Replication materials for "Measuring Commuting and Economic Activity inside Cities with Cell Phone Records"

Gabriel E. Kreindler and Yuhei Miyauchi

Date: June 12 2021

Packages necessary for replication

STATA: (Windows or macOS)
- Stata 16
- geodist
- winsor
- ppmlhdfe
- grc1leg (net install grc1leg,from(http://www.stata.com/users/vwiggins/))
- gtools (version 1.7.5 18Apr2020)
- ols_spatial_HAC.ado v3 2018 (from http://www.fight-entropy.com/2010/06/standard-error-adjustment-ols-for.html)
- estout
- coefplot

R: (Windows or macOS)
- R version 4.0.4
- ggplot2
- dplyr
- tidyr
- boot
- readstata13
- lfe
- knitr
- foreign
- readxl
- stargazer
- geosphere
- Hmisc
- pastecs
- FENmlm
- caTools
- glmnet

- glmnetUtils
- gbm
- randomForest
- zeallot

Setting the path to the replication folders

STATA
Set the $cellphone_root global (in Windows, in "C:\ado\profile.do" which runs each time Stata is opened). The global should point to the main replication folder

R
Set `BGDSLKCELLPHONE_DATA ` in `source(paste0(Sys.getenv("HOME"), "/Rinit/profile.R"))`. It should point to the main replication folder

Set `BGDSLKCELLPHONE_CODE_UTIL` in `source(paste0(Sys.getenv("HOME"), "/Rinit/profile.R"))`. It should point to the `ado` folder in the replication folder.

Data sets not included in replication package

The following data are not included in the replication package due to restrictions on sharing this data:
1. Individual-level cell phone transaction level data
2. Exact cell phone tower locations

Due to the second restriction, we anonymized the latitudes, longitudes, and distance to central business districts (CBDs) of each tower. Therefore, the results produced by the public data are slightly different from our paper, whenever we report Conley standard errors or when we control for distance to the CBD.

All code processing the raw underlying data is included in the replication package.

To obtain access to these restricted data, interested readers can contact:
1. For Sri Lanka data, the LIRNEasia think tank (https://lirneasia.net/dap)
2. For Bangladesh, the Shibasaki & Sekimoto research lab (http://shiba.iis.u-tokyo.ac.jp/home_en/)

Part 1: Coding

To code the data for analysis, please run the following scripts in sequence.

Note that the gravity analysis is included here (as some coding uses the estimated destination fixed effects).

0. Coding raw CDR data

*****Note: raw CDR input data for this folder not included in the public repository. Consequently, the scripts in this section cannot be run using the public replication package.*****
Folder: `0-code-raw-data`
Scripts in this folder: hadoop Java code to classify the raw CDR data into
(A) daily trips and
(B) home-work data.

Home Work Classification SLK

Run
`0-code-raw-data\SLK-coding\workHomeTower\WorkHomeTowerMonthly.java`
then
`0-code-raw-data\SLK-coding\workHomeTower\WorkHomeTowerMonthlyCombine.java`

Output:
- `data_raw_slk\flows-daily-trips\140910_tripsX.csv` for X=0,1,2

Daily Trips Classification SLK

Run
`0-code-raw-data\SLK-coding\tripsMinMax\TripsMinMax.java`
then
`0-code-raw-data\SLK-coding\tripsMinMax\TripsMinMaxCombine.java`

Output:
`data_raw_slk\flows-home-work\part-0000X.csv` for X=0,1,2

For Bangladesh, in each folder, there is a `.sh` script that executes all the Java and Hadoop files. The paths have to be correctly adjusted to execute the code properly.

Daily Trips Classification BGD

Scripts: `BGD-coding\daily_commuting_matrix`
This code: Construct the tower-pair and day level commuting matrix.
Output: `data_raw_bgd/flows-daily-trips/commuter_matrix_YYYY_MM/`

Scripts: `BGD-coding\daily_commuting_panel`
This code: Construct the user, tower-pair, and day level commuting matrix.
Output: `data_raw_bgd/flows-daily-trips/commuting_panel/`

Daily Trips and Home Work Classification BGD

Scripts: `BGD-coding\home_work_and_ML_covariates\`
This code: Construct the user-level home and work classification, as well as the covariates for machine learning
Output:
- `data_raw_bgd/flows-home-work/user_home_office_list.csv` (with `data_raw_bgd/home_work_panel/userid_table.csv` as converter between new user ID and original user ID)
- `data_raw_bgd/ML/tower_entropy.csv`, `data_raw_bgd/ML/tower_user_info.csv`: covariates for machine learning

1. Travel Time Coding

*****Note: the input and some of the output data for this folder not included in public repository because they contain tower coordinates*****
Folder: `1-code-travel-time`
Scripts in this folder: coding and interpolating travel time data collected from Google Maps.
(more details in `1-travel-time-coding/readme.md`)

Travel Time Coding SLK

Script: `1-travel-time-coding\process googlemap data before interpolation SLK.do` **(***cannot run***)**

Output: (in `data_coded_slk\travel-times\`)
-`all tower pair within 50km before interpolation.csv` (*****Not included*****)
-`random 90000 towe pair within 50km - google prediction before interpolation.csv` (*****Not included*****)

Script:
`1-travel-time-coding\googlemap_interpolate\src\interpolate\GoogleMapInterpolateSLK.java` **(***cannot run***)**
This will interpolate duration, duration_in_traffic, distance_in_traffic to all tower pairs with positive commuting flows within 50 km. The bandwidth is set to be 0.1 km.

Output: (in `data_coded_slk\travel-times\`)
-`all tower pair within 50 km after interpolation.csv` **(***Included***)**
- `all tower pair within 50 km after interpolation auxiliary.csv` **(***Not included***)**

Travel Time Coding BGD

Script: `1-travel-time-coding\process googlemap data before interpolation BGD.do` **(***cannot run***)**

Output (in `data_coded_bgd\travel-times\`):
-`all tower pair in Dhaka before interpolation.csv` **(***Not included***)**
-`random tower pair - google prediction before interpolation.csv` **(***Not included***)**

Script:
`1-travel-time-coding\googlemap_interpolate\src\interpolate\GoogleMapInterpolateBGD.java` **(***cannot run***)**
This will interpolate duration, duration_in_traffic, distance_in_traffic to all tower pairs with positive commuting flows within 50 km. Bandwidth is set to be 0.03 km due to higher density of towers than SLK.

Output (in `data_coded_bgd\travel-times\`):
-`all tower pair in Dhaka after interpolation.csv` **(***Included***)**
- `all tower pair in Dhaka after interpolation auxiliary.csv` **(***Not included***)**

2. Code Distances and Dates

Folder: `2-code-other`

Scripts in this folder: code holidays and hartals in Bangladesh, and geographic tower properties in both countries.

Holidays and hartal dates in Bangladesh

Script: `code-dates-bgd.do` **(***can run***)**
Uses Hartal date definitions from Ahsan and Iqbal (2015).
Output:
- `data_coded_bgd\other\dates_igc.dta` **(***Included***)**

Distance to CBD

Script: `code-distance-to-CBD.do` **(***cannot run***)**
In each city, compute the distance from each tower to the CBD.
*****Note: the output data in the public repository has random noise (normal with mean zero and SD 1km) added to the distance to the CBD. Results relying on distance to CBD may be slightly different compared to the paper.*****
Output
- `data_coded_bgd/other/dist2cbd.dta` **(***Included***)**
- `data_coded_slk/other/dist2cbd.dta` **(***Included***)**

3. Code Census Data

Folder: `3-code-census`
Scripts in this folder: code census data (education and income proxy based on PCA of housing characteristics) in both countries.

Code Census SLK

*****Note: input data not available in public repository in order to not disclose tower locations.*****
Script: `slk_census_education.do` **(***cannot run***)**
Output: `data_coded_slk/census/censuspop_tower_education.dta` **(***Included***)**

Script: `slk_census_pca.do` **(***can run***)**
Output: `data_coded_slk/census/censuspop_tower_allvars.dta` **(***Included***)**

Code Census BGD

*****Note: input data not available in public repository in order to not disclose tower locations.*****
Script: `bgd_census_pca.do` **(***cannot run***)**
Output: `data_coded_bgd/census/censuspop_tower_allvars.dta` (*****Included*****)

4. Coding Commuting Flows

Folder: `4-code-flows`
Scripts in this folder: prepare commuting flows between pairs of towers, adding travel time. There are two versions for each country: "home-work" based on the classified home and work towers for each user, and "daily trips" based on identified trips within each day. See section "1. Cell-Phone Data and Commuting Flows" in the paper for more details.

Code Commuting Flows BGD

Script: `code-bgd-flows-daily-trips.do` **(***cannot run***)**
Output:
`data_coded_bgd/flows/daily_trips_intermed_idlevel_2013-XX.dta` **(***Not included***)**
`data_coded_bgd/flows/daily_trips_odmatrix.dta` **(***Included***)**

Script: `code-bgd-flows-home-work.do` **(***cannot run***)**
Output: `data_coded_bgd/flows/home_work_odmatrix.dta` **(***Included***)**

Code Commuting Flows for Hartal Analysis BGD

Script: `code-bgd-flows-daily-trips-panel.do` **(***cannot run***)**
Code daily trips for the Hartal analysis. An observation is a unique user ID and date, with information about the origin and destination towers for the daily trip that day. (This includes "stationary" trips if origin=destination.)
Output:
`data_coded_bgd/flows/daily_trips_panel.dta` **(***Included***)**
`data_coded_bgd/flows/commuting_panel/commuting_panel_XX_X.dta` **(***Not included***)**

Code Commuting Flows SLK

Script: `code-slk-flows-daily-trips.do` **(***cannot run***)**
Output:
`data_coded_slk/flows/daily_trips_odmatrix.dta` **(***Included***)**

`data_coded_slk/flows/daily_trips_intermed_idlevel` (***Not included***)

Script: `code-slk-flows-home-work.do` (***cannot run***)
Output: `data_coded_slk/flows/home_work_odmatrix.dta` (***Included***)

Coding Commuting Flows for Gravity Analysis

Script: `code-gravity-flows.do` (***can run***)
Additional coding to commuting flows before running gravity analysis.
Output:
- `data_coded_slk/flows/daily_trips_odmatrix_gravity.dta` (***Included***)
- `data_coded_slk/flows/home_work_odmatrix_gravity.dta` (***Included***)
- `data_coded_bgd/flows/daily_trips_odmatrix_gravity.dta` (***Included***)
- `data_coded_bgd/flows/home_work_odmatrix_gravity.dta` (***Included***)

Script: `code-gravity-skills.do` (***can run***)
Additional coding to commuting flows by skill. Uses output above and census data on education.
Output:
- `data_coded_bgd/flows/home_work_odmatrix_2kills.dta` (***Included***)
- `data_coded_slk/flows/home_work_odmatrix_2kills.dta` (***Included***)

5. Code DHUTS survey data

Folder: `5-code-dhuts`
Script in this folder: Read raw DHUTS travel survey data, code income, occupation, education level, commuting zones

Script: `coding_raw_dhuts.R` (***cannot run***)
Output: `data_coded_bgd/dhuts/coded_dhuts.rds` (***Included***)

Script: `coding_dhuts_at_czones.Rmd` (***can run***)
Output: `data_coded_bgd/dhuts/...` (***Included***)

Script: `code-DHUTS.do` (***cannot run***)
Used in commuting validation (section 10) (***Included***)
Output: `data_coded_bgd/dhuts/coded_dhuts_czone_pairs.dta` (***Included***)

6. Code Features for Machine Learning Analysis

Folder: 6-code-ML
Script in this folder: Construct covariates that are used as inputs for the machine learning predictions
Script: `create_ML_covariates.R` **(***cannot run***)**
Output: data_coded_bgd/ML/covariates_df_ML.Rds (*****Included*****)

7. Gravity Analysis - Estimating Destination Fixed Effects

Folder: `7-analysis-gravity`
Scripts in this folder: Run gravity equations, generate and save destination fixed effects, and generate Table 1 and Table H4.

Gravity Equation Table 1

Script: `table_1.do` **(***can run***)**
Output destination fixed effects:
- `data_coded\dfe_bgd_home_work.csv` (*****Included*****)
- `data_coded\dfe_bgd_daily_trips.csv` (*****Included*****)
- `data_coded\dfe_bgd_skills.csv` (*****Included*****)
- `data_coded\dfe_slk_home_work.csv` (*****Included*****)
- `data_coded\dfe_slk_daily_trips.csv` (*****Included*****)
- `data_coded\dfe_slk_skills.csv` (*****Included*****)

Output tables:
- `tables\table_1\table_1_main.tex` (*****Included*****)
- `tables\table_C2\table_C2_col1.tex` (*****Included*****)

Figure 2 (Smooth Destination Fixed Effects)

*****Note: input file with tower coordinates not included in public release to not disclose tower locations. *****
Script: `dfe_smoothing_for_map.do` **(***Cannot run***)**
Output: `maps/dfe_bgd_home_work_smoothed.csv` **(***Not included***)**
Output: `maps/dfe_slk_home_work_smoothed.csv` **(***Not included***)**

Gravity Equation Robustness

Script: `table_H4.do` (***Can run***)
Output destination fixed effects:
- `data_coded\dfe_bgd_robust_close_towers.csv` (***Included***)
- `data_coded\dfe_bgd_robust_logvol.csv` (***Included***)
- `data_coded\dfe_bgd_robust_logvol_plus1.csv` (***Included***)
- `data_coded\dfe_bgd_robust_nonparam.csv` (***Included***)

- `data_coded\dfe_slk_robust_close_towers.csv` (***Included***)
- `data_coded\dfe_slk_robust_logvol.csv` (***Included***)
- `data_coded\dfe_slk_robust_logvol_plus1.csv` (***Included***)
- `data_coded\dfe_slk_robust_nonparam.csv` (***Included***)
- `data_coded\dfe_slk_robust_traffic.csv` (***Included***)

Output table:
- `tables\table_H4\table_H4.tex` (***Included***)

8. Coding of model-predicted income

- Folder: Coding model-predicted income at workplaces and residential locations from gravity equation estimates (from 7-analysis-gravity)

Workplace Income Coding

Script: `workplace_income_coding.Rmd` (***Can run***)
- Output: /data_coded_bgd/workplace_income/dhuts_…: predicted income aggregated at workplace locations (***Included***)

Residential Income Coding

Script: `residential_income_coding.Rmd` (***Can run***)
- Output: /data_coded/residential_income.Rdata: predicted residential income at the tower level (***Included***)

Analysis

Each folder described below can be run independently of others, provided that all coding blocks above have been run.

9. Descriptive Statistics of Cell Phone Data (Table H1)

Folder: `9-analysis-stats`
Script: `table_H1.do` (***Can run***)
Output table: `tables/table_H1/sample_size_stats.tex` (***Included***)

10. Validation of commuting flows from CDR data

Folder: `10-analysis-commuting-validation`
Scripts in this folder: compare commuting flows and residential populations from cell phone data with analogues from the household transportation survey DHUTS and with census data.

Table H2. Comparison of Commuting Flows from Survey Data and Cell Phone Data

Script: `code-daily-trips-odmatrix-DHUTS.do` (***Can run***)
Output: `data_coded_bgd/dhuts/daily_trips_odmatrix_dhuts.dta` (***Included***)

Script: `code-home-work-odmatrix-DHUTS.do` (***Can run***)
Output: `data_coded_bgd/dhuts/home_work_odmatrix_dhuts.dta` (***Included***)

Script: `code-prep_figure_H2a_table_H2.do` (***can run***)
Output: `data_coded_bgd/dhuts/merged_comparison.dta` (***Included***)

Script: `table_H2.do` (***can run***)
Output table: `tables/table_H2/comparison_dhuts_v0_hw.tex` (***Included***)

Figure H2. Commuting Flows from Survey Data and Cell Phone Data

Scripts: `figure_H2a.do` and `figure_H2b.do` (***can run***)
Output figures:
- `figures/figure_H2/figure_dhuts_comp_full_appendix` (***Included***)
- `figures/figure_H2/figure_bgd_comm_hw` (***Included***)
- `figures/figure_H2/figure_slk_comm_hw` (***Included***)

- `figures/figure_H2/figure_both_comm_hw` **(***Included***)**

Table H3. Comparison of Residential Population from Cell Phone Data and Population Census

*****This analysis uses tower coordinates for Conley SEs. The public version script performs analysis without Conley SEs and with random tower coordinates.*****
Script: `table_H3.do` **(***can run partiall (without Conley SEs)***)**
Output:
- `data_coded_bgd/census/table_H3_population_CDR_census` **(***Included***)**
- `data_coded_slk/census/table_H3_population_CDR_census` **(***Included***)**
Output table:
- `tables/table_H3/table_H3.tex` **(***Included***)**
(Also runs equations without Conley standard errors.)

11. Validation of model-predicted income

Folder: `11-analysis-income-validation`

Table 2 (panel A), Table H5, Table H6, Table D1: Income Validation at Workplaces

Script: `workplace_income_analysis.Rmd`**(***can run***)**
This file: workplace income validation
Output:
- Model prediction and survey data in Dhaka (Table H5) **(***Included***)**
- Robustness regression table (Table H6) **(***Included***)**
- Survey income under different assumptions about shocks and travel cost (Table D1) **(***Included***)**
- Raw correlation between model prediction and survey data in Dhaka (Table 2A) **(***Included***)**

Table 2 Panel B and Table C3: Workplace Income Validation by Skill

Script: `table_2_table_C3.do` **(***can run***)**
Notes:
- requires `data_coded/dfe_bgd_skills_MLE.csv` which is generated by `table_C2.do` (see section 12 below)

Output tables:

- `tables\table_2\table_2B.tex` (***Included***)
- `tables\table_C3\table_C3.tex` (***Included***)

Table D2 and Table H7: Workplace Income Validation at Individual Level

Script: `workplace_income_analysis_structural.Rmd` (***can run***)
This file: Workplace Income Validation Analysis with Different Assumptions of Shocks and Travel Costs (Appendix D)
Output:
- parameter estimates (Table D2) (***Included***)
- individual-level validation regression (Table H7) (***Included***)

Table 4A and Table H8: Residential Income Validation

Script: `residential_income_analysis.Rmd` (***can run***)
Output:
- Raw correlation between model prediction and survey data (Table 4A) (***Included***)
- Regression table (Table H8) (***Included***)

Table H9: Robustness of Residential Income Validation

Script: `residential_income_analysis_robustness.Rmd` (***can run***)
Output:
- Regression table for robustness (Table H9) (***Included***)

Table 3, Table 4B and Table F1: Comparison with Machine Learning Predictions

Script: `residential_income_analysis_ML.Rmd` (***can run***)
Output
- ML vs model prediction for workplace income (Table 3) (***Included***)
- ML vs model prediction for residential income (Table 4B) (***Included***)
- Robustness to different tuning parameters (Table F1) (***Included***)

12. Analysis of Model with Skills

Folder: `12-analysis-skills`
Scripts in this folder: estimate model with skill heterogeneity and perform income validation with (skill-specific) destination fixed effects.

Table C1. Numerical Simulation to Check Estimation Procedures

Script: `simulation2skills.do` (***Can run***)
Output table: `tables\table_C1\simulation_gravity_main.tex` (***Included***)

Table C2. Gravity Equation with Skills: MLE Estimation

Script: `table_C2.do` (***Can run***)
Output:
- `tables\table_C2\table_C2_col2.tex` (***Included***)
- `tables\table_C2\table_C2_col4.tex` (***Included***)

Note: Columns 1 and 3 of Table C2 are identical to columns 3 and 6 in Table 1 and are generated in `7-analysis-gravity\table_1.do`

13. Hartal Analysis

Note: raw microdata used to run the hartal analysis is not available in the public repository due to its sensitive nature. Only tower- or tower-pair level aggregate commuting data from the cell phone data is available.
Folder: `13-analysis-hartal`
Scripts in this folder: additional coding and analysis for hartal section.

Coding

Script: `code-home-work-idlevel.do` (***Cannot run***)
Output: `data_coded_bgd\flows\home_work_idlevel.dta` (***Not included***)

Script: `code-daily-trips-panel-hartal-part1.do` (***Cannot run***)
Output: `data_coded_bgd\flows\daily_trips_panel_hartal` (***Not included***)

Script: `code-daily-trips-panel-hartal-part2.do` (***Cannot run***)
Output: `data_coded_bgd\flows\daily_trips_panel_hartal_coded` (***Not included***)

Analysis

Script: `table_5.do` **(***Cannot run***)**
Output table: `tables\table_5\main_table_heterogeneity_5.tex`
(*Included***)**

Script: `table_G1_hartal_frequent_caller_sample.do` **(***Cannot run***)**
Output table: `tables\table_5\main_table_heterogeneity_G1.tex`
(*Included***)**

Script: `figure_G1.do` **(***Can run***)**
Output figure: `figures/figure_G1/figure_G1_hartal_event_TW` **(***Included***)**

Script: `figure_G2.do` **(***Can run***)**
Output figure:
- `figures/figure_G2/figure_G2_hartal_dates_TW_novdec` **(***Included***)**
- `figures/figure_G2/figure_G2_hartal_dates_TW_augsep` **(***Included***)**

14. Other Analysis and Robustness

Folder: `14-analysis-robustness`
Scripts in this folder: run gravity and/or income validation with various assumptions.

Table E1. Gravity overidentification and validation

Script: `table_E1_gravity_overid.do` **(***Can run***)**
Output table:
- `tables\table_E1\table_E1_panel_A_exact_sample_size.tex` **(***Included***)**
- `tables\table_E1\table_E1_panel_B.tex` **(***Included***)**

Figures H3 and H4

Script: `figure_H3H4.do` **(***Can run***)**
Output figures:
- `figures\figure_H3H4\figure_H3_r2_dist_both` **(***Included***)**
- `figures\figure_H3H4\figure_H4_r2_popden_both` **(***Included***)**

Figures H5

First run
Script: `figure_H5_code_grids.do` (***cannot run (uses tower coordinates)***)
Output:
- `data_coded_slk/other/tower_grid_cells_destination.dta` (***Included***)
- `data_coded_bgd/other/tower_grid_cells_destination.dta` (***Included***)
Note: input file with tower coordinates not included in public release to not disclose tower locations.

Script:
- `figure_H5_code_aggregate_robustness.do` (***Can run***)
- `figure_H5.do` (***Can run***)
Output figure:
- `figures/figure_H5/figure_H5_aggregation` (***Included***)

Code to generate each figure and table in the paper

Table 1
Title: Gravity Equation Estimation Results
Code: `7-analysis-gravity\table_1.do`

Figure 1
Title: Estimated log Wages in Dhaka and Colombo
Code: uses output from `7-analysis-gravity\dfe_smoothing_for_map.do`
Note: cannot be replicated with data in public relieve to not disclose tower locations

Table 2
Title: Average Workplace Income: Model Prediction and Survey Data in Dhaka
Panel A
Title: Raw Correlation
Code: `11-analysis-income-validation\workplace_income_analysis.Rmd`

Panel B
Title: Raw Correlation By Skill
Code: `11-analysis-income-validation\table_2_table_C3.do`

Table 3
Title: Average Workplace Income: Model Prediction and Survey Data in Dhaka Comparison with supervised learning using features derived from cell-phone data
Code: `11-analysis-income-validation\residential_income_analysis_ML.Rmd`

Table 4
Title: Average Residential Income: Model Prediction and Residential Income Proxy
Panel A
Title: Raw Correlation
Code: `11-analysis-income-validation\residential_income_analysis.Rmd`

Panel B
Title: Comparison with supervised learning using features derived from cell-phone data (Dhaka)
Code: `11-analysis-income-validation\residential_income_analysis_ML.Rmd`

Table 5
Title: The Heterogeneous Impacts of Hartal on Commuting

Code: `13-analysis-hartal\table_5.do`

Table C1
Title: Numerical Simulation Check: Estimating Gravity with Two Skill Groups
Code: `12-analysis-skills\simulation2skills.do`

Table C2
Title: Gravity Equation with Skills: Estimation Results
Code:
- `7-analysis-gravity\table_1.do`
- `12-analysis-skills\table_C2.do`

Table C3
Title: Average Workplace Income by Skill: Model Prediction and Survey Data in Dhaka
Code: `11-analysis-income-validation\table_2_table_C3.do`

Table D1
Title: Robustness of Workplace Income Validation with Different Assumptions on Id- iosyncratic Shocks and Travel Cost
Code: `11-analysis-income-validation\workplace_income_analysis.Rmd`

Table D2
Title: How Pref. Shocks and Travel Time Affect Income: Estimated Structural Parameters
Code:
`11-analysis-income-validation\workplace_income_analysis_structural.Rmd`

Table E1
Title: Overidentication: Estimating on “Close” and “Far” Tower Samples
Code: `14-analysis-robustness\table_E1_gravity_overid.do`

Table F1
Title: Predicting Workplace Income: Choosing Hyperparameter with Cross-Validation
Code: `11-analysis-income-validation\residential_income_analysis_ML.Rmd`

Figure G1
Title: Impact of Hartal on Commuting to Work
Code: `13-analysis-hartal\figure_G1.do`

Figure G2
Title: Commuting by Calendar Date (Hartals, Holidays and Weekends)

Code: `13-analysis-hartal\figure_G2.do`

Table G1
Title: The Heterogeneous Impacts of Hartal on Commuting: Frequent Commuter Sample
Code: `13-analysis-hartal\table_G1_hartal_frequent_caller_sample.do`

Table H1
Title: Cell Phone Data Coverage at User-Day Level
Code: `9-analysis-stats\table_H1.do`

Figure H2
Title: Commuting Flows from Survey Data and Cell Phone Data
Panel A
Title: Survey vs Cell Phone Data
Code: `10-analysis-commuting-validation\figure_H2a.do`

Panel B
Title: Commuting Flows vs Home-Work Flows
Code: `10-analysis-commuting-validation\figure_H2b.do`

Table H2
Title: Comparison of Commuting Flows from Survey Data and Cell Phone Data
Code: `10-analysis-commuting-validation\table_H2.do`

Table H3
Title: Comparison of Residential Population from Cell Phone Data and Population Census
Code: `10-analysis-commuting-validation\table_H3.do`

Figure H3
Title: Distance to CBD and R^2
Code: `14-analysis-robustness\figure_H3H4.do`

Figure H4
Title: Population Density and R^2
Code: `14-analysis-robustness\figure_H3H4.do`

Figure H5
Title: Prediction R^2 and Geographic Aggregation Level
Code: `14-analysis-robustness\figure_H5.do`

Table H4

Title: Gravity Equation Robustness: Destination Fixed Eects
Code: `7-analysis-gravity\table_H4.do`

Table H5
Title: Average Workplace Income: Model Prediction and Survey Data in Dhaka
Code: `11-analysis-income-validation\workplace_income_analysis.Rmd`

Table H6
Title: Robustness: Average Workplace Income and Survey Income Comparison
Code: `11-analysis-income-validation\workplace_income_analysis.Rmd`

Table H7
Title: Individual Income: Model Predictions and Survey Data
Code:
`11-analysis-income-validation\workplace_income_analysis_structural.Rmd`

Table H8
Title: Average Residential Income: Model Prediction and Residential Income Proxy
Code: `11-analysis-income-validation\residential_income_analysis.Rmd`

Table H9
Title: Robustness: Average Residential Income and Census Income Proxy
Code:
`11-analysis-income-validation\residential_income_analysis_robustness.Rmd`