Readme:

Nonparametric welfare and demand analysis with unobserved individual heterogeneity

Sam Cosaert and Thomas Demuynck

January 23, 2017

Abstract

The *HeterogeneityRUM* package contains the data and programs necessary to replicate the results in our paper. Pathnames need not be changed when files are saved in folder *CodesPublishedVersion* on the *C:/* drive.

1 Data and programs

HeterogeneityRUM consists of three data folders and four program folders. The data are drawn from the Diary Survey of the Consumer Expenditure Survey (CE), collected for the Bureau of Labor Statistics by the U.S. Census Bureau (see *https://www.bls.gov/cex/*). We use the 1994–2007 waves. The data from 1996 to 2007 can be downloaded from the public–use microdata (PUMD). Data before 1996 can be purchased from the Bureau of Labor Statistics. We refer to *https://www.bls.gov/cex/csxform.pdf* for requesting access to these data.

The data folders (*CEX3*, *CEX3seas* and *CEX8*) contain, respectively, the data for the deseasonalized setting with three goods, the nondeseasonalized setting with three goods and the deseasonalized setting with eight goods.

The program folders (*MatlabAnalysis3goods*, *MatlabAnalysisSeasonal* and *MatlabAnalysis8goods*) follow this structure. Finally, the program folder *Subsampling* contains the programs to produce the results in Table 1 of the online appendix. We mainly use STATA for data management and Matlab for the computation of our bounds.

1.1 Data

This subsection explains the construction of our final sample represented by the 'Engelparameters' files. Readers who are more interested in the actual analysis can start from subsubsection 'Analysis main text'. The final samples are saved in (sub)folders *Engelparameters* (or *Engelparametersbb*) of the program folders.

The data folders (*CEX3*, *CEX3seas* and *CEX8*) consist of

- *Diary* folders, which contain three files 'fmldxx.dta', 'memdxx.dta' and 'uccgroupsxx.dta'. Files 'fmldxx.dta' (household information) and 'memdxx.dta' (information on household members) are raw Dairy Survey data. Finally, 'uccgroupsxx.dta' aggregates commodities in 'expdxx.dta' by means of the universal classification codes (UCC titles) in 'uccdxx.txt'. Because the data for 1994–1995 are not publicly available, we refer to the website *https://www.bls.gov/cex/csxform.pdf*

- *Year* folders, which contain do–files for data cleaning using STATA and the resulting intermediate samples 'CEXxxfam.dta'

- price files from the Bureau of Labor Statistics. These data are publicly available at *https://www.bls.gov/data/*

- (*DataStruct*, which saves intermediate output)

- Matlab files that produce the mat–files in *EngelInputs.* These files are necessary in the final 'preparatory step' (see infra) before the actual analysis

To create the mat–files in *EngelInputs*, first run all the programs in the *Year* folders. More specifically, run do–files

'CEXmakedataxx_1_restrictsample.do',

'CEXmakedataxx_1a_mergemembers.do',

'CEXmakedataxx_1b_mergemembers.do',

'CEXmakedataxx_2_aggregateexpnd.do' and

'CEXmakedataxx_3_mergesampleexpnd.do'.

These do–files limit the sample based on demographic characteristics (1), merge information from the family and individual files (1a and 1b), aggregate expenditures from the detailed uccgroups (2) and finally merge household information and expenditure data (3).

Subsequently, run 'runcollebounds_1_dataCEX.m' and 'runcollebounds_2_engelCEX.m'. The other m–files are subfunctions of these programs. This creates the 'EngelInputsxxfam.mat' files. The analysis can then proceed with Matlab.

1.2 Programs

The program folders (*MatlabAnalysis3goods*, *MatlabAnalysis8goods*, *MatlabAnalysisSeasonal* and *Subsampling*) contain the codes to replicate all tables and figures in the paper and the online appendix. *Note that the Tomlab package is necessary to successfully run the Matlab programs.* If desired, one can easily replace Tomlab/cplex with another linear program solver.

Preparatory step The programs that will be discussed in the next subsubsection use mat–files 'Engelparameters.mat' (for our bounds) or 'Engelparametersbb.mat' (for BBC bounds) as inputs. m–file 'find_pilot_h_and_lambda.m' creates these files by merging the information in *EngelInputs* and the residuals from a first–stage regression of total expenditure on income. These codes, and the resulting mat–files, can be found in folders *Engelparameters* or *Engelparametersbb*, respectively. The data set is now ready for analysis.

Analysis main text The programs to replicate the results in the main text are collected in *MatlabAnalysis3goods*. The prefixes indicate to which Table or Figure the m–files contribute. The main functions are:

- 'Table1_priceindices.m': computes the Laspeyres, Paasche and Tornqvist price indices
- 'Table1_runbb.m': computes BBC bounds on money metric utility
- 'Table1_runbounds.m': bounds the distribution of money metric utility

'Table1_runbb.m' produces the results in the second–to–last column while 'Table1_runbounds.m' produces the results in the final column. Recall to divide the bounds by the median income level in the reference year. Furthermore, the bounds in the last column are the average bounds across different quantiles.

- 'Figure4_av_shares2_3goods.m': provides summary statistics on budget shares
- 'Figure5_distribution_of_col2_3goods.m': bounds the distribution of cost of living (run 'Table1_runbounds.m' first and save results)
- 'Figure6_av_col_for_quantiles2_3goods.m': computes cost of living for different income quantiles
- 'Figure7_comp_var2_3goods.m': bounds compensating variation (run 'Table1_runbounds.m' first and save results)
- 'Figure8_bounds_on_demand_3goods.m': bounds the distribution of demand responses
- 'OA_Figure1_distribution_of_col2_3goods_per_income.m': bounds the distribution of cost of living for poor households (results in online appendix) (run 'Table1_runbounds.m' first and save results)

Finally, 'OA_Table1_Subsampling_HeterogeneityRUM.m' in folder *Subsampling* produces confidence intervals for our nonparametric lower and upper bounds on money metric utility.

Robustness: nondeseasonalized expenditures To replicate the nondeseasonalized results in the online appendix, consider the following main programs from *MatlabAnalysisSeasonal*:

- 'OA_Table2_priceindices.m': computes the Laspeyres, Paasche and Tornqvist price indices (non–ds)
- 'OA_Table2_runbb.m': computes BBC bounds on money metric utility (non–ds)
- 'OA_Table2_runbounds.m': bounds the distribution of money metric utility (non–ds)
- 'OA_Figure2_distribution_of_col2_3goods.m': bounds the distribution of cost of living (non–ds) (run 'OA_Table2_runbounds.m' first and save results)

Robustness: 8 commodities The results with 8 commodities can be replicated by means of *MatlabAnalysis8goods*.

- 'OA_Table3_priceindices.m': computes the Laspeyres, Paasche and Tornqvist price indices (8 goods)
- 'OA_Table3_runbb.m': computes BBC bounds on money metric utility (8 goods)
- 'OA_Table3_runbounds.m': bounds the distribution of money metric utility (8 goods)
- 'OA_Figure3_av_shares2_8goods.m': provides summary statistics on budget shares (8 goods)
- 'OA_Figure4_distribution_of_col2_8goods.m': bounds the distribution of cost of living (8 goods) (run 'OA_Table3_runbounds.m' first and save results)