Code Documentation

Banning Products with Low Energy Efficiency: Market Response and Adjustment Cost

February 16, 2024

Thiess Buettner, Anne Kesselring

Contents

1 Data Sources **1**

1.1 Unit Sales and Attribute Information (GfK) 1

1.2 Variables 1

1.3 Generation of Estimation Sample 2

1.3.1 Bunching Analysis 2

1.3.2 Spatial Analysis 2

1.4 Simulation for Welfare Analysis 3

2 Code Description **4**

2.1 Stata "∗.do" Files for Tables and Figures 4

2.1.1 Data and Bunching 4

2.1.2 Spatial Analysis 4

2.1.3 Welfare Analysis 5

2.2 Additional Output for Robustness 5

2.3 Other Information 7

3 Software Information **7**

3.1 Requirements 7

3.2 Run Time 7

1 Data Sources

1.1 Unit Sales and Attribute Information (GfK)

The data set is provided by the market research company Gesellschaft für Konsumforschung (GfK) Retail and Technology GmbH headquartered in Nuremberg, Germany. For contact details, see https://www.gfk.com/en-us/home. Due to a non-disclosure agreement and the proprietary nature of the data, the data set for this project cannot be made publicly available. Interested researchers can request the data directly from the company subject to the company's approval and conditions. The specific data we are working with is GfK's POS Retailer Panel for major domestic appliances, specifically white goods (Gesellschaft für Konsumforschung Retail and Technology GmbH). The data we employ in estimation are for washing machines, reported at monthly frequency, covering the perid from 2004 to 2017.

1.2 Variables

The initial data set contains the following variables with relevance for the paper:

1. Unit sales: Sum of unit sales over all retailers for a given product in a given country, monthly frequency.
2. ID: Product-specific identifier over time and across countries.
3. Month.
4. Year.
5. Country: Austria, Germany, Hungary, Poland, Czech Republic, Croatia, Slovenia.
6. Product attributes: The two characteristics we make use of are a product's energy consumption (in kWh/year) and rated capacity (in kg).
7. The assignment variable, the Energy Efficiency Index EEI, is not reported in the GfK data. We calculate it from energy consumption and capacity according to the formula in the regulation (Commission Delegated Regulation (EU) No 1061/2010, see Online Appendix B.2).

All auxiliary variables used for the spatial analysis and the welfare simulation are obtained from this initial data set.

1.3 Generation of Estimation Sample

All programming and data work for this paper was performed using STATA statistical software. The following explains how the data sets for estimation are created and combined. Section 2 then describes the code pieces used for estimation.

1.3.1 Bunching Analysis

The file `bunching_1_bins.do` aggregates the original product-level data to bins on the basis of the EEI. To run the file, the directories in lines 30-52 must be adjusted. The resulting data set `bunching_2_collapse_1512.dta` is the data used for the subsequent bunching analysis.

1.3.2 Spatial Analysis

The final sample for the spatial analysis is built in a series of six steps. These steps could be combined, but given the novelty of the methodology, they were kept separate to facilitate cross-checks between the co-authors. Each step generates an intermediate file for the next step. Section D.4 of Online Appendix explains the process in more detail and contains additional step-by-step instructions that explain how the spatial analysis could be adapted for other contexts (beyond replication).

1. The file `spatial_1_raw_s.do` builds the lattice of the product space from the product-level raw data (directories in lines 13 and 19 must be changed).

2. The file `spatial_2_collapse_s.do` completes the lattice structure and collapses the data at two levels as intermediate data sets (directories in lines 18 and 21). First, at the region-year level (saving of *.dta in line 307). This intermediate data set is `spatial_2_collapse_s.dta`, which is the input file for step 3 below and the welfare analysis later in the paper. Second, at month-country-region level (saved in line 406): `spatial_2_collapse_mc.dta`. This is the input for step 4 below.

3. The third step is the implementation of the four different definitions of neighborhood in separate do-files with the common stub `spatial_3_*` (see Online Appendix D.3). The files build the neighborhood structure, and save the corresponding matrices in *.dta format. `spatial_3_fan.do` implements the V-shaped sectors (directories line 16-27, saving in line 333 and 348). `spatial_3_beam.do` implements the strip-based neighborhood (directories line 17-28, saving in line 313 and 335). `spatial_3_4corners.do` implements the intersection neighborhood (directories line 16-27, saving in line 331 and 346). `spatial_3_grid.do` implements the step function, i.e. grid-based neighborhood (directories line 16-27, saving in line 229 and 353).

4. The file `spatial_5_transform_s.do` generates the time lags of the sales variable with the intermediate data at monthly level (input from step 2). The order between steps 4 and 5 is switched here due to a change in the process from an earlier version. Note that file `spatial_5*` has to be run before `spatial_4*`.

5. The file `spatial_4_spgen.do` generates the spatial lag variables for the V-shaped, strips, and intersection neighborhoods. This first requires the transformation to *.stswm format (directories in lines 16 and 28). This is a special format for spatial objects in Stata that can be used in combination with the `spgen` command for spatial transformation of variables. Before transformation, the available matrices have to be called up into memory (commented out in lines 30 to 103, must be activated if directories change). This do-file generates the final version of `spatial_regs_s_monthly.dta`, on which the regression is run. The file `spatial_4_spgen_grid.do` generates the spatial lag variables for the step function, i.e. the grid-based neighborhood, using the same process, resulting in `spatial_regs_grid_monthly.dta` as the final file for the regression.

6. The final estimation samples were obtained from the previous step: `spatial_regs_s_monthly.dta` and `spatial_regs_grid_monthly.dta`. Before regression, the outcome and spatial variables are transformed using the logarithmic and inverse hyperbolic sine transformations.

1.4 Simulation for Welfare Analysis

We have two different levels of granularity in the welfare analysis. For the 3D-graphs and contour plots, we work with a parameter space of 20x20, i.e. 20 values of κ, and 20 of σ. This is generated in `welfare_simulation_20x20.do` (directories in lines 11 and 17). The file creates four matrices for four outcomes of interest: V (private utility), E (energy savings), P (assignment to restricted segment) and W (sales weights), that are saved. These have values of σ as rows and values of κ as columns in the data structure (wide format).

For the figures in Table 5, we work with a finer granularity. This data is not saved separately, the simulation is created with the do-file `welfare_analysis_table.do` from the input file `spatial_2_collapse_s.dta`, which is product space data at the region-year level (see step 2 in previous section). There is no additional matrix structure needed for this step.

2 Code Description

The following section describes the contents of the code files that create the analysis. The do-files are in the folder `main analysis in paper`. The corresponding output logs (*.smcl format) are contained in the sub-folder `output files\logs`. The graphs (*.gph format) are provided in the sub-folder `output files\logs` and the tables (in *.tex format) are contained in the sub-folder `output files\tables`.

2.1 Stata "∗.do" Files for Tables and Figures

2.1.1 Data and Bunching

The file `descriptives.do` generates the descriptive statistics in Table 1, as well as Figures 1, 2, 3, and 7. The directories in lines 18-14, and line 94 have to be adjusted. The corresponding log file is `descriptives1512.smcl`.

The file `bunching_2_regress.do` generates the bunching analysis in Table 2. The directories in lines 30-37 have to be adjusted. The corresponding log file is `bunching_2_regress_2703.smcl`. The additional error structures in the Online Appendix, Table E-1 are implemented in the file `bunching_3_errors.do`. The directories in lines 21-28 have to be changed, the output log is `bunching_3_errors_2703.smcl`.

2.1.2 Spatial Analysis

The file `spatial_6_regs_s.do` implements the regressions for the V-shaped, strips, and intersection neighborhoods. To run the file, the directories in line 27 and 29 have to be adjusted. The corresponding log file containing output is `spatial_6_regs_s_2703.smcl`. Naming conventions are: S for V-shaped sectors, B for strips, C for intersection. For each of these three neighborhood types, the estimates have a numeric index for each value of ρ, e.g. $B1$-$B8$ for strips. In all cases, 2 indicates the horizontal, and 6 the vertical direction.

The file `spatial_6_regs_grid.do` implements the regressions for the grid-based neighborhoods (directories in lines 26 and 28). The corresponding log file is `spatial_6_regs_grid_2703.smcl`. Naming conventions are W (west) for directions in the second quadrant (i.e. increasing size), and O (east, German: Ost) for directions in the first quadrant (i.e. decreasing size). The numeric indices indicate the grid search pattern, e.g. $W5$ is an interval of 5 cells in the step function (see Online Appendix D.3.4). For the two polar cases, H is the horizontal, and R is the vertical direction.

These two do-files described above create the output in Tables 3 and 4 of the paper and the robustness checks in Online Appendix E, with the exception of the $RMSE$ and $Vuong$ statistics (see 2.2 below).

2.1.3 Welfare Analysis

The file `welfare_analysis_table.do` generates the content in Table 5. We choose the values of σ at even-spaced intervals, and then iteratively search for the κ that delivers the same energy savings. The simulation takes the 2011 (pre-ban) yearly values (directory in line 16 must be adjusted), and calculates welfare effects based on the policy adjustment in accordance with the theory section (Equations (5), (6), and (7) in the paper). This process was initially done iteratively and manually (try and error) to match the energy savings to those of the actual regulation at given σ. We started

with two decimal places of κ and then narrowed down until we matched energy savings to the nearest integer. The do-file here only contains the final parameter combinations that we found and that are reported in Table 5. The corresponding output log is `welfare_analysis_table.smcl`.

The file `welfare_3D_plot_complete.do` generates the Figures 8, 9, and 10. This includes both the 3D plot (command `3dplot` has be installed first), and the contour plots. The file takes the 20x20 matrices of the four outcomes (wide format, see section 1.4). They are first merged by parameter value into a single data set, and the data is re-arranged to long format to simplify plotting. For Figures 8, and 9, the plotted values require no further transformation (except scaling for visual representation, see below). For Figure 10, the outcomes E and V are transformed to the ratio of changes in both quantities. In the long format, this is simplified with the use of panel notation (command `xtset`). Changes are computed as "lags" in the outcome between subsequent values of σ.

For the 3d plots, the y-axis (with the outcome) has to be scaled to obtain the cubic shape. The scale factors differ by outcome, and are simply calculated by hand outside of the do-file based on the range of the respective variable. For the contour plots, there is no scaling needed. Minor edits to the labels and rotation were done manually in the *.gph format before saving the graphs as the picture files contained in the paper. For the black and white version of the 3D graphs, additional graphic editing software was used to improve the representation of the color scheme when they are printed in shades of gray.

2.2 Additional Output for Robustness

As part of the revision, we conducted additional tests and statistics for the spatial analysis that were added to the existing Tables. This concerns the cross-validation, Vuong-Tests, country interactions, and overlapping matrices (additional spatial lags) referred to in section 5.2. The do-files for these items are in the folder `revision additional tests`, including the logs containing the output. We keep them in a separate folder and follow the naming conventions for the steps in section 1.3.2 for cases where we had to reconstruct several steps of the spatial process to adjust the data. The naming conventions of the estimates are kept the same as in the main results of the spatial analysis.

1. The file `spatial_6_regs_s_vuong.do` implements the Vuong test as an alternative model selection criterion for the V-shaped, strips, and intersection neighborhood (directories in line 40 to 59). The file produces the test statistics reported in the Tables 3 and 4, as well as Tables E-3, E-4, E-5 of the Online Appendix. The log file `spatial_6_regs_s_1502_vuong.smcl` contains the full output, a summary is reported for simplicity in `results_summary_s_1502_vuong.smcl`.

2. The file `spatial_6_regs_s_grid_vuong.do` implements the Vuong test for the step function (directories in lines 30-40, and 49). This produces the test statistics in the lower panel of Table 4, and Table E-6 of the Online Appendix. The log file

spatial_6_regs_s_grid_1702_vuong.smcl contains the full output, a summary is reported for simplicity in results_summary_s_grid_1702_vuong.smcl.

3. The file spatial_6_regs_s_xval.do implements the cross-validation as an alternative model selection criterion for the V-shaped, strips, and intersection neighborhood (directories in line 27-34, and 46). The file produces the $RMSE$ reported in the Tables 3 and 4, as well as Tables E-3, E-4, E-5 of the Online Appendix. The log file spatial_6_regs_s_1602_xval.smcl contains the full output, a summary is reported for simplicity in results_summary_s_1602_xval.smcl.

4. The file spatial_6_regs_s_grid_xval.do implements the cross-validation for the step function (directories in line 29-40, and 48). This produces the $RMSE$ in the lower panel of Table 4, and Table E-6 of the Online Appendix. The log file spatial_6_regs_s_grid_1702_xval.smcl contains the full output, a summary is reported for simplicity in results_summary_s_grid_1702_xval.smcl.

5. The file spatial_6_regs_s_overlap.do generates the robustness check with the additional spatial lags in Online Appendix Tables E-11, E-12, E-13 (directories in lines 30 to 41 must be changed). The corresponding log file is spatial_6_regs_s_0102_overlap.smcl. To achieve this, the do-file spatial_4_spgen_0102_overlap.do must first be run to re-create the data with additional spatial lags without overlap (see step 5 of section 1.3.2). Directories to change are in lines 24-32, the log file is spatial_4_spgen_overlap_0102.smcl.

6. The file spatial_6_regs_s_grid_overlap.do generates the same additional spatial lags for the step function in Online Appendix Table E-14 (directories in lines 30 to 41). The corresponding log file is spatial_6_regs_s_grid_1702_overlap.smcl. To achieve this, the do-file spatial_4_spgen_grid_1702.do must first be run to re-create the data with additional spatial lags (see step 5 of section 1.3.2). Directories to change are in lines 24-34, the log file is spatial_4_spgen_overlap_grid_1702.smcl.

7. The file spatial_6_regs_countries.do conducts the robustness test with country interaction reported in the last paragraph of section 5.2 (text, no separate table). The directories in lines 27-46 must be adjusted. The log file with the output is spatial_6_regs_s_2103_countries.smcl.

8. The file spatial_6_regs_s_gap.do excludes products close to the compliance line as a robustness check. This is not mentioned with specific numbers in the paper, but referred to as a concern in the penultimate paragraph of section 5.2. To achieve this, the do-file spatial_4_spgen_gap.do must first be used to re-create the adjusted spatial lags (see step 5 of section 1.3.2). Both do-files have log-files with the same name in *.smcl format that contain the output.

2.3 Other Information

Note that the remaining Figures 4,5, and 6 in the paper are conceptual, and do not contain empirical data. This also holds for the Online Appendix Tables B-1, C-1.

3 Software Information

3.1 Requirements

The analysis was done in STATA (version 15, later upgraded to 16). We employ the following additional packages, which are downloaded from the Boston College Statistical Software Components (SSC) archive: `carryforward` to fill in the grid structure, `matsave` for the intermediate handling of the weight matrices, `spgen` to create the spatial lag variables, `reghdfe` to run the regressions, and `graph3d` to create the 3D-plots. To run the spatial analysis, the maximum variable size has to be increased. We changed the default permanently to 30000.

3.2 Run Time

The analysis was initially run on a Dell Latitude E7470 notebook (make year 2017). The run time of the entire script including all variations is about 4 hours in total. On a new desktop computer, the time would be reduced to less than 2 hours. The most time-consuming steps are the files `spatial_3_*.do`, due to the creation of the neighborhood matrices. These files take approximately 15min each. The subsequent creation and loading of the spatial weight matrices in `*.stswm` format is less time intensive, but takes up substantial memory. These steps may become unresponsive when calling data from a server connected through VPN. The time needed to iteratively identify the combinations in Table 5 is not included in the run time estimate.