README file for JAERE *Third-best carbon taxation*

Overview

In the replication paper for our article "Third-best carbon taxation: trading off emission cuts, equity, and efficiency", we provide the following files for estimation and simulation:

- Estimation of the EASI demand system (Stata)
- Estimation of labour supply (Stata)
- Matching consumption bundles with their carbon content (Stata)
- Random selection of simulation subset (Stata)
- Simulation of policy changes (GAMS)
- Final processing, formatting, and plotting (EXCEL)

Data Availability and Provenance Statements

To reproduce our results, you need access to the German Income and Consumption Survey (EVS), in particular files AAGSHB 2013, AAGSHB 2008, HBEIGS 2003 and EIHB1998. The EVS is collected by the German statistical office but is proprietary. Researchers based in Germany can request access to the necessary data file under: https://www.forschungsdatenzentrum.de/de/haushalte/evs. Commodity prices used in the estimation are provided by the German statistical office and are available under: https://www.destatis.de/EN/Themes/Economy/Prices/Consumer-Price-Index/_node.html

We also used heating degree days from the EUROSTAT (Statistics | Eurostat (europa.eu))

Carbon content of consumption is compiled in the Environmental-Economic Accounts (EEA). The latest available data covers the period 2010-2015. We are using the year 2013, which matches the latest version of the microdata available for this period (see https://www.destatis.de/EN/Themes/Society-Environment/Environment/Material-Energy-Flows/Publications/Downloads-Material-Energy-Flows/co2-emissions-pdf-5851306199004.pdf?__blob=publicationFile).

- ☐ This paper does not involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

  If box above is checked and if no simulated/synthetic data files are provided by the authors, please skip directly to the section on Computational Requirements. Otherwise, continue.

Statement about Rights

- ☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- ☐ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

(Optional, but recommended) License for Data

Summary of Availability

- ☐ All data **are** publicly available.
- ☒ Some data **cannot be made** publicly available.
- ☐ **No data can be made** publicly available.

Details on each Data Source

The following table provides a summary of the data sources in this estimation.

The data sources we used are all from the German statistical office. The German Income and Consumption Survey (EVS) for 2013, 2008, 2003, and 1998 is available upon request to this Organization. We also use price indices that can be freely downloaded from their website. Heating degree days are also freely available from Eurostat. These two last data sets are available at monthly basis. Regarding heating degree days, we use data at county level.

Finally, we use CO2 emission factors that they provide directly to us for this analysis with more desegregation. We provide the estimated footprints for different commodities in this package. Aggregated data on emissions is available under: https://www.destatis.de/EN/Themes/Society-Environment/Environment/Material-Energy-Flows/Publications/Downloads-Material-Energy-Flows/co2-emissions-pdf-5851306199004.pdf?__blob=publicationFile

Table 1 summarises all the raw data sets needed to replicate our article. The table also provides the links where the data can be obtained and more details on their characteristics.

Table 1. Row data used in the estimation and their sources

Data.Name	Data.Files	Location	Provided	Citation
EVS 2013	evs_aagshb2013_slr.csv	German statistical office. Available upon request	False	German Income and Expenditure Survey (EVS) - German Federal Statistical Office
EVS 2008	gf3_08_zew_f.csv			
EVS 2003	gf3_03_zew_f.csv			
EVS 1998	gf7_93_zew_f.csv	Homepage - German Federal Statistical Office (destatis.de)		
Index price	Available in different formats	German statistical office	False	Consumer price index - German Federal Statistical Office (destatis.de)
Heating degree days	Available in different formats	Eurostat	False	Statistics | Eurostat europa.eu)
Emission factors	Emission factors_270721_submission.xls		True	Destatis (2019)*

* Destatis (2019) is Environmental Accounting. Direct and indirect CO2 emissions in Germany 2010 – 2015. Federal Statistical Office. Available at: Environmental-Economic Accounting - Direct and indirect CO2 emissions in Germany 2010 - 2015 (destatis.de)

Dataset list

With the data described above, the following data sets are created for our analysis.

Table 2. Data sources used by the econometric analysis and simulations

Data file	Source	Notes	Provided
all_years_distribution_base.dta	German statistical office	This file appends the different cross-sections of the EVS data to estimate the demand systems.	No
all_years_distribution_2.dta	German statistical office	This data has more details on the working hours of the household.	No
prices_evs300818.dta	Eurostat	Information on the index prices for different commodities	No
hdd060416.dta	German statistical office, own calculations	Number of heating degree days	No
pf_p_2.dta pf_q_2.dta	German statistical office, own calculations	Our estimation of aggregated carbon footprints for different footprints is based on the data provided by the German statistical office. See file: factors_0505_20_submission.xls	Yes
Simulation analysis Simul_data.xls	German statistical office, own calculations	This is a subsample of the EVS data 2013 (the latest available year when we estimated). This is to reduce the computational burden when doing the simulation.	No

Computational requirements

Software Requirements

- Stata (code was last run with version 18)
  - estwrite (as of 2005)
  - xml_tab (as of 2006)

The program "Setup.do" will install all dependencies locally, and should be run once.

- GAMS software
  - Solver CONOPT4.

Controlled Randomness

For the simulation analysis, a random subsample of 1000 units is used. In the file EVS simulation 2020_s, we have set the unit reproducibility always to be the same. This is done in line 921 of this file, where we indicate that the sample has to be reproducible by setting a seed.

Memory and Runtime Requirements

We used the standard Windows 10 operating system with 64-bit operating system, x64-based processor. RAM capacity of 16.0 GB (15.7 GB usable).

Summary

Approximate time needed to reproduce the analyses on a standard (2024) desktop machine:

- ☐ <10 minutes
- ☐ 10-60 minutes
- ☐ 1-2 hours
- ☒ 2-8 hours
- ☐ 8-24 hours
- ☐ 1-3 days
- ☐ 3-14 days
- ☐ > 14 days
- ☐ Not feasible to run on a desktop machine, as described below.

Description of programs/code

- Data preparation_demand_systems.do: This file appends the cross sections of the survey and includes price indices to estimate the demand systems
- Data preparation_labour supply.do: This file creates the database to estimate the labour supply function parameters. It has a detailed representation of the working hours.

  EASI Estimation_s.do: This file estimates the parameters from the EASI demand system and cross-price and expenditure elasticities.
- EASI labour_Estimation_s.do: This file estimates the labour supply function using the parameters from the EASI demand system
- LES_EVS_ Estimation_s.do: This file estimates the parameters of the linear expenditure system.
- LES_EVS_labour_Estimation_s.do: This file estimates the parameters of the labour supply function using the linear expenditure system.
- EVS simulation_Estimation_s.do: This file creates the Excel file with 1000 units for the simulation exercise (Simul_data.xls). Once the do files "*EASI labour_2020_s.do*" and "EASI Estimation_s.do'' are run, an Excel file with information on footprints worked hours, budget shares and other socioeconomic variables is created for the simulation exercise. We have set a seed in the Stata file so the user can replicate the file random sample of 1000 units.
- LES 2020_s.do: It produces the parameters for the linear demand system.

(Optional, but recommended) License for Code

Instructions to Replicators

These are the general tasks the provides files replicate and the software they use:

- Estimation of the EASI demand system (Stata)
- Estimation of labour supply (Stata)
- Matching consumption bundles with their carbon content (Stata)
- Random selection of simulation subset
- Simulation of policy changes (GAMS)
- Final processing, formatting, and plotting (EXCEL)

Step 1: After running the do file "Set up.do", the user needs to run the data creation files first (Data preparation_demand_systems.do and Data preparation_labour supply.do)

Step 2: Estimate the parameters from the EASI demand systems (EASI Estimation_s.do) and then the labour supply parameters (EASI labour_Estimation _s.do)

Step 3: Estimate the parameters for the linear demand system (LES Estimation_s.do) and after the labour supply function (LES_EVS_labour_Estimation_s.do).

Step 4: Create the data for the simulation exercise (EVS simulation_ Estimation _s.do)

Step 5: Load the estimation output and survey data into GAMS (param_EASI.gms, param_LES.gms, and data_reading.gms)

Step 6: Simulate policy experiments (policy_simulations.gms)

Step 7: Conduct final data processing and graphing (Third_best_replication.xlsx)

Details

I. Estimation of the EASI demand system:

Data files input: AAGSHB 2013, AAGSHB 2008, HBEIGS 2003, price index and heating degree days

I.a) Preparing data for the estimation of the demand system (Stata do file "Data preparation_demand_systems.do'')

I.b) Estimating parameters of EASI demand system (Stata do file "EASI Estimation_s.do'')

Data output: Stata file with the data on expenditure and commodity prices ("all_years_distribution_base.dta''). This file also generates the output file "output_EASI.xml''. This format is suitable for GAMS. But it can also be read in Excel. Table B2 provides the parameters displayed in this file. This file also generates the summary statistics provided in Table B1 and the matrices with own price and expenditure elasticities provided in Tables 1, B3, and B4. When running this file, a Stata file will also be generated ("EASI_ouput.sters") to estimate the labour supply.

II. Estimation of labour supply:

Data input: EVS files AAGSHB 2013, AAGSHB 2008 and HBEIGS 2003, price index and heating degree days

II.a) Preparing data for the estimation (Stata do file Data: "Data preparation_labour supply.do'')

II.b) Estimating parameters of labour supply (Stata do file "EASI labour_Estimation_s.do'')

Data output: "Data preparation_labour supply.dta" Stata file with the parameters from the labour supply "output_labour_EASI.xml''. It provides the parameters displayed in Table 2. This file also generates the output "EASI_labour_ouput.sters'' used when generating the simulation data.

Note that a similar procedure needs to be followed to estimate the parameters for the linear expenditure system. By running the do files "LES_EVS_Estimation_s.do'' and "LES_EVS_labour_Estimation_s.do'' the file "output_LES_ouput.xlm'' and "output_labour_LES.xlm'' will be generated to be used by GAMS. These parameters do not appear directly in the article but are used to estimate the figures in Appendix F.

III. Matching consumption bundles with their carbon content:

III.a) Matching environmental accounts with EVS goods classification and aggregation to match EASI system. The file "Emission factors_270721_submission.xls'' documents the

procedure to obtain the files “pf_p_2.dta“ and “pf_q_2.dta’’ used for the Stata do file “EVS simulation_Estimation_s.do’’ to generate the simulated data.

III.b) Producing Table 3 and Engle curves (figures B1 and B3)

Data output: Data files with footprints in kg/euro and kg/quantity consumed and Engel curves for several commodities.

IV. Preparation of data for simulation

IV.a) Routine to prepare a random sample of 1000 units (EVS simulation_2020_s.do)

Data output: An Excel file with 1000 units for the simulation exercise (“Simul_data.xls’’). Once the do file “EVS simulation_Estimation_s.do” is run, an Excel file with information on footprints, worked hours, budget shares, and other socioeconomic variables are created for the simulation exercise. We have set a seed in the Stata file so the user can replicate the file's random sample of 1000 units.

V. Simulation of policy scenarios

Data input: Replication_simulation.zip (containing files “param_EASI.gms”, “param_LES.gms”, “data_reading.gms”, “param_EASI.xlsx”, and “param_LES.xlsx” in sub-folder “simul_data” and “policy_simulation.gms” and “Third_best_replication.xlsx” in main directory). Note that the file “Simul_data.xls“ file created in step IV needs to be copied into the “simul_data” sub-directory.

V.a) Load EASI and LES demand system and household characteristics into GAMS (GAMS files “param_EASI.gms”, “param_LES.gms”, and “data_reading.gms” to be run in this order in the “simul_data” sub-directory). Note that before the files can be executed, the data file “param_EASI.xlsx“ and “param_LES.xlsx“ in the subfolder have to be updated with the parameters of the demand system estimation using file “output_EASI.xml’’ and “output_LES_output.xml’’. See example files in Replication_simulation.zip for correct format “param_EASI - example.xlsx” and “param_LES - example.xlsx”. In addition, parameters for the labour supply function using the EASI and LES can be found in the files: “output_labour_EASI.xlm” and “output_labour_LES.xml”.

Data output: several “.gdx” files in subdirectory “data”.

V.b) Simulating policy scenarios by executing “policy_simulation.gms” in the mail directory. Note that this requires access to the solver CONOPT4.

Data output: Third_best_replication.txt

VI. Final processing, formatting, and plotting

Data input: “Third_best_replication.txt”, “Third_best_replication.xlsx”

VI.a) Copy/Overwrite the content of “Third_best_replication.txt” into the ‘Data Input’ sheet of “Third_best_replication.xlsx”.

VI.b) The Excel file “Third_best_replication.xlsx” is set up to automatically conduct simple formatting and aggregation calculations, including the calculation of EVs. The file contains information used for plotting policy scenarios. Some of the figures of the manuscript are also produced as examples.

Note that all Stata files described in steps I-IV are contained in Replication_estimation.zip, and Steps V-VI require those contained in Replication_simulation.zip.

List of tables and programs

The provided code reproduces:

- ☒ All numbers provided in text in the paper
- ☒All tables and figures in the paper
- ☐ Selected tables and figures in the paper, as explained and justified below.

Table 4 displays the output generated by each output generated by the estimation routine. All of these files are also provided in the replication package.

Table 4. Output files available in this replication package

Figure / Table #	Program	Line #	Output file	Time (hrs)
Statistical analysis				
Table 1	EASI Estimation_s.do	1046	Elasticity.xls	1
Table 2	EASI labour_Estimaton_s.do	825	output_labour_EASI.xml	0.3
Table 3	EVS simulation_Estimation_s.do	867, 886	table_3_1.xls, table_3_2.xls	0.1
Table B1	EASI Estimation_s.do	324	Summary.xls	0.1
Table B2	EASI Estimation_s.do	871	output_EASI.xml	0.5
Table B3, B4	EASI Estimation_s.do	1055	Elasticities.xls	1
Figure B1	EVS simulation_Estimation_s.do	829	Several PNG files	0.2
Figure B2	EVS simulation_Estimation_s.do	838	Several PNG files	0.2
Simulation				0.8
Tables 4,5 Figures 1, 2, 3,4 ,5, C1, C2, C3,C4, D1, D2, D3, D4, E1, E2, F1,F2, F3	Policy_simulations.gms, Third_best_replication.xlsx		Several tables and graphs in Excel format	

Acknowledgements

This work was done while Miguel Tovar was a postdoctoral researcher at the ZEW—Leibniz Centre for European Economic Research. He thanks the Federal Statistical Office of Germany for granting access to the data back then.