Readme file for replication package of
"Estimating Substitution Patterns and Demand Curvature
in Discrete-Choice Models of Product Differentiation"
Review of Economics and Statistics (2024)

Cameron Birchall, Debashrita Mohapatra, Frank Verboven
May 8, 2024

This file describes the code and data used to replicate the results in the paper and appendix.

Data Availability and Sources

Statement about rights:

We certify that the authors of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Summary of availability:

Some of the data cannot be made publicly available

Details on data and source:

The data are scanner data on cereal sales (value and volume) at the product level (EAN code, description and package size), weekly during 2011-2013, aggregated over 6 provinces in the Netherlands. The data are proprietary, they were purchased from IRI in 2015. The contact person at that time was Eric Broers, Eric.Broers@iriworldwide.com, from IRI Nederland, website: www.iriworldwide.nl.

Computational Requirements

Software requirements:

1. Stata 16 or above (data management and presentation of results)
2. Matlab (demand estimation)

Memory and runtime requirements:

The files can be run on a standard 4-core computer with 48GB RAM. Running all the files requires approximately 14 hours.

Description of Programs/Code

There are three main folders:

\1) Data

- **\Raw dataset**

This subfolder contains two raw data sets in Stata format:

    - Cereals – S4932.dta (proprietary): IRI scanner data at the product level, weekly and per province on the following variables: province (market), week, week number ("week_num"), month, year, sales value ("pq"), sales volume ("q") and package size ("size").
    - Product_description_11.dta: Information on product description of each EAN code. The included variables are: product identifier ("EAN"), product description ("Description"), (broad and detailed) product type, (umbrella) brand, package type and package size.
- MasterData_Cereal.mat: The main dataset used for the analysis (in Matlab format), after cleaning the raw data (as done below in folder "\2) Code").
- MasterData_Cereal.dta: : The main dataset used for the analysis (in Stata format), after cleaning the raw data (as done below in folder "\2) Code").

\2) Code

- **\Stata code**
    - 1) Create panel.do: This do-file starts from the raw data sets to create the final datasets. It calls four do-files:
      1a) Clean outliers.do
      1b) Box-Cox data management.do
      1c) Instruments.do
      1d) Clean for estimation.do
      This involves tasks such as cleaning data by removing outliers, generating market share variables for the logit, and Box-Cox logit model, and generating several sets of instruments.
        - log-log.do & Own-elasticity vs price for log-log (Figure 1): estimates log-log (constant elasticity) demand model and plots the elasticity estimates against price to generate Figure 1 (in folder "\3) Graphs and Outputs").
        - Figures and Tables_in_paper.do: uses the results from the demand estimation to generate Figure 2, Figure A.2, Figure A.3, Table A.4, Table 1, and Figure A.1 (in folder "\3) Graphs and Outputs")

- **\Matlab code**
    - Run the main file "**a1main_file.m**" to replicate results from the four specifications. An outline of the steps follows using the dataset MasterData_Cereal.mat in folder "\1Data".
    - prepare_data.m: prepares the dataset, constructs the instruments for analysis
    - Run the section "**1. Logit Estimation**" to get the results from the logit specification. It uses the following files:
        - logit_exercise.m: performs various calculations and estimations using logit estimation, including parameter estimation, standard errors, market share, variance-covariance matrix computation, demand residuals analysis, elasticity matrix, own price elasticity derivation, generates derivative matrix, marginal costs estimation, pass-through rates, and the generation of confidence intervals for own price elasticity using bootstrap techniques.
        - logitregression.m: estimates parameters from logit specification
        - price_elasticity_logit.m: computes the Elasticity matrix, Own price elasticities, and Derivative matrix from logit specification.

        - MC_Passthrough_logit.m: computes the marginal costs and pass-through rates.
    - Run the section "**2. Simple Box-Cox Estimation**" to get results from the Simple Box-Cox specification. It uses the following files:
        - get_SE_SimpleBoxCox.m: estimates the parameters and the standard error.
        - price_elasticity_Simple_Boxcox.m: computes the Elasticity matrix, Own price elasticities, and Derivative matrix from the simple box-cox specification.
        - MC_Passthrough_simple_BoxCox.m: computes the marginal costs and pass-through rates.
    - Run the section "**3. BLP Estimation**" to get results from the BLP specification. It uses the following files:
        - get_paral_se_BLP.m: estimates the parameters and the standard error.
        - price_elasticty_BLP.m: computes the Elasticity matrix, Own price elasticities, and Derivative matrix from BLP specification.
        - MC_Passthrough_BLP.m: computes the marginal costs and pass-through rates.
    - Run the section "**4. BLP and Box-Cox Estimation**" to get results from the BLP specification. It uses the following files:
        - get_paral_se_BLP_Boxcox.m: estimates the parameters and the standards error.
        - price_elasticity_BLP_boxcox.m: computes the Elasticity matrix, Own price elasticities, and Derivative matrix from BLP and Box-Cox specification.
        - MC_Passthrough_BLP_boxcox.m: computes the marginal costs and pass-through rates.
    - The analysis uses other subfunctions for computation, such as BLP contraction mapping, market shares, demand derivatives, the value of the objective function:
        - computey.m, computey_Simple_BoxCox.m, contraction.m, derivative_simple_BoxCox_bymarket.m, derivativeBLP_boxcox_bymarket.m, derivativeBLP_bymarket.m, expmu_simple_BoxCox.m, expmu.m, fminsearchbnd.m, get_derivative_matrix_logit.m, getranddraw.m, marketlocs.m, mktshares.m, obj_BLP_boxcox.m, obj_BLP_DemandOnly.m, obj_simpleBoxcox.m, shares.m
    - Export_dataset_results_to_stata.m
        - Prepares the MATLAB analysis results from the four specifications (Logit, Simple box-cox, BLP, BLP and Boc-Cox) for export to Stata for further analysis and to generate graphs and tables.
        - Stores the results in folder "\3) Graphs and Outputs\OutputFromMatlab"

\3) Graphs and Outputs

- **\Graphs**
    - Fig 1 - own-elasticity vs own-price (log-log).png: Refers to "Figure 1: Descriptive own-price elasticity vs. own-price" in paper
    - Fig 2 - own_elasticity_vs_own_price.jpg: Refers to "Figure 2: Own-elasticity vs. own-price" in paper

- Fig A.1 - consumption vs income.jpg: Refers to “Figure A.1: Cereal consumption per potential consumer vs. income” in paper
- Fig A.2 - own_elasticity_vs_own_price_NoLogit.jpg: Refers to “Figure A.2: Own-elasticity vs. own-price (without simple logit)” in paper
- Fig A.3 - Percentage_markup_vs_own_price.jpg: Refers to “Figure A.3: Markups vs. own-price” in paper

- **\Outputs**
  Contains various datasets that are directly exported from analysis done in Matlab, as Text files. This contains: a) Diversion ratios between different products, b) Price elasticities from bootstrap, c) Estimated own price elasticities d) Pass-through rates, from the following specifications:
  - Logit specification:
    - logit_diversion_ratios.txt
    - Logit_price_elasticity_bootstrap.txt
  - BLP specification
    - blp_diversion_ratios.txt
    - BLP_price_elasticity_bootstrap.txt
  - Simple Boxcox specification
    - simple_boxcox_diversion_ratio.txt
    - Simple_boxcox_price_elasticity_bootstrap.txt
  - Joint BLP and Boxcox specification
    - blp_boxcox_diversion_ratios.txt
    - BLP_boxcox_price_elasticity_bootstrap.txt
  - own_price_elast_all_specifications.txt
  - OwnPriceElasticity_PassthroughRate.txt
  - log-log.xls: The dataset is used to estimate descriptive own-elasticity per product via instrumental variables log-log regression (in log-log.do)