Replication Files for "Subways and Road Speed"

Yizhen Gu, Chang Jiang, Junfu Zhang, and Ben Zou

February, 2020

Files in the replication folder provide replication codes, log files, results, and replication datasets.

Openicpsr project ID: openicpsr-115681.

1 Data Availability Statement

This paper uses data from four sources. (1) Hourly speed data at the road segment level from Baidu Maps (**Baidu Maps Inc**, 2016-2018). (2) Beijing Household Travel Survey in 2010 and 2015 (**Beijing Transportation Institute**, 2010, 2015). (3) Data on public transit in a panel of Chinese cities (**National Bureau of Statistics**. 2010-2017). (4) 2016 Beijing Transportation Annual Report (**Beijing Transportation Institute**. 2016).

(1) Data from Baidu Maps

Hourly speed data are proprietary of Baidu Maps. Baidu Maps makes some of their data available through commercial or academic collaboration. Applications should be submitted through their portal (website in Chinese). Data sets included in this replication packet have been vetted and approved for posting by Baidu Maps.

(2) Beijing Household Travel Surveys (北京城市家庭出行调查**)**

Beijing Household Travel Surveys (Beijing Transportation Institute, 2010, 2015) are conducted by the Beijing Transportation Institute (BTI). This survey is conducted every year at least since 2005. Once every five years, the survey has a larger sample. We use those larger samples in 2010 and 2015. The micro-level data are restricted use, and the BTI does not maintain a webpage with detailed information on the availability of the data. The BTI publishes an annual yearbook, *Beijing Transportation Annual Reports* (**北京交通发展年报**), partly based on data from this survey. These annual reports are available to download here. For more information, please visit BTI's website. Phone, email, and other contact information is available at the contact page (All websites in Chinese). As far as we know, researchers are encouraged to apply for BTI's data through formal collaborations between their institutions and the BTI.

(3) Data on public transit in a panel of Chinese cities

Information on public transit in Chinese cities are from series of the *Statistical Yearbook of Chinese Cities* (中国城市统计年鉴) (National Bureau of Statistics, 2010-2017). One easy way to access these data is through the National Data Portal, run by the National Bureau of Statistics (website in Chinese). Copy right information is available here.

(4) Beijing Transportation Annual Report (北京交通发展年报)

The *Beijing Transportation Annual Reports* (Beijing Transportion Institute, 2016) is available in PDF format on BTI's website (Website in Chinese). We use the 2016 *Report* for some values listed in Appendix Table C.1.

Raw Data and Cleaning Codes

We provide sample raw data and cleaning codes in folder "ReplicaCodes/DataCleaning". For details see the "_Readme.txt" file in that folder.

Except for restricted-use Beijing Household Transportation Surveys, we provide datasets used to generate results in the paper and the online appendix. Details are provided below. Data sources are cited at the end of the README.

All websites were last accessed on 2/18/2020.

2 Computational Requirements

2.1 Software Requirements

Codes were run on Stata (version 14.2).

Additional package needed:

- `labutil`, `reghdfe`, `ftools`, `estout`, `outtable`. _Setup.do installs all these packages.

- `ols_spatial_HAC.ado` and `reg2hdfespatial.ado` are needed for AppB_FigB6.do. These ado-files are included in folder "Ado."

    - For more information about `ols_spatial_HAC.ado`, see the following post by Solome Chiang (last accessed 2/18/2020).

    - For more information about `reg2hdfespatial.ado`, see the following post by Thiemo Fetzer (last accessed 2/18/2020).

2.2 Memory and Runtime Requirements

The code was last run on a 8-core Intel-based desktop with Windows 10 Enterprise, 32GB of RAM, and 3TB of fast local storage.

Running through all codes took approximately 30 hours.

3 Replication Codes, Log Files, and Results

- Do-files in folder "ReplicaCodes" are codes used to generate all tables and figures in the paper and the online appendix.

    - Note: IndModeSample.dta and HhdVKTSample.dta needed to run Tab6_Cols123.do and Tab6_Cols45.do are from restricted-use data, and are not provided in the replication packet.

- _RunAll.do is the master do-file that generates all the results.

    - to run the code, change to the local directory following command `cd` in _RunAll.do.

- Log files of running these codes are provided in folder "LogFiles."

- Results (tables and figures) are stored in folder "TablesFigures."

- Data are stored in folder "Data."

3.1 List of Data Files

- BaseSamp.dta — Baseline sample, used to generate most results in the paper and the online appendix.

- CityChars.dta — City characteristics.

- ExtendSample.dta — Sample including all road segments, mainly used to generate Figure 6.

- HhdVKTSample.dta — Household characteristics and car usage from Beijing Household Transportation Surveys, used to generated Table 6.

    - Not provided in the replication packet.

- IndModeSample.dta — Individual characteristics and transportation mode choices from Beijing Household Transportation Surveys, used to generated Table 6.

    - Not provided in the replication packet.

- PublicTransits.dta — City-year panel on urban transportation and public transit, used to generate Appendix Table B.4.

- RushNonrushHours.dta — Sample data with both rush-hour and non-rush-hour speed (baseline sample only uses rush-hours), used to generate Figure 6.

- SubwayLines.dta — A list of subway lines in China as of 12/31/2017.

- Subway_Lines_Cases.dta — 45 treated lines used in the analysis.

- UrbanTransitReports.dta — Original data on city-level public transit.

- WithAgainstTraffic.dta — Sample data with both segment-hour observations that are in the same direction of the rush-hour traffic, and those that in the opposite direction of the rush-hour traffic, used to generate Figure 6.

- 2017_part18.dta — sample raw data, used to generate Appendix Figure A.4.

- linkInfo_new.dta — raw data on road segment type, name and location, used to generate Appendix Figures A.5 and B.6.

- HourlySample.dta — hourly speed sample, used to generate Appendix Figure A.5 and Appendix Table B.5.

3.2 List of Codes and Results

Caution: Some code may take a long time to run. Please see the included log files for an estimated running time.

- _Setup.do

  - installs additional packages used in the replication codes.

- _RunAll.do — Run all the do-files that generate all the results. Specific do-files introduced below.

  - to run the code, change to the local directory following command cd.

- Fig1.do — Generates Figure 1.

- Fig2.do — Generates Figure 2.

- Fig3.do — Generates Figure 3.

- Fig4.do — Generates Figure 4.

- Fig5A.do, Fig5B.do, Fig5C.do — Generates Figure 5.

- Fig6.do — Generates Figure 6.

- Figure 7 is not empirical, there is no code associated with it.

- Footnote20.do — Generates in-text numbers cited in Footnote 20.

- Table 1 is not empirical, there is no code associated with it.

- Tab2.do — Generates Table 2.

- Tab3.do — Generates Table 3.

- Tab4.do — Generates Table 4.

- Tab5.do— Generates Table 5.

- Tab6_Cols123.do, Tab6_Cols45.do— Generates Table 6.

  - Not replicable. IndModeSample.dta and HhdVKTSample.dta needed to ru the code are from restricted-use data, and is not provided in the replication packet.

- Appendix Figures A.1, A.2 and A.3 are not empirical, there is no code associated with them.

- AppA_FigA4.do — Generates Appendix Figure A.4.

- AppA_FigA5.do — Generates Appendix Figure A.5.

- Appendix Table A.1 is not empirical, there is no code associated with it.

- AppB_FigB1FigB2FigB3TabB1.do — Generates Appendix Figure B.1, Figure B.2, Figure B.3, and Appendix Table B.1

    - Figure B.1 includes:
        * AppB_SeasonalityCityChars_A.pdf
        * AppB_SeasonalityCityChars_B.pdf
        * AppB_SeasonalityCityChars_C.pdf
        * AppB_SeasonalityCityChars_D.pdf
        * AppB_SeasonalityCityChars_E.pdf
    - Figure B.2 includes AppB_SeasonalityCityChars_F.pdf
    - Figure B.3 includes
        * AppB_SeasonalityCityChars_G.pdf
        * AppB_SeasonalityCityChars_H.pdf
    - Table B.1 is summarized in AppB_DiffSeasonality_Tab.tex

- AppB_FigB4.do — Generates Appendix Figure B.4.

    - AppB_FigB4_A.pdf: Panel A.
    - AppB_FigB4_C.pdf: Panel B.
    - AppB_FigB4_B.pdf: Panel C.
    - AppB_FigB4_D.pdf: Panel D.

- AppB_FigB5TabB2.do — Generates Appendix Figure B.5 and Table B.2.

    - Figure B.5 includes AppB_FigB5.pdf.
    - Table B.2 is summarized in AppB_TabB2A.tex.

- AppB_FigB6.do — Generates Appendix Figure B.6

- AppB_FigB7.do — Generates Appendix Figure B.7.

- AppB_TabB3.do — Generates Appendix Table B.3.

- AppB_TabB4.do — Generate Appendix Table B.4.

- AppB_TabB5.do — Generate Appendix Table B.5.

- Appendix Table C.1 is not empirical, there is no code associated with it.

- Folder "DataCleaning" lists codes used for data cleaning. These codes need original raw data which are not provided in the replication package. See "_Readme.txt" in the folder for details.

References of Data Sources:

Baidu Maps Inc. August 1, 2016 to January 31, 2018. "Hourly Speed Data at Road Segment Level." Baidu Maps Inc. Data application portal: https://jiaotong.baidu.com/contactus/.

National Bureau of Statistics. 2010-2017. "Statistical Yearbook of Chinese Cities." (中国城市统计年鉴) *Statistical Press of China*. ISBN: 978-7-5037-6192-8. Digital version available at http://data.stats.gov.cn/easyquery.htm?cn=E0105.

Beijing Transportation Institute. 2010, 2015. "Beijing Household Travel Surveys." (北京城市家庭出行调查) Beijing Transportation Institute.

Beijing Transportion Institute. 2016. "Beijing Transportation Annual Report." (北京交通发展年报) Beijing Transportation Institute. Available at http://www.bjtrc.org.cn/List/index/cid/16.html.

(All websites in Chinese, last accessed on 2/18/2020)