READ ME for replication of results in

Incumbency Advantages: Price Dispersion, Price Discrimination and Consumer Search at Online Platforms

February 22, 2023

This document describes how we obtained and constructed the data used in our study, as well as the production of the tables and graphs in the paper. The code is written in Stata 14.

Data sources

We use the following source data files:

- From ene't we received
  - Incumbent base tariffs
    GV_HH_3500_kWh_NEU.csv
  - Cheapest overall tariffs
    HH_3500_kwh_NEU.csv
  - Incumbent cheap tariffs
    BGV_HH_3500_kWh.csv
  - Data indicating whether there is an overlap of incumbency areas in a zip code
    2012_PLZ_ORT_Haupt_GV_weitere_GV.xlsx
  - Search queries for electricity tariffs
    wbs_csv_windows_1252_tb_Anfragen.csv
  - Search queries for gas tariffs
    WSB__keiner_31.12.2011.csv

WSB_2012.csv

WSB_2013.csv

WSB_2014.csv

- From Acxiom we received
  - Socio-economic data

    ZEW_Liefer_2009-2014_neu.xls
  - Google Trends data for the term "Strompreisvergleich"

    multiTimeline.csv

1 Script

Steps:

1. We first import the csv file *HH_3500_kWh_NEU.csv* which was extracted from ene't based on a specific request from us. It contains all the data on the cheapest electricity rates overall as well as costs components. The data are at the zip code month level and cover the period 2011 to 2014. We collapse the data to the year zip code level. Save as *incumbent_base.dta*.
2. We then import the csv file *GV_HH_3500_kWh_NEU.csv* which was also extracted from ene't based on a specific request from us. It contains all the data on the incumbent baseline tariffs overall as well as costs components. The data are also at the zip code month level and cover the period 2011 to 2014. We collapse the data to the year zip code level. Save as *incumbent_cheap.dta*.
3. We then import the csv file *BGV_HH_3500_kWh.csv*. It was also extracted from ene't based on a specific request from us. It contains all the data on the cheapest tariffs offered by the local incumbents. The data are at the zip code year level already. Save as *cheapest_overall.dta*.
4. Next merge the three price data sets by zip code year. Save as *prices.dta*
5. Import the xlsx file *2012_PLZ_ORT_Haupt_GV_weitere_GV.xlsx* which indicates for each zip code whether the zip code is affected by an overlap of incumbency areas. Save as *multiple_incs.dta*.

6. Import the xls file *ZEW_Liefer_2009-2014_neu.xls* which contains socio-economic data at the year zip code level which we received from Acxiom based on a specific request. Save as *acxiom.dta*.

7. Import the csv file *wbs_csv_windows_1252_tb_Anfragen.csv* which contains the search queries for electricity tariffs at the zip code-timestamp level (timestamp is the time of the search query with second precision), keep only queries from private customers, only keep the first query per search session. Save data set *search_electricity_prep.dta* collapse the number of search sessions as well as the number of search sessions that start with a search query for a green electricity tariff (*mm_oeko*) at the zip code year level. Save as *search_electricity_final.dta*.

8. Next prepare the search data for gas tariffs. They have been received from Acxiom as four csv files, one for every year from 2011 to 2014. Import csv file *WSB__keiner_31.12.2011.csv* first, save as *search_data_gas_2011.dta*. Proceed analogously with the years 2012, 2013 and 2014. Append the four yearly data sets and save as *search_data_gas_all.dta*. Generate search sessions for gas tariffs using the variables enetsessionid, plz, ort, day and month. This is necessary because we detected that the same session ID is given to multiple sessions sometimes and in contrast to the search data for electricity tariffs we do not have information of the search order. We then keep only one search query per session. Then collapse the number of search queries for gas tariffs at the zip code year level. Save as *search_gas_final.dta*.

9. Load *prices.dta* and merge with *acxiom.dta*, *search_electricity_final.dta* and *search_gas_final.dta* by zip code year.

10. Compute search intensity for eletricity tariffs as the number of search queries for electricity tariffs in a zip code year minus those for green tariffs and divide by the number of households. Compute search intensity for gas tariffs as the number of search queries for gas tariffs in a zip code year divided by the number of households. Drop outliers.

11. Adjust prices for value added taxes by multiplying them by 1.19 (19% VAT).

12. Generate Figure 4 of the paper using *spmap*.

13. Manually generate variable *bfut* which contains information on the EEX base future 1-year ahead based on information from EEX. Compute costs and markups for the incumbent's

base tariff, the incumbent's cheapest tariff and the cheapest overall tariff using. Compute price dispersion, price discrimination and online price dispersion.

14. Merge *multiple_inc.dta* by zip code year and drop zip codes if there is an overlap of incumbency areas.

15. Compute price zones which are defined as areas which have the same incumbent and all three prices do not vary within the respective area.

16. Compute Hausman-type instruments as the average search intensity in zip codes surrounding the focal zip code but only consider surrounding zip codes that are not in the same price zone.

17. log-transform all variables.

18. Label variables as they appear in the paper.

19. Generate Figure 2.

20. Generate Figure 6 which consists of two panels. For the right panel use the data set *search_electricity_prep.dta* constructed in step 7. Collapse the number of search queries at the year month week level and generate an image showing the development of search queries per week over time. Next clear the data and import the Google Trends data for the term "Stronmpreisvergleich" using the csv file *multiTimeline*. Collapse at the year month level and generate and image showing the development of Google searches for the term "Strompreisvergleich" over time.

21. Generate Table 1 with summary statistics of all variables. Only consider those that are also used in the estimations.

22. Run regressions reported in Table 2 and Table 3 and generate the correspondig tables.

23. Run regressions reported in the Appendix and generate the corresponding Tables B1-B7.

2 Data access

The data used in the paper are from commercial providers like e'net and Acxiom and are not publicly available. Independent researchers, however, can be given access to the data subject

to the signing of a data usage contract. The usage contract allows independent researchers to access the data at ZEW Center for European Economic Research Mannheim, however, it is not allowed to take the data set away from ZEW,

Alternatively, the data can be purchased at a negotiated price from ene't (info@enet.eu) and Acxiom (dataguru_DE@acxiom.com)

Websites:

- ene't
  https://www.enet.eu/downloads/
- Acxiom
  https://www.acxiom.de/wp-content/uploads/2021/05/Data_Catalogue_Germany.pdf