ReadMe for "The Cost of Privacy: Welfare Effects of the Disclosure of COVID-19 Cases"

by David Argente, Chang-Tai Hsieh, Munseob Lee

This replication package contains Stata codes and Matlab codes to replicate Figure 1, 2, 3 and Table 1, 2, 3 in the paper. Note that all Stata codes generating Figure 1, 2 and Table 1 are under the Data folder and all Matlab codes generating Figure 3 and Table 2, 3 are under the Model folder.

There are two proprietary data we used:

1. Inflow-outflow data
    a. This is obtained from the SK Telecom's Geovision (big data service platform). You can contact them either by phone (1599-6011 from Korea) or contact us menu on the website (http://www.geovision.co.kr/).
    b. We used daily inflow-outflow data across 25 districts in Seoul from January 2020 to May 2020. Those are file names provided by the Geovision:
        i. SEOUL_TO_SEOUL_202001
        ii. SEOUL_TO_SEOUL_202002
        iii. SEOUL_TO_SEOUL_202003
        iv. SEOUL_TO_SEOUL_202004
        v. SEOUL_TO_SEOUL_202005
    c. We used monthly inflow-outflow data (weekday and weekend separately) across 25 districts in Seoul in November 2019. Those are file names provided by the Geovision:
        i. 201911_H (weekend data)
        ii. 201911_W (weekday data)

2. Digitized visits data
    a. This is obtained from the DS4C (Data Science for COVID-19 in South Korea). You can contact them by visiting the website (https://www.kaggle.com/kimjihoo/coronavirusdataset). In late 2020, they decide to pull out digitized visits data from the website for privacy concern.
    b. We used digitized visits from February 2020 to May 2020.
        i. datasets_PatientRoute

Stata Program (under the Data folder)

Figure 1, 2, and Table 1 can be produced by changing the path and running the following five codes sequentially. All output files are already saved under the Output folder. We used the StataMP 15 under the Windows OS for final replication.

Codes

1. Step 01 - Data Cleaning.do
    a. This file cleans SK Telecom Inflow-Outflow data, COVID individual cases data, COVID individual visits data, and Seoul Survey.

2. Step 02 - Epsilon Kappa Nu Estimation.do
    a. This file calculates the following five parameters:
        i. Epsilon for weekday
        ii. Epsilon for weekend
        iii. Kappa
        iv. Nu for weekday
        v. Nu for weekend
    b. Nu estimates are reported in Columns 1 and 2 of Table 1.

3. Step 03 - Delta Xi Upsilon Estimation.do
    a. This file calculates the following six parameters:
        i. Delta for the young
        ii. Delta for the old
        iii. Xi for the young
        iv. Xi for the old
        v. Upsilon for the young
        vi. Upsilon for the old
    b. Gravity equation coefficients are reported in Columns 3 and 4 of Table 1.

4. Step 04 – Figure 1.do
    a. This file creates Figure1.png in the Output folder.

5. Step 05 – Figure 2.do
    a. This file creates Figure2a.png, Figure2b.png and Figure2c.png in the Output folder.

Input (under the Raw_Data folder)

Note that two proprietary data are not included in the replication package.

1. covid19_Seoul_asof_0531.dta
    a. Basic information about 861 confirmed cases from January to May 2020.

2. district_list.dta
    a. This file assigns numbers 1-25 for 25 districts. These numbers will be used instead to identify each district in the Stata and Matlab codes.

3. Seoul_Survey.dta
    a. Micro-data of 2018 Seoul Survey containing 42,991 individuals.

4. d_GIS.dta
    a. Distance between 25 districts in Seoul in kilometers calculated from the ArcGIS software.

Matlab Program (under the Model folder)

Figure 3 and Table 2, 3 can be produced by adding input, output, subcode folders to the path and running the following nine codes sequentially. All output figures and tables are saved under the Output folder. We used the Matlab R2020a under the Windows OS for final replication.

Codes

1. Calibrate_E.m
    a. This code calibrates mean parameters.
    b. This code calls
        i. Calibrate_omega_function_young_weekday.m
        ii. Calibrate_omega_function_young_weekend.m
        iii. Calibrate_omega_function_old_weekday.m
        iv. Calibrate_omega_function_old_weekend.m
    c. This code generates
        i. calibrated_E.mat

2. Calibrate_beta_d_I.m
    a. This code calibrates a transmission rate and a detection rate.
    b. This code calls
        i. Calibrate_beta_d_I_function.m
        ii. Calculate_pi_function.m
        iii. Calculate_pi_weekend_function.m
        iv. Calculate_V_function.m

3. Simulate.m
    a. This code simulates the economy for 1,000 days
        i. To compare different information disclosure strategy, run this code by changing lines 33-37.
    b. This code calls
        i. draw_frechet.m
        ii. Calculate_utility_per_capita_function.m
        iii. Calculate_utility_per_capita_weekend_function.m
    c. This code generates
        i. Figure3a.png, Figure3b.png, Figure3c.png, Figure3d.png
        ii. cnt_original.mat, cnt_delta1_0.mat, cnt_delta2_0.mat, cnt_delta1_0_delta2_0.mat

4. Counterfactual_lockdown.m
    a. This code simulates the economy for 1,000 days with a lockdown.
    b. This code calls
        i. Calculate_utility_per_capita_lockdown_function.m

        ii. Calculate_utility_per_capita_weekend_lockdown_function.m
    c. This code generates
        i. cnt_lockdown.mat

5. Counterfactual_table.m
    a. This code generates
        i. Table2.mat

6. Sensitivity_beta.m
    a. This code generates
        i. sen_low_beta_original.mat, sen_low_beta_delta1_0_delta2_0.mat, sen_high_beta_original.mat, sen_high_beta_delta1_0_delta2_0.mat
    b. To simulate low and high beta under different disclosure strategies, change line 8 and line 33-35.

7. Sensitivity_beta_table.m
    a. This code generates
        i. Table3a.mat

8. Sensitivity_detection.m
    a. This code generates
        i. sen_detection_low_moment_original.mat, sen_detection_low_moment_delta1_0_delta2_0.mat, sen_detection_high_moment_original.mat, sen_detection_high_moment_delta1_0_delta2_0.mat
    b. To simulate low and high detection rates under different disclosure strategies, change line 8-9 and line 33-35.

9. Sensitivity_detection_table.m
    a. This code generates
        i. Table3b.mat

Inputs (under the Input folder)

1. moments_from_Seoul_survey.mat
    a. This mat file contains spatial distribution by age group over 25 districts in Seoul plus home sector.

2. distance_GIS.mat
    a. This mat file contains physical distance across 25 districts in Seoul calculated from ArcGIS.

3. initial_infection.mat
    a. This mat file contains initial condition of the simulation; total four detected cases in four different districts.

4. infection_May_31st.mat
    a. This mat file contains actual number of cumulative total detected cases by May 31st.