commuting flows observed in mobile phone location data predicts the heterogeneous spread of the virus across neighborhoods in Seoul. We also project the effect of public disclosure on the transmission of the virus over the next two years. Relative to a scenario where there are no changes in commuting patterns, public disclosure of information lowers the number of COVID patients by 60,000 over two years.

We also endogenize the commuting flows in the SIR model. In the model, the flow of people across neighborhoods

Received for publication October 6, 2020. Revision accepted for publication April 26, 2021. Editor: Amit K. Khandelwal.

*Argente: Pennsylvania State University; Hsieh: University of Chicago; Lee: University of California, San Diego.

We thank Fernando Alvarez, Jingting Fan, David Lagakos, Marc-Andreas Muendler, Valerie A. Ramey, and Nick Tsivanidis for helpful comments. We use proprietary data from SK Telecom and thank Geovision at SK Telecom and Brian Kim for their assistance with the data.

A supplemental appendix is available online at https://doi.org/10.1162/rest_a_01095.

[1] An asterisk * indicates that the establishment's name was disclosed.

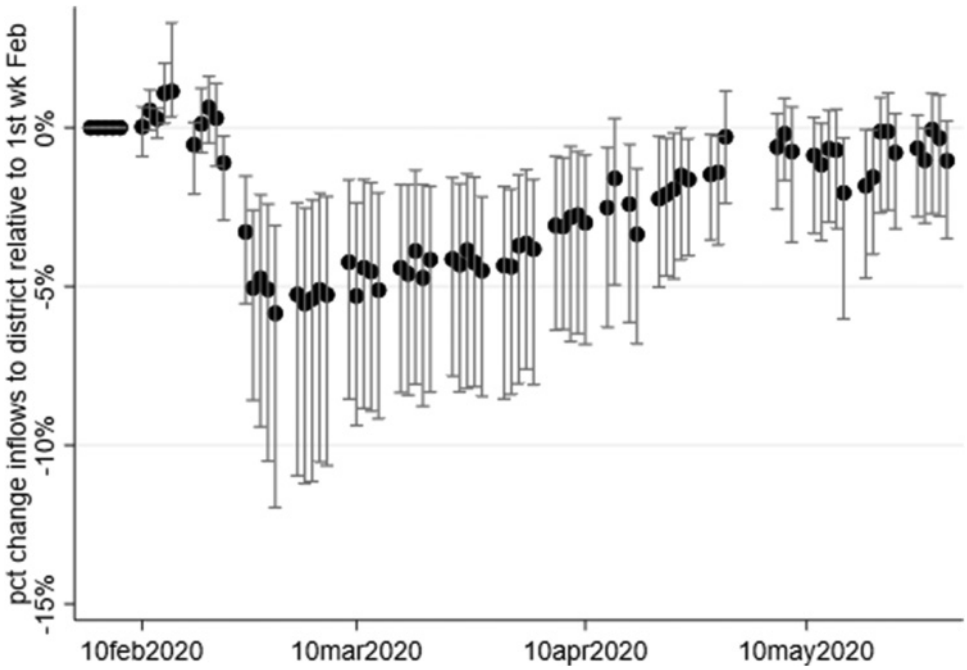
[2] "Seoul's Radical Experiment in Digital Contact Tracing," *New Yorker*, April 17, 2020.

[3] We describe the mobile phone data in section III.

The Review of Economics and Statistics, January 2022, 104(1): 176–186

https://doi.org/10.1162/rest_a_01095

FIGURE 1.—CHANGE IN WEEKDAY INFLOWS INTO DISTRICTS IN SEOUL



The figure shows the mean of the percent change in the inflow of people into each district of the city of Seoul, relative to the first week of February, calculated from SK Telecom's mobile phone data. Confidence bands indicate 90th and 10th percentiles.

generates economic gains from the optimal match of people with the place of work and leisure. We use the economic commuting model to estimate the economic losses from the disruption of commuting flows shown in figure 1. Over the next two years, the predicted disruption in commuting patterns under South Korea's current strategy will lower economic welfare an average 0.14% per day compared to a scenario with no changes in commuting flows.

We then compare South Korea's current strategy to a hypothetical lockdown that results in the same number of infections over the next two years. There are two advantages of a disclosure strategy relative to a lockdown. First, a lockdown does not discriminate among locations. All coffee shops are shut down, not only the ones visited by people who later tested positive for COVID. Second, a lockdown forces everybody into social isolation regardless of the cost the lockdown imposes on them. In contrast, in our model, people self-select into social distancing based on their perceived risks from COVID and costs of social distancing. As a consequence, under a lockdown, optimal commuting patterns are severely disrupted for a large number of people, and the cost of the disruption is about four times as large as in the disclosure scenario.

Our approach combines the SIR meta-population model, where movements of the population transmit the virus across space, with a quantitative model of internal city structure as in Ahlfeldt et al. (2015), where the flows of people across the city are endogenous. Recent papers that develop a similar model are Antràs, Redding, and Rossi-Hansberg (2020), Fajgelbaum et al. (forthcoming), and Cuñat and Zymek (2020). Our work differs in that in our model, commuting costs depend on the local information, and we use detailed mobile phone data to discipline the sensitivity of commuting flows to information about COVID cases. We do not study the optimal control problem as do recent papers on the COVID-19 pandemic by Alvarez, Argente, and Lippi (2021) and Farboodi, Jarosch, and Shimer (2020). Instead, we focus on comparing the policy of disclosing information to both a policy without information disclosure and a lockdown policy.

The paper proceeds as follows. Section II presents the SIR model with commuting choices. Section III discusses the data and how we calibrate the parameters of the model. In section IV, we compare our benchmark model with one with no disclosure and one with a lockdown. Section V concludes.

II. SIR Model with Endogenous Population Flows

In this section, we develop a model where the virus spreads across neighborhoods due to commuting flows and where the commuting flows are the endogenous outcome of individuals maximizing their utility by choosing where to work and to enjoy their leisure. Specifically, we adopt the canonical meta-population SIR model to analyze how the release of public information of COVID-19 cases affects the spread of the virus in Seoul.[4] Individuals live in different neighborhoods in Seoul, and the virus spreads when they commute to different neighborhoods, including the neighborhood where they reside. They can also choose to stay at home. In that

[4]See Keeling et al. (2010) and Keeling and Rohani (2011).

case, they do not interact with other people, and as a result, the virus does not spread.[5]

Specifically, there is an exogenous number of individuals in each neighborhood. Individuals are classified as susceptible (S—at risk of contracting the disease), infectious (I—capable of transmitting the disease), quarantined (Q—infected but quarantined and not transmitting the disease), and recovered (R—those who recover or die from the disease). We further differentiate individuals by age a and residential neighborhood i. Time t is defined as a day. We will show that commuting flows are different on weekdays rather than versus weekend. Because of this, we distinguish between the days that fall on weekdays versus weekends. The total number of nonquarantined residents of neighborhood i at time t of age a is $N_i^a(t) \equiv I_i^a(t) + S_i^a(t) + R_i^a(t)$.

The change in the number of infected residents of neighborhood i of age a at time t is

$$\Delta I_i^a(t) = \beta \sum_{j \neq \text{home}} \left[\frac{\sum_s \sum_a \pi_{sj}^a(t) I_s^a(t)}{\sum_s \sum_a \pi_{sj}^a(t) N_s^a(t)} \times \pi_{ij}^a(t) S_i^a(t) \right] - \gamma I_i^a(t) - d_I I_i^a(t). \quad (1)$$

The first term in equation (1) is the number of susceptible residents of i who get infected during the day, where the key endogenous variables are the commuting flows π_{ij}, which denote the share of residents of neighborhood i who commute to neighborhood j. In the absence of these flows, a susceptible person does not come into contact with an infected person. When a susceptible person comes into physical contact with an infected person, the "matching parameter" β denotes the probability the susceptible person gets infected.[6] The second and third terms in equation (1) are the infected who recover (or die) and the infected who are detected and quarantined, respectively, and the exogenous parameters are the recovery rate of infected people γ and the detection rate d_I.[7]

Next, we follow Ahlfeldt et al. (2015) to endogenize the commuting flows π_{ij} as the result of utility-maximizing commuting choices.[8] We assume individuals make commuting choices every day and distinguish between weekdays and weekends. Specifically, utility of a worker of age a who lives in i and works in j during the weekdays is $U_{ij}^a(t) = z_j^{a,wd}/d_{ij}^a(t)$ where $z_j^{a,wd}$ is idiosyncratic productivity from working in j on weekdays and $d_{ij}^a(t)$ is the cost of commuting from i to j. On weekends, the utility of the same person from commuting to j is $U_{ij}^a(t) = z_j^{a,wn}/d_{ij}^a(t)$ where $z_j^{a,wn}$ denotes idiosyncratic preferences from leisure in neighborhood j during the weekends.

In the absence of a pandemic, we assume that commuting costs depend only on the travel distance between i and j. We incorporate the effect of disclosure of COVID cases on commuting flows during the pandemic by assuming that the cost of commuting to neighborhood j also depends on information about the number of infected individuals in that neighborhood. Specifically,

$$\ln d_{ij}^a(t) = \kappa \tau_{ij} + \delta^a \ln C_j(t) + \xi^a \ln V_j(t) + \zeta^a(t). \quad (2)$$

Here τ_{ij} denotes travel distance between i and j, $C_j(t)$ is the number of residents of j confirmed as COVID patients in the two weeks prior to time t, $V_j(t)$ is the number of visits by confirmed COVID patients to neighborhood j in the two weeks prior to t, and $\zeta^a(t)$ captures the change in commuting costs that are independent of destination-specific information. For every infected person who is detected, we know the exact location of the places she visited the days before the infection was confirmed. We count each distinct place observed as a visit.[9] The sensitivity of the commuting cost to travel distance τ_{ij}, confirmed cases C_j, and visits V_j are governed by κ, δ^a, and ξ^a, respectively. When δ^a and ξ^a are positive, disclosure of COVID cases and visits in neighborhood j increases the cost of commuting to that neighborhood. Finally, the change in commuting costs due to forces that are orthogonal to detailed geographical information on detected individuals is captured by $\zeta^a(t)$.

An individual's commuting choice then boils down to a discrete choice problem. Given her idiosyncratic productivity, she chooses to work on weekdays in the neighborhood that maximizes her income net of the commuting cost. Similarly, given her idiosyncratic preferences, she chooses the neighborhood that maximizes her leisure net of the commuting cost on weekends. We further assume that a person's idiosyncratic productivity $z_j^{a,wd}$ is drawn from an i.i.d. Fréchet distribution with mean parameter $E_j^{a,wd}$ and dispersion parameter ϵ^{wd}. Similarly, her idiosyncratic preferences during the weekend are drawn from an i.i.d. Fréchet distribution with mean parameter $E_j^{a,wn}$ and dispersion parameter ϵ^{wn}. Note that $E_j^{a,wd}$ and $E_j^{a,wn}$ vary across neighborhoods, capturing mean differences in the quality of jobs and leisure activities across neighborhoods.

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

$$\pi_{ij}^{\text{a}}(t = weekday) = \frac{E_j^{a,wd} d_{ij}^a(t)^{-\epsilon^{wd}}}{\sum_s E_s^{a,wd} d_{is}^a(t)^{-\epsilon^{wd}}}. \quad (3)$$

[5] Individuals can get infected in the neighborhood where they live. In the context of our model, home can be seen as a different neighborhood where individuals do not interact with other people and where the virus does not spread.

[6] The home sector is one of the destinations. Once a susceptible person chooses to stay at home, she doesn't get infected. The summation over the destination neighborhoods in equation (1) excludes the home sector.

[7] The changes of the other state variables are $\Delta S_i^a(t) = -\beta \sum_{j \neq \text{home}} \left[\frac{\sum_s \sum_a \pi_{sj}^a(t) I_s^a(t)}{\sum_s \sum_a \pi_{sj}^a(t) N_s^a(t)} \times \pi_{ij}^a(t) S_i^a(t) \right]$, $\Delta Q_i^a(t) = d_I I_i^a(t) - \rho^a Q_i^a(t)$, $\Delta R_i^a(t) = \gamma I_i^a(t) + \rho^a Q_i^a(t)$, and $\Delta N_i^a(t) = N_i^a(t-1) - \Delta Q_i^a(t)$. $1/\rho^a$ is the amount of time a quarantined person is isolated.

[8] See also Monte, Redding, and Rossi-Hansberg (2018) and Tsivanidis (2019).

[9] South Korea discloses the information on all the places visited by every infected person who is detected.

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

$$\pi^{a}_{ij}(t = weekend) = \frac{E_j^{a,wn} d_{ij}^{a}(t)^{-\epsilon^{wn}}}{\sum_s E_s^{a,wn} d_{is}^{a}(t)^{-\epsilon^{wn}}}. \quad (4)$$

Equations (3) and (4) say that commuting flows to j from residents of i are increasing in $E_j^{a,wd}$ (during the weekday) and $E_j^{a,wn}$ (during the weekend) and decreasing in d_{ij}^{a} relative to other neighborhoods. More COVID cases in j relative to the other neighborhoods thus lower commuting flows to j. These commuting flows are the critical endogenous variables in the SIR model in equation (1) where a decline in π lowers the transmission of the virus throughout the city.

Finally, when individuals choose commutes that maximize their utility, expected utility of an individual living in neighborhood i is

$$\mathbb{E}[U_i^{a}(t = weekday)] = \Gamma\left(1 - 1/\epsilon^{wd}\right)\left(\sum_s E_s^{a,wd} d_{is}^{a}(t)^{-\epsilon^{wd}}\right)^{1/\epsilon^{wd}} \quad (5)$$

during the weekday and

$$\mathbb{E}[U_i^{a}(t = weekend)] = \Gamma\left(1 - 1/\epsilon^{wn}\right)\left(\sum_s E_s^{a,wn} d_{is}^{a}(t)^{-\epsilon^{wn}}\right)^{1/\epsilon^{wn}} \quad (6)$$

during the weekend where $\Gamma(\cdot)$ is a gamma function. Expected utility of residents of neighborhood i is a harmonic mean of the mean parameter of the Frechet distribution in all neighborhoods and decreasing in the cost of all accessing neighborhoods from location i. Expected welfare differs between residents of different neighborhoods depending on commuting costs from that neighborhood to all the other neighborhoods of the city. We refer to the weighted average of equations (5) and (6) across the residents in all the neighborhoods of Seoul as economic welfare.[10]

In summary, the total number of detected cases and visits that are publicly disclosed affect the commuting costs to each neighborhood, as specified in equation (2). The changes in commuting costs are reflected in the commuting flows (equations [3] and [4]) and affect not only the spread of the virus through their feedback into the SIR dynamics (equation [1]), but also individuals' economic welfare (equations [5] and [6]). The model highlights the important trade-offs of the public disclosure policy. The availability of information increases the commuting costs and lowers economic welfare, since it disrupts the optimal match of people with the place of work and leisure, but it also reduces the transmission of the virus across neighborhoods.

[10] See appendix A for a derivation of equations (3–6).

III. Model Calibration and Simulation

In this section, we describe the data we use and how we calibrated the parameters of the model in the previous section. We then simulate the effect of the disclosure of information on the transmission of the virus and economic losses in Seoul over the next two years.

A. *Calibration*

We infer the economic parameters from data on commuting flows. We measure commuting flows from proprietary data provided by the SK Telecom mobile phone company.[11] Based on the location of mobile phones, SK Telecom calculates daily bilateral flows of people by age group and gender in the 25 districts of Seoul.[12] We have data on daily bilateral commuting flows across Seoul's districts from January 2020 to May 2020. We also have the data on the monthly average of daily bilateral commuting flows from January 2019 to December 2019, which we use to calibrate parameters related to the prepandemic period.

The elasticity of commuting flows to commuting costs is ϵ^{wd} (weekdays) or ϵ^{wn} (weekends). Before the pandemic, the commuting cost depends only on distance τ_{ij} and the cost of distance κ. Specifically, the commuting flow equations (3) and (4) prior to the pandemic can be written as

$$\ln \pi_{ij}(t = weekday) = -\kappa\epsilon^{wd}\tau_{ij} + \theta_i + \theta_j, \quad (7)$$

$$\ln \pi_{ij}(t = weekend) = -\kappa\epsilon^{wn}\tau_{ij} + \theta_i + \theta_j, \quad (8)$$

where θ_i and θ_j are residence and commuting destination fixed effects and the elasticity of commuting flows with respect to distance is the product of ϵ and κ. Table 1 estimates equations (7) and (8) from data on bilateral commuting flows from SK Telecom in November 2019 (before the pandemic) and the distance (in kilometers) among the 25 districts in Seoul. Table 1 reports $\kappa\epsilon^{wd} = 0.1413$ for weekday commutes (column 1) and $\kappa\epsilon^{wn} = 0.1666$ during the weekends (column 2).[13] Commuting flows during the weekend are more sensitive to distance compared to commutes during the weekdays.

We next estimate ϵ^{wd} from the coefficient of variation in wages within each of the 25 districts in Seoul.[14] Specifically,

$$\frac{\text{Variance}}{\text{Mean}^2} = \frac{\Gamma(1 - \frac{2}{\epsilon^{wd}})}{\Gamma(1 - \frac{1}{\epsilon^{wd}})^2} - 1, \quad (9)$$

[11] SK Telecom has the largest share (42% in January 2020) of South Korea's mobile phone market.

[12] A person's movement is included when she stays in the origin district for more than two hours, commutes to another district, and stays in that district for more than two hours. If a person moves multiple times during the day, the data record the main movement. The data exclude daily flows between Seoul and locations outside Seoul.

[13] We get similar coefficients with the Poisson pseudo-maximum-likelihood (PPML) estimator. For example, $\kappa\epsilon^{wd}$ is estimated to be 0.1493.

[14] We use data from the 2018 Seoul Survey. This survey collects commuting and labor market information of household members over age 15 from 20,000 households in Seoul.

TABLE 1.—COMMUTING FLOW EQUATION ESTIMATION BEFORE AND DURING PANDEMIC

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

C_j is the number of COVID cases in j, V_j is the number of COVID visitors in j, and **weekend** is an indicator variable for a day that falls on a weekend. The table shows the results of estimating equations (7) and (8) (columns 1 and 2) and equation (10) (columns 3 and 4). The dependent variable in columns 1 and 2 is the monthly average of daily commuting probability from district i to district j in November 2019. The dependent variable in columns 3 and 4 is the change in the daily commuting probability from district i to district j from February 1 to May 31, 2020, relative to the first week of February 2020. Column 3 includes only people under the age of 60 and column 4 only those above 60 years of age.

where Γ is a gamma function. When we filter the within-district wage dispersion data through this equation, we get $\epsilon^{wd} = 4.1642$. This estimate of ϵ^{wd} combined with the estimate of the elasticity of commuting flows during the weekdays implies that $\kappa = 0.0339$. We then use this estimate of κ to infer $\epsilon^{wn} = 4.9144$ from the elasticity of commuting flows to distance during the weekends. Note that $\epsilon^{wn} > \epsilon^{wd}$, which says that the variance of a person's productivity across districts is larger than the variance of her preferences for leisure across districts. It also implies that the responsiveness of weekend flows to a given change in commuting costs will be larger than that of the weekday flows.

We next use the equation for the commuting flows during the pandemic to estimate how commuting costs change in response to disclosure of information on COVID cases. Figure 2 shows the variation across districts in Seoul in the number of COVID cases and visits that we use to estimate the elasticity of commuting costs to information. Panels a and b show the mean of the total number of confirmed cases for the past two weeks (C_j) and the total number of visits by confirmed cases for the past two weeks (V_j) in each district. In March, when weekly inflows declined the most (figure 1), the mean of the total cases increased up to around nine and the mean of the total visits increased up to around forty. The variation across the districts also increased when the mean increased. Confidence bands in both panels show the dispersion across districts. This variation allows individuals to switch destination as the prevalence of the virus increases over specific districts. Figure 2c shows the correlation between cases and visits. Each dot indicates total number of cases and total number of visits in each district for the past two weeks. Overall, the correlation between C_j and V_j is positive but clearly not equal to 1.

The elasticities of the weekday commuting flows with respect to the number of cases C_j and visits V_j are $\delta^a \epsilon^{wd}$ and $\xi^a \epsilon^{wd}$, respectively. The corresponding elasticities of the weekend flows to C_j and V_j are $\delta^a \epsilon^{wn}$ and $\xi^a \epsilon^{wn}$. Specifically, we estimate these elasticities from estimating the following commuting flow equation on data during the pandemic,

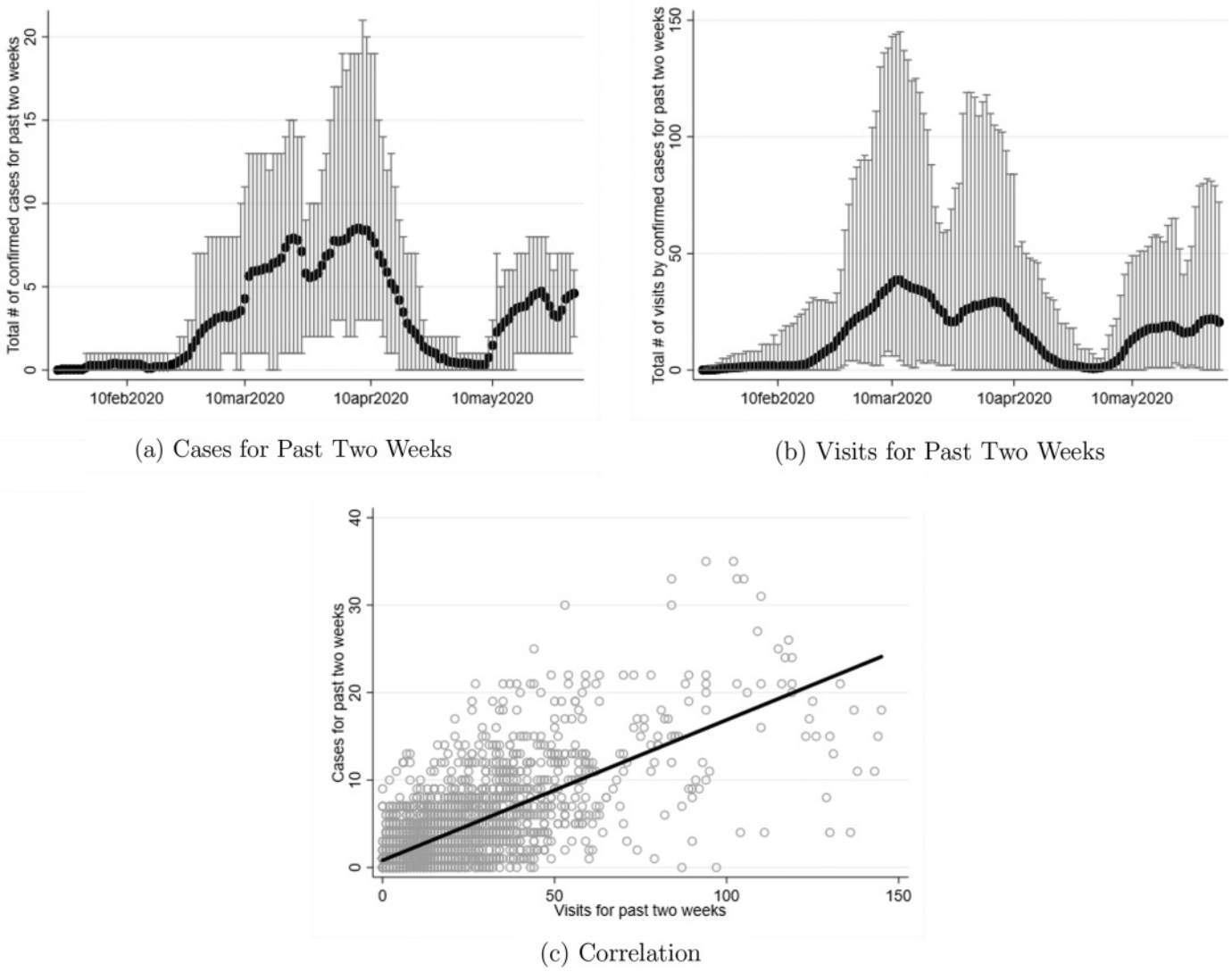
$$\begin{aligned}\Delta \ln \pi^a_{ij}(t) = {} & \delta^a \epsilon^{wd} \ln C_j(t) + \delta^a (\epsilon^{wn} - \epsilon^{wd}) \ln C_j(t) \\ & \times \textbf{weekend} + \xi^a \epsilon^{wd} \ln V_j(t) \\ & + \delta^a (\epsilon^{wn} - \epsilon^{wd}) \ln V_j(t) \times \textbf{weekend} + \varphi^a \\ & \times \textbf{weekend} + \theta^a_i + \lambda^a_j + \zeta^a(t), \qquad (10)\end{aligned}$$

where $\zeta^a(t)$ are the date fixed effects.

The dependent variable is the daily change in the commuting flows relative to the first week of February 2020 computed from SK Telecom's data and **weekend** is an indicator variable for a day that falls on a weekend.[15] The parameters in equation (10) are shown in columns 3 and 4 in table 1. Using the estimates of ϵ^{wn} and ϵ^{wd} derived earlier along with the elasticities of the commuting flows with respect to the disclosure of cases and visits in table 1, the elasticities of commuting costs with respect to the disclosure of information are $\delta = 0.00209$ and $\xi = 0.00138$ for people under age 60 and $\delta = 0.00247$ and $\xi = 0.00096$ for people over age 60.

[15] We account for serial correlation through two-way clustering, bootstrapped to account for the small number of districts.

FIGURE 2.—CASES AND VISITS IN EACH DISTRICT



(a) Cases for Past Two Weeks

(b) Visits for Past Two Weeks

(c) Correlation

(a) The mean of the total number of confirmed cases for past two weeks across 25 districts in Seoul. (b) The mean of the total number of visits by confirmed cases for past two weeks. Confidence bands indicate 90th and 10th percentiles. (c) The correlation between cases and visits for past two weeks.

We take three messages from the estimates of δ and ξ calculated from the commuting flows during the pandemic. First, δ and ξ are negative. This says that part of heterogeneity across neighborhoods in Seoul in the change in commuting inflows shown in figure 1 can be explained by the disclosure of COVID cases. Second, weekend commuting flows are more sensitive to COVID cases compared to weekday flows, perhaps because the comparative advantage across locations is weaker for leisure activities compared to work. This is consistent with the finding that in the prepandemic period, $\epsilon^{wn} > \epsilon^{wd}$. Third, commuting flows of those over the age of 60 are more sensitive to information on COVID cases. We interpret this as evidence that the perceived cost of getting infected is larger for the elderly compared to the young.

The estimated date fixed effects ($\zeta^a(t)$) are mostly significant. However, this information is only available until May 31. In order to simulate the aggregate commuting change after that period, we estimate the elasticity of the date fixed effects with respect to the total number of cases in Seoul,

$$\zeta^a(t) = \upsilon^a C(t) + \varepsilon^a(t),$$

where $C(t) = \sum_{j=1}^{25} C_j(t)$. We then use the estimated aggregate commuting change $\hat{\zeta}^a(t)$ to calculate the time-varying commuting cost in the simulation.

The last set of parameters of the commuting model age the mean parameters of the Fréchet distributions of productivity and leisure $E^{a,wd}$ and $E^{a,wn}$ estimated for each district and home sector. Using the estimates of κ, ϵ^{wn}, and ϵ^{wd}, we

estimate $E^{a,wd}$ and $E^{a,wn}$ from data on the commuting flows among the 25 districts of Seoul and the initial home sector shares.[16] We calculate the initial home sector shares from the Seoul Survey and Time Use Survey. During the weekdays, we find that 17% of people under age 60 and 61% of people over age 60 are not commuting but earning positive income. During the weekends, we assume that people who spend less than five hours on outside activities are at the home sector. The inferred shares are 46% for people with under age 60 and 67% for people over age 60.

Turning to the parameters of the SIR model, the rate (per day) at which infected people either recover or die is set to $\gamma = 1/18$, reflecting an estimated duration of illness of eighteen days.[17] We estimate ρ^a from the average duration of hospital care in Ferguson et al. (2020), which is eight days if critical care is not required and sixteen days if critical care is required. Ferguson et al. (2020) also estimate that 6.3% of those between the ages of 40 and 49 require critical care, whereas 27% of those between the ages of 60 and 69 require it. Using these estimates, we set $\rho^a = 1/8.5$ for the young and $\rho^a = 1/10.2$ for the old.

The remaining parameters in the SIR model are β and d_I. We calibrate these parameters internally using two moments from the data. First, we target the total number of detected cases in Seoul (861) until May 31, 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 ranges from 88.7% to 93.6%. We target a fraction of 90% undetected infections, which is also consistent with the estimates for the United States in Hortaçsu, Liu, and Schwieg (2021). The calibrated values of β and d_I are 0.1524 and 0.0163, respectively; we show sensitivity analysis to these values below.[18]

Finally, we also estimate the number of deaths from COVID. To do this, we set the fatality rate to 0.21% for the young and 2.73% for the old. We obtain these estimates from the Korean Centers for Disease Control and Prevention, which estimate fatality rates for the groups between 40–49 years of age and 60–69 years of age.

B. Simulation of SIR Commuting Model

We now simulate the SIR commuting model over two years assuming public disclosure of all COVID cases. Our initial conditions are the first 4 cases confirmed in the city of Seoul placed at the districts where the people infected reside. 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 who are undetected follow the predicted commuting patterns of the model. The first day of the simulation is January 30, 2020. In the data disclosed by local authorities, infected people who are detected report visiting two distinct districts on average in their entire travel logs.

We first evaluate the performance of the SIR model in terms of explaining the heterogeneity in the spread of the virus across neighborhoods in Seoul. Figure 3a plots the cumulative number of reported COVID cases in each district in Seoul by the end of May against the model's prediction of the number of cases in each district. 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. Figure 3b shows the dynamics of infected people that will take place in the upcoming months as predicted by the model. The figure shows that according to our model, infection increases up to around day 400 and decreases after that. The peak of infection will involve less than 6% of the population at the same time. Relative to the people below 60 years of age, there are fewer people above 60 years of age who are infected at a given time. This is because older people are more likely to stay in the home sector and respond more to information on the number of confirmed cases and visits in each district.

Figure 3c shows the change in inflows by district during weekdays and weekends. As disease spreads, inflows to each district decline. Before the pandemic, individuals working from home or from their residence location face the same commuting costs. However, during the pandemic, the commuting cost of working from their residence location increases as individuals are exposed to the spread of disease, while commuting costs of "work from home" stay at 0. Therefore, over time, the work-from-home option becomes more attractive. At the peak of disease, as the home sector share increases, average inflows are 12% lower during weekdays and 25% lower during weekends. Consistent with empirical evidence in figure 1, we also find significant heterogeneity across districts in the change in inflows in each district. This can be seen in the 95% confidence interval of the change in inflows into each district. Figure 3d shows the change in economic welfare, aggregating weekdays and weekends and residents of different neighborhoods of Seoul. Economic welfare declines because workers realize the cost of getting infected and change their commuting behavior. At the peak of disease, economic welfare declines by 0.27%.

IV. Welfare Effects of Disclosure

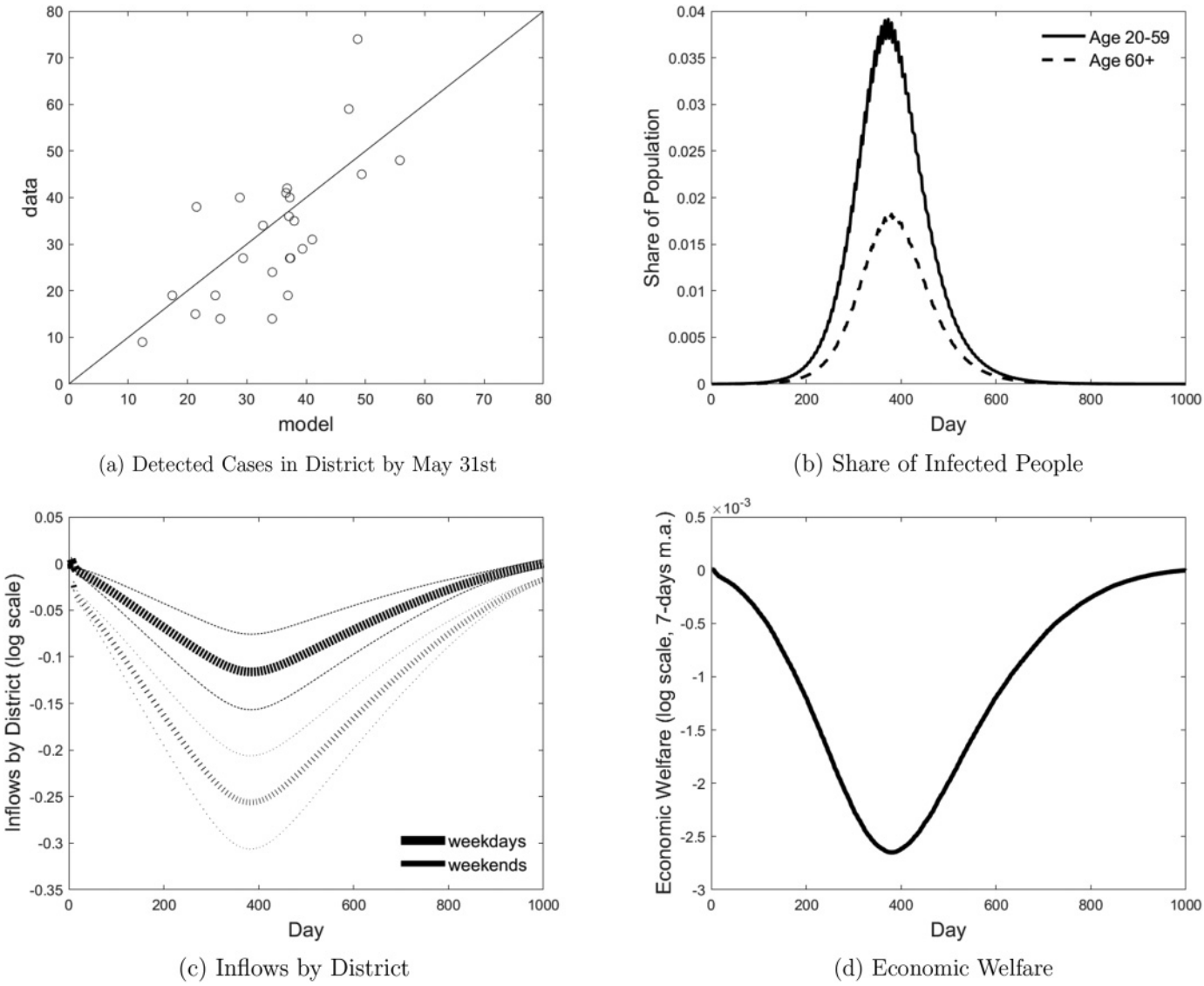
In this section, we evaluate the effectiveness of information disclosure and lockdown. First, to quantify the effectiveness of the disclosure policy, we simulate the model without allowing commuting to respond to destination-specific information. Still, we capture aggregate decline in commuting flows that could come from multiple sources such as recommended

[16] We normalized the geometric average of the location parameter $E^{a,wd}$ and $E^{a,wn}$ to 1.

[17] This is the value estimated from early evidence from COVID cases in China reported in Wang et al. (2020).

[18] There has not been large-scale randomized testing in South Korea. Song et al. (2020) conducted a small-scale seroprevalence study of COVID-19 for Daegu, the fourth largest city in Korea. The number of undiagnosed cases in Daegu is estimated to be a dozen times more than the number of confirmed cases based on PCR testing. Thus, in this case, the share of undetected infections is higher than 90%. Section IVA shows sensitivity analysis for parameter values consistent with this estimate.

FIGURE 3.—SIMULATED EFFECT OF DISCLOSURE ON COMMUTING, COVID CASES, AND ECONOMIC LOSSES



(a) Detected Cases in District by May 31st

(b) Share of Infected People

(c) Inflows by District

(d) Economic Welfare

(a) The total number of confirmed cases in the data (*y*-axis) and in the model (*x*-axis). Each dot is a district of Seoul. Note the 45 degree line. (b) The share of infected population for the young (ages 20–59) and the old (age 60+). (c) The change in inflows by district on weekdays and weekends, along with 95% confidence intervals. (d) The change in economic welfare as a seven-day moving average.

social distancing and closure of schools.[19] We call this the case without information disclosure. Second, we quantify the effectiveness of a lockdown policy, another widely used mitigation strategy, relative to the information disclosure case.

As described in the previous section, detailed information disclosure can be summarized into two components: (a) the total number of confirmed cases in each district for the past two weeks, $C_j(t)$, and (b) total number of visits by confirmed cases to each of the districts for the past two weeks, $V_j(t)$. We first examine the no-information-disclosure case, in which both $C_j(t)$ and $V_j(t)$ are not available. In this case, the commuting flow equation depends only on the physical distance from origin to destination. Then we evaluate a partial disclosure case, where either $C_j(t)$ or $V_j(t)$ is not disclosed; the partial disclosure case reported below presents the chained results of these two scenarios.

Table 2 reports the total number of detected cases, the total number of deaths, and the economic welfare losses over two years under different disclosure policies. Under partial and no information disclosure, we find more detected cases. The difference between the full information disclosure and partial or no disclosure scenarios is also significant when comparing the total number of deaths. The scenario with no disclosure of information yields 11% more deaths compared to the full disclosure scenario. This is because individuals above 60 years of age, who are more vulnerable to the virus, are also more

[19] Although there was no general lockdown in the city of Seoul (businesses, supermarkets and other retailers have remained open throughout the pandemic), there have been restrictions on schools and universities, which were closed soon after the outbreak, as well as mandatory social distancing.

TABLE 2.—COMPARISON OF FULL DISCLOSURE WITH NO DISCLOSURE AND LOCKDOWN

	No Disclosure	Partial Disclosure	Full Disclosure (Korea Case)	22% Lockdown Days 280 to 380
Total number of cases	840,709	811,032	780,907	780,692
Total number of deaths	20,744	19,739	18,743	20,488
Age 20–59	6,687	6,473	6,255	6,106
Age 60+	14,057	13,265	12,489	14,381
Welfare loss per day(%)	0.07	0.10	0.14	0.50
Age 20–59	0.07	0.10	0.13	0.64
Age 60+	0.08	0.12	0.16	0.07

The table reports the total number of detected cases, the total number of deaths, and the welfare losses over two years in the city of Seoul under no disclosure, partial disclosure, information disclosure (Korea case), and 22% lockdown from day 280 to 380. The economic welfare losses are shown in percent.

sensitive to the information disclosed and have altered their commuting patterns significantly in response.

The public health benefits from information disclosure come at the cost of economic welfare losses. We calculate average economic welfare loss per day. The daily economic welfare loss for the young (old) is 0.10% (0.12%) under partial disclosure and 0.13% (0.16%) under full disclosure. Under the partial or full disclosure, 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.

We impose a lockdown policy assuming that in this case, no information is disclosed. Under the lockdown policy, the standard mitigation strategy in most countries at the time South Korea implemented the disclosure policy, 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 stay in the home sector during weekdays and weekends. 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. We assume that the lockdown policy is implemented from day 280 to 380, when the spread of disease is the fastest.[20] To compare the disclosure policy and lockdown policy, we choose a 22% lockdown, which gives the same total number of cases over two years as the full information disclosure case.

Table 2 reports the total number of detected cases, the total number of deaths, and the economic welfare losses over two years under the lockdown policy. In this scenario, the total number of deaths is higher, especially among the old. The economic 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. 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 who have recovered from the disease as they are mandated to stay home under the lockdown, whereas under full information, those who have higher health risks are those who choose to stay at home.

[20] A lockdown implemented earlier delays infections and herd immunity, but has no impact on the total number of infected once herd immunity is achieved.

A. *Sensitivity Analysis*

We check the robustness of our results by checking their sensitivity to alternative values for the transmission rate β and the daily detection rate d_I. First, we run a simulation and a counterfactual without disclosure both increasing and decreasing β by 20%. Remember, however, that we choose β to match the total number of reported cases in Seoul by the end of May 2020, so changing β implies that the model no longer matches this data moment. Putting this aside, the top panel in table 3 shows the effect of changing β. When β is lower, the model delivers much lower numbers of cases and economic losses, and information disclosure lowers the number of COVID cases by 28%. However, higher β results in a higher number of cases and economic losses. Information disclosure is still effective but less so, reducing the number of cases by 5%.

We next examine the sensitivity of the results to the effectiveness of the testing regime. Specifically, we recalibrate 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 (0.1682) and d_I (0.0357) are calibrated to be higher than in our baseline calibration (0.1524 and 0.0163) so that they also jointly match the total number of cases by May 31.[21] 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 cases declines significantly. Thus, information disclosure is more effective in this case. When we target 95% of undetected infections, β (0.1515) and d_I (0.0076) are calibrated to be lower than the benchmark calibration. With a lower detection rate, information disclosure

[21] Although by construction our model replicates well the total number of cases in Seoul, it does not capture other measures taken in the city to successfully contain the virus in the months that followed, which include the closure of high-risk business (e.g., nightclubs, gyms), increases in the social distancing level, and the banning of gatherings of more than fifty people.

TABLE 3.—SENSITIVITY TO TRANSMISSION AND DETECTION RATES

	20% Lower $\beta = 0.1219$		20% Higher $\beta = 0.1829$	
	No Disclosure	Full Disclosure	No Disclosure	Full Disclosure
Total number of cases	81,314	58,384	1,143,903	1,090,291
Welfare loss per day(%)	0.04	0.06	0.05	0.11
	Fraction of Undetected = 0.8		Fraction of Undetected = 0.95	
	$\beta = 0.1682, d_I = 0.0357$		$\beta = 0.1515, d_I = 0.0076$	
	No Disclosure	Full Disclosure	No Disclosure	Full Disclosure
Total number of cases	907,202	776,173	565,072	538,609
Welfare loss per day(%)	0.08	0.15	0.05	0.11

The table reports the total number of detected cases and the economic welfare losses over two years in the city of Seoul under information disclosure and no disclosure with lower and higher β and with lower and higher fraction of undetected infections. For different fractions of undetected infection, β and d_I are recalibrated to match this new target moment, along with the number of total cases by May 31 (861). The economic welfare losses, compared to the no-disclosure case, are shown in percent.

is less effective because there is less information disclosed on the local cases and visits. So information disclosure is more effective when combined with testing that increases the share of the infected that are detected.

Finally, our estimate of the elasticity of commuting costs to information disclosure in table 1, combined with our baseline estimate of the detection rate d_I of 1.63%, implies that the effective transmission rate remains significantly above 1 until Seoul reaches herd immunity about 400 days after the pandemic starts (see figure 3). Put differently, information disclosure lowers the transmission of the virus, but not by enough to end the pandemic before herd immunity is reached. What would it take to bring the effective reproduction rate below 1 immediately? If the detection rate is $d_I = 0.0163$, we calculate that the sensitivity of commuting costs to information disclosure would have to be ten times larger than what we estimate in table 1 to immediately lower the effective reproduction rate below 1. If the detection rate is twice as large, so $d_l = 0.0326$, and the elasticity of commuting costs to information disclosure is only twice as large as what we estimated, this would be enough to immediately bring the effective reproduction rate below 1. The point here is that information disclosure helps but is not a panacea by itself. But combined with other measures, it can be a useful complement to end the pandemic.

Finally, we extend the model to introduce a switching cost paid by individuals adjusting their commuting decisions. This might be relevant particularly during weekdays (e.g., job-search frictions) and could lead us to underestimate the true commuting elasticities. Appendix C shows that a switching cost of 1% discount in utility, when individuals change commuting decisions, leads to 33% higher commuting elasticities. Relative to our benchmark model, results for this model are qualitatively similar. The model that includes switching costs reduces cases by 3 additional percentage points under the full disclosure scenario.

V. Conclusion

This paper uses an SIR model with multiple subpopulations and an economic model of commuting choice between the subpopulations 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 60,000 and deaths by 2,000 in Seoul over two 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 72% lower.

We do not attempt to measure the cost of the loss of privacy from disclosure of COVID cases or other inefficiencies arising from the disclosure of information—one example, the revenue losses of restaurants and coffee shops visited by infected individuals. These costs, whenever such measures are available, can be weighted against the benefits of public disclosure we provide here.

In our analysis, the community (in Seoul) reaches herd immunity within the next two years. We assume no vaccine will be available in the next two years. The analysis will obviously be different, and the trade-offs 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 can be an effective way to reduce the transmission of the disease while minimizing the economic cost of social isolation. Public dissemination of information is one way to accomplish that, but there obviously can be more effective ways to target social distancing.

REFERENCES

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf, "The Economics of Density: Evidence from the Berlin Wall," *Econometrica* 83 (2015), 2127–2189. 10.3982/ECTA10876

Alvarez, Fernando E., David Argente, and Francesco Lippi, "A Simple Planning Problem for COVID-19 Lockdown, Testing and Tracing," *American Economic Review: Insights* 3 (2021), 367–382. 10.1257/aeri.20200201

Antràs, Pol, Stephen Redding, and Esteban Rossi-Hansberg, "Globalization and Pandemics," NBER working paper 27840 (2020).

Cuñat, Alejandro, and Robert Zymek, "The (Structural) Gravity of Epidemics," CESifo working paper 8295 (2020).

Fajgelbaum, Pablo, Amit Khandelwal, Wookun Kim, Cristiano Mantovani, and Edouard Schaal, "Optimal Lockdown in a Commuting Network," *American Economic Review: Insights* (forthcoming).

Farboodi, Maryam, Gregor Jarosch, and Robert Shimer, "Internal and External Effects of Social Distancing in a Pandemic," NBER working paper 27059 (2020).

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," MRC Centre for Global Disease Analysis (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," *Journal of Econometrics* 220:1 (2021), 106–129.

Keeling, Matt J., and Pejman Rohani, *Modeling Infectious Diseases in Humans and Animals* (Princeton, NJ: Princeton University Press, 2011).

Keeling, Matt J., Leon Danon, Matthew C. Vernon, and Thomas A House, "Individual Identity and Movement Networks for Disease Metapopulations," *Proceedings of the National Academy of Sciences* 107 (2010), 8866–8870. 10.1073/pnas.1000416107

Monte, Ferdinando, Stephen J., Redding, and Esteban Rossi-Hansberg, "Commuting, Migration, and Local Employment Elasticities," *American Economic Review* 108 (2018), 3855–90. 10.1257/aer .20151507

Song, Suk-Kyoon, Duk-Hee Lee, Jun-Ho Nam, Kyung-Tae Kim, Jung-Suk Do, Dae-Won Kang, Sang-Gyung Kim, and Myung-Rae Cho, "IgG Seroprevalence of COVID-19 among Individuals without a History of the Coronavirus Disease Infection in Daegu, Korea," *Journal of Korean Medical Science* 35 (2020).

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* (2020). MedRxiv.

Tsivanidis, Nick, "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogota's TransMilenio," UC Berkeley mimeograph (2019).

Wang, Huwen, Zezhou Wang, Yinqiao Dong, Ruijie Chang, Chen Xu, Xiaoyue Yu, Shuxian Zhang, Lhakpa Tsamlag, Meili Shang, Jinyan Huang et al., "Phase-Adjusted Estimation of the Number of Coronavirus Disease 2019 Cases in Wuhan, China," *Cell Discovery* 6 (2020), 1–8. 31934347