README file for replication of "An Engel Curve for Variety" by Nicholas Li

All data is freely available. I include the original downloads in raw format for convenience. However, in some cases you must register to use the data.

- ICRISAT VDSA district data
- NSS consumption, employment and time use data
- Colorado Nielsen Homescan data
- American Time-Use Survey
- UK National Food Survey data
- Spanish Encuesta de Presupuestos Familiares

Below I list the programs for replicating all results in the paper and web appendix. All analysis was conducted using STATA 13.1 64 bit. Most of the raw data files are in ASCII format but a few are already in STATA format.

The programs need to be run in order as they use outputs from earlier programs (except f_other.do which uses the non-Indian data).

To run the programs, set local macro 'rootdir' to the root directory – and all of the output figures and tables will be created there. The .tar files need to be opened and then unzipped in directories with the same names (e.g. "districts," "Round66") for the programs to run as is. Note that the total unzipped data is over 10GB.

Replication files:

1. f_other.do
    - Purpose: Non-Indian data for Tables and Figures
    - Raw data inputs: (from Other folder) Colorado Nielsen data from Aguiar and Hurst (2007), American Time-Use data 2008, UK National Food Survey (2000), Spanish EPF (1985-1996)
    - Outputs: Figure A.1, Figure A.2, Table A.4, Table A.5
2. read_in_NSS_sch1.do
    - Purpose: Read in raw NSS schedule 1 data and output consumption and household files for rounds 38, 43, 55, 61, 66
    - Inputs: Raw NSS data files (from Round38-Round66 folderS)
    - Outputs: STATA files used as inputs for other programs
3. read_in_TUS.do
    - Purpose: Read in raw Time-Use data

- Inputs: Raw Time-Use data (from folder TUS)
- Outputs: STATA files used as inputs for other programs

4. common_consumption_NSS.do

   - Purpose: Harmonize definition/labeling of consumption varieties and quantities and also states/regions/districts across rounds for NSS schedule 1
   - Inputs: STATA files created earlier
   - Outputs: STATA files used as inputs for other programs

5. common_hh_NSS_sch1.do

   - Purpose: Harmonize NSS schedule 1 household variables across rounds
   - Inputs: STATA files created earlier and NIC/NCO crosswalks (in root directory) to get common industry/occupation codes
   - Outputs: STATA files used as inputs for other programs

6. common_hh_NSS_sch10.do

   - Purpose: Harmonize NSS schedule 10 (employment) variables across rounds, including industry codes; also generate food retail/distribution employment data
   - Inputs: STATA files created earlier and raw data from folders Round38-Round66 (a few data files already in STATA format)
   - Outputs: STATA files used as inputs for other programs

7. prices_indexes_final.do

   - Purpose: Create price indexes for within region across rounds, across sectors within region-round, and over districts/time
   - Inputs: STATA files created earlier, district code concordance file
   - Outputs: STATA files used as inputs for other programs

8. f_stylizedfacts.do

   - Purpose: Create Figure 1, Table 1, Appendix Figures A.2 and A.3
   - Inputs: STATA files created earlier
   - Outputs: Figure 1, Appendix Figures A.2 and A.3, summary statistic output used to fill Table 1

9. f_variety_districts.do

   - Purpose: Create Table 2, Appendix Table A.1 and also prep district variables for shopping time-analysis, region rainfall
   - Inputs: STATA files created earlier, ICRISAT VDSA data and district concordances/codebooks (in folder "districts")
   - Outputs: Table 2, Appendix Table A.1, STATA files used as inputs for othe programs

10. f_variety_mechanisms.do

    - Purpose: Create Tables 3 and 4 and Appendix Tables A.2 and A.3, Figure A.5 and A.6
    - Inputs: STATA files created earlier
    - Outputs: Tables 2 and 3, Appendix Tables A.2 and A.3, Figure A.5 and A.6

11. f_elasticity_final.do

    - Purpose: Estimate sigma parameter and generate Table 5, Appendix Tables A.6, A.7 and A.8
    - Inputs: STATA files created earlier (including regional rainfall)
    - Outputs: Table 5, Appendix Tables A.6, A.7, A.8

12. f_ceswelfare_overtime.do

    - Purpose: Estimate model parameters (except sigma) and main welfare results/decomposition over 1983-2009
    - Inputs: STATA files cretaed earlier
    - Outputs: Table 6, Appendix Figure A.7, A.8, A.9, A.10, A.11, Table A.9

13. f_ceswelfare_ru.do

    - Purpose: Estimate model parameters (except sigma) and main welfare results/decomposition for rural-urban comparison in 2009
    - Inputs: STATA files created earlier
    - Outputs: Table 7, Appendix Table A.9

14. f_parameters.do

    - Purpose: Examine relationship between estimated variety costs and local retail/income/density variables
    - Inputs: STATA files created earlier
    - Outputs: Appendix Table A.14

15. f_bootstrap_se.do

    - Purpose: Estimate boot-strapped standard errors for region-level results and create relevant tables and figures
    - Inputs: STATA files created earlier
    - Outputs: Appendix Figures A.12 and A.13, Tables A.10-A.13