Overview

The code in this replication package constructs all figures in 'Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market,' by Assad, Clark, Ershov, and Xu using Python, Matlab, and Stata. The 19 Python files described below clean and prepare the data. A Python and Matlab file, described below, allocates stations into geographical markets using a clustering algorithm. The Stata file described below executes the analysis and produces the 10 tables and 4 figures in the main text, and 16 tables and 10 figures in the online appendix. This is done using an additional 40 Stata files (also described below). The replicator should expect the code to run for at least 72 hours.

Data Availability and Provenance Statements

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.

Summary of Availability:

- Some data **cannot be made** publicly available.

Details on each Data Source:

- *Gas Price Data:*

E5, E10 and diesel gas price data from January 2016 to December 2018 can be downloaded from the website Tankerkönig in three pieces. There are two large data files that can be downloaded from: (1) https://creativecommons.tankerkoenig.de/history/history.dump.1.gz (2) https://creativecommons.tankerkoenig.de/history/history.dump.2.gz As well as daily price data and station data that can be downloaded from Azure: https://dev.azure.com/tankerkoenig/_git/tankerkoenig-data. These data need to be downloaded into the directory "…/replication_package/Data/tankerkoenig-data/". The data are available (with attribution) under a Creative Commons license (CC BY 4.0).

- *Intraday crude oil data:*

Intraday crude oil data can be purchased from the provider FirstRate from the website https://firstratedata.com/i/futures/CL. These data need to be downloaded into "…/replication_package/Data/Crude_data/" and unzipped.

- *Weather data:*

10-minute interval weather data for German weather stations from January 2016 to December 2018 can be downloaded from the German Weather Service website (dwd.de). The data are available for download in weather-station specific ZIP files from the online directory https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate

/10_minutes/air_temperature/historical/. The data should be downloaded into the directory "…/replication_package/Data/Weather/". These data are freely available (with attribution) – see https://www.dwd.de/EN/ourservices/cdc/cdc.html for more detail.

- *Daily Crude oil average prices and EU/US exchange rates:*

These data were downloaded from the US Federal Reserve website. Daily crude oil prices were downloaded from the webpage https://fred.stlouisfed.org/series/DCOILBRENTEU as "…/replication_package/Daily_Crude_Oil/Crude_Oil.csv". Daily EU/US exchange rates were downloaded from the webpage https://fred.stlouisfed.org/series/DEXUSEU as "…/replication_package/Daily_Crude_Oil/EUR_USD.csv". These data are free to use.

- *Wholesale Gas Price Data:*

These data were purchased from the German company O.M.R. OIL MARKET REPORT GmbH. They are only licensed to be used after purchasing. The data are provided in the form of .xls MS Excel files, one for each fuel type. The files need to be reformatted (into "long" from "wide" format) and downloaded into the folder "…/replication_package/Data/Wholesale Price Data/".

- *1990s Gas Station Electronic Payments Adoption Data (Kent Data)*

These data were previously purchased by one of the authors of this paper (Robert Clark) from Kent Marketing (now Kalibrate), for the paper Carranza, Clark and Houde (JIE, 2015). This data is not publicly available and must be purchased from Kalibrate.

- *Autobahn Gas Station Data*

This is data identifying gas stations that are located on the Autobahn (German highway). This information was ccraped from the website https://www.raststaetten.de using code in the file ".../replication_package/Data/Autobahn/Autobahn_GasStations_Scrape.ipynb". Data was saved into the same folder in the csv file "…/replication_package/Data/Autobahn/all_AB_stations.csv".

- *Broadband Speed Data*

The data comes from the EU Commission's Joint Research Centre netBravo initiative (https://netbravo.jrc.ec.europa.eu/en/OpenData/Index). Raw data is crowd-sourced. Users throughout Europe either download a mobile app or use a web application to test their speed and the strength of their broadband signal. Raw data of individual tests is aggregated monthly and geographically by to the 1km×1km grid level. More details about the raw data structure are provided by the JRC (http://netbravo.jrc.ec.europa.eu//assets/netBravo/Open%20Data%20Format.pdf). Aggregated data from August 2016 to the end of our sample (December 2018) is available to download through the OpenData initiative. This data is saved in .csv files in the folder "…/replication_package/Data/Germany Broadband Speed Data/".

- *Demographics Data*

Demographics data comes from Eurostat (https://ec.europa.eu/eurostat/home). Data for 2016-2018 is at the year- Nomenclature of Territorial Units for Statisticslevel 3 (NUTS3) level. Germany's NUTS3 regions all begin with the letters "DE" and follow with a three character alphanumeric code. When necessary, we also collected this information at the moreaggregate NUTS2 level ("DE" followed by a two character alphanumeric code). Eurostat data is free to use (with sourcing). The data is provided in .csv files in the folder: "…/replication_package/Data/Germany Demographics Data/".

- *Daily Weather Data (for estimation)*

Average daily temperature and precipitation data, as well as the standard error of daily precipitation and temperature, at the German weather-station/day level for 2016-2018 was collected from the German Weather Service website (dwd.de). The data are available to download from the folder https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/ These data are freely available (with attribution) – see https://www.dwd.de/EN/ourservices/cdc/cdc.html for more detail. The data have been downloaded as .csv files into the folder: "…/replication_package/Data/Germany Weather/".

Computational Requirements

Software requirements:

1. Python 3.8 or above (with packages numpy, pandas, pyarrow and parquet installed)
2. Stata 16 or above (with packages ftools, gtools, reghdfe, ivreghdfe, outreg2, estout, sutex2, egenmore, parmest, geodist installed)
3. Matlab 2022 or above

Memory and runtime requirements:
The data files involved are very large and data cleaning consumes substantial computer memory. A computer with at least 10 cores, >64GB RAM and at least 1 TB memory is recommended. Approximate time needed to reproduce the analyses on a standard 2021 desktop machine is at least 3 days.

Description of Programs/Code and Instructions to Replicators

The data are cleaned and assembled using Python code below, in order. The processed data are then analyzed using Stata code below. We describe what the replicator needs to do in each of the programs ahead of running the program files. All program files are located in the folder "…/replication_package/".

Data Cleaning/Preparation (in Python):

Python Instructions
In the init_path.py file, change the first line to replace "PATH_OF_CURRENT_FOLDER" by the current directory (…/replication_package/). Then run the Python script **"AP_master_file_python.py".** This script runs all of

the python scripts below in order. Note, the data files involved are very large and data cleaning consumes substantial computer memory. A computer with at least 64GB RAM is recommended.

AP_master_file_python.py runs the following files:

Data_1_Raw_Cleanup.py:
- download gas price data from the data dump as well as Azure
- basic clean up & save to HDF data files by day
- clean up gas station information

Data_2_Gas_Station.py:
- Merge gas station meta data extracted from both Tankerkonig & Azure

Data_3_Price_History.py:
- Creates price history for every gas station in the data

Data_4_Price_History_FullDay.py:
- Creates additional price history files

Data_5_Dates.py:
- Creates date files (including ids for weekends, holidays, etc).

Data_6_Gas_Station_Rival.py:
- Define gas station rivals

Data_7_Shocks_Weather.py:
- Define weather shocks: temperature

Data_8_Shocks_Oil.py:
- Define oil shocks

Data_9_Price_Change_Count.py:
- for each station, count the number of price changes per day.

Data_10_PC_Response.py:
- For each rival's price change, calculate how long it takes for the station to update its own price.

Data_11_Shocks_Supply_and_Price.py:
- Merge weather shocks & gas price: calculate responses to weather shocks
- Merge oil shocks & gas price: calculate responses to oil shocks
- Calculate weekly averages for all 4 shocks and save files

Data_12_structural_break_supply_shock_data_split.py:

- Split the Stata dta files into individual files for each station and measure (for parallelization)

Data_13_Hourly_Prices.py:
- Creates time specific price data (9am, noon, 5pm, 7pm).

Data_14_Price_Average.py:
- Creates average daily prices for all stations in the markets for all fuels.

Creating Geographic Markets based on driving time using a clustering algorithm:
- "Data_15_Market_Definition.py" creates a matrix of the driving time distance between all the stations – this file also calls on STATA file "Data_15a_travel_time.do" to estimate driving distances.
    - IMPORTANT: Need to change PATHS in Stata file.
- Then run "Data_15b_cluster_stations.m" file in Matlab. This file reads in the distance matrix and calculates the hierarchical clustering algorithm that produces the final market definitions. This creates a file where each station in the data is assigned to a geographic market.
    - IMPORTANT: Need to change PATH in line 3 and line 29 of the Matlab file to current project path.
- For convenience, the completed market file using our main specification is provided in "/replication_package/Data/Market/all/20_80.csv".

Data_16_Price_Reponsiveness.py:
- Creates price responsiveness data (number of times rival in the same geographic market responded within 5 minutes). Note: this requires the geographic market definitions generated above.

Data_17_Appendix_Big_Breaks.py:
- Defines adopting stations that experienced "big" structural breaks in the adoption markers. Note: prior to running this file, you need to run "Stata_3_break_combinations.do" before this file, since it defines the "baseline" set of adopters.

Data Analysis (in STATA):

STATA Instructions

In the AP_master_file_stata.do file, change the first line to replace "PATH_TO_MAIN_DIRECTORY_WHERE_THE_CODE_IS" by the current directory (…/replication_package/). Then run the script in STATA. This script executes the Stata files described below in order. All intermediary and final data files are saved into the
folder .../replication_package/Data/processed/. All figures are saved into the folder .../replication_package/Figures/. All tables are saved into the folder .../replication_package/Tables/.

AP_master_file_stata.do :
- runs following .do files:

Stata_1_date_concordance.do:

- Generates a file that creates a concordance between the dates used in Python (y2kw) and Stata's date/time.

Stata_2_structural_breaks.do:

- For each station, run QLR test for structural breaks for all four shocks for all fuel types.

Stata_3_break_combinations.do:

- Calculate the combinations of breaks happening within four weeks of one another for each station for each fuel type. Defines AP adopters in files such as "…/replication_package/Data/processed/adoper_measures_4week_`fuel'.do".

Stata_4_Data_Prep_.do:

- Prepares covariates for regressions. As part of that process, it runs files "Stata_4_Data_Prep_a_..." to "Stata_4_Data_Prep_j…". These files do the following:
  - Stata_4_Data_Prep_a_ZIP_NUTS3.do creates a DTA file with correspondence between the ZIP codes that stations belong to and NUTS3 regions.
  - Stata_4_Data_Prep_b_demographics.do creates a DTA file of NUTS3 based demographics.
  - Stata_4_Data_Prep_c_wholesale_e5.do creates a DTA file of daily e5 wholesale prices by region and matches it to gas stations.
  - Stata_4_Data_Prep_d_weather.do creates a DTA file of daily local precipitation and temperature averages and standard deviations and merges it to gas stations.
  - Stata_4_Data_Prep_e_broadband_speed.do creates a DTA file of geographic /time period based broadband speeds.
  - Stata_4_Data_Prep_f_mean_hourly_prices.do creates a DTA file of daily gas station prices at different points during the day (9am, noon, 5pm and 7pm).
  - Stata_4_Data_Prep_g_price_responsiveness.do creates a DTA file which calculates the number of times a station responds to its rival increasing or decreasing their price within 5 minutes.
  - Stata_4_Data_Prep_h_top5_brands.do creates a DTA file listing the top 5 brands in the market (by the number of gas stations) and all of the stations belonging to them.
  - Stata_4_Data_Prep_i_nearest.do creates a DTA file with the distance between each station j and its nearest rival.

- Stata_4_Data_Prep_j_autobahn.do creates a DTA file of the stations that are on the Autobahn (German highway).

After this, the file combines average daily prices, wholesale prices, gas station location/brand, and the various covariates calculated above. Then, the file collapses the data to monthly averages. The file also does the same, but for duopoly and triopoly markets, and collapses that data into monthly averages as well. The final files used for estimation (for the main text results) are "…/replication_package/Data/processed/monthly_duo_trio_mkt_data_e5_merged.dta" and "…/replication_package/Data/processed/monthly_station_data_e5_merged.dta"

Stata_5_main_text_results.do:

- This file calculates summary statistics (Tables 1 and 2) and the regressions show in the tables in the main text of the paper (Tables 3-10). The tables are all saved into the folder: "…/replication_package/Tables/".

Stata_6_main_text_figures.do:

- This file performs the analysis that produces the figures in the main text (Figures 1-3, and Figure A1) and saves them into the folder ".../replication_package/Figures/".

Stata_7_appendix_Figure_B1_B2.do:

- This file creates figures B1 and B2 (based on the Lekkerland stations that adopted dynamic pricing in convenience stores).

Stata_7_appendix_Figure_C1.do:

- This file creates Figure C1 which shows the distribution of stations across the geographic markets.

Stata_7_appendix_Figure_D1.do:

- This file creates a descriptive figure of an example of a station with and without a structural break in the number of price changes (Figure D1).

Stata_7_appendix_Figure_D2.do:

- This file creates a figure (Figure D2) showing the difference in structural break markers between stations that experienced structural breaks and stations that did not.

Stata_7_appendix_Figure_D3.do:

- This file creates a figure (Figure D3) showing the distribution of QLR test Fstats for a structural break test in the number of price changes for four stations in the data (stations 22, 29, 31 and 163).

Stata_7_appendix_Figure_D4.do:

- This file creates Figure D4, which shows a histogram of pairwise combinations of structural breaks in markers.

Stata_7_appendix_Figure_D5:
- This file creates Figure D5, which shows the distribution of AP adoption shares by market in December 2018.

Stata_7_appendix_Figure_D6:
- This file creates Figure D6, which are histograms of structural break distributions for Diesel gas for the number of price changes and responsiveness to local rivals.

Stata_7_appendix_Figure_F1:
- This file creates Figure F1, which shows a time series of the share of Canadian gas stations belonging to different brands that adopted electronic payment technology in the 1990s.

Stata_7_appendix_Table_D1:
- This file creates Table D1, which tests the difference in covariates between brands based on their adoption shares.

Stata_7_appendix_Table_E1:
- This file creates Table E1, which estimates the "flat" regressions (2 time periods for each gas station, averaging all the data before mid-2017 and after mid-2017).

Stata_7_appendix_Table_E2:
- This file creates Table E2, which estimates weighted OLS regressions, where each observation is weighted by an inverse of the average of the noise in the structural break tests.

Stata_7_appendix_Table_E3:
- This file creates Table E3, which estimates the effects on a non-adopting station in a duopoly/triopoly market of their rivals adopting AP.

Stata_7_appendix_Table_E4:
- This file creates Table E4, which estimates the effects of AP adoption on prices at different points during the day, but not breaking up the sample into monopolist and non-monopolists (as in the main text).

Stata_7_appendix_Table_G1:
- This file creates Table G1, which performs various robustness checks on the sample used in the main text (e.g., not including Shell stations, not including highway stations).

Stata_7_appendix_Table_G2:
- This file creates Table G2, which replicates Table 5 in the main text, but uses a market definition based on ZIP codes rather than clustering.

Stata_7_appendix_Table_G3:

- This file creates Table G3, which replicates Table 7 in the main text, except using a market definition based on ZIP codes rather than clustering.

Stata_7_appendix_Table_G4:

- This file creates Table G4, which replicates Table 5 in the main text, except using a variety of different definitions of who is an adopter of AP (e.g., only stations with structural breaks in both Diesel gas and E5 gas). NOTE: BEFORE THIS FILE CAN RUN, IT'S IMPORTANT TO RUN THE FOLLOWING PYTHON FILE: Data_17_Appendix_Big_Breaks.py.

Stata_7_appendix_Table_G5:

- This file creates Table G5, which replicates parts of Table 5 in the main text, except using alternative instruments based on different geographic classifications: e.g., the share of brand adopters outside of station j's NUTS3 area.

Stata_7_appendix_Table_G6:

- This file creates Table G6, which replicates Table 5 in the main text, except using an alternative instrument based on broadband speed changes in the local area.

Stata_7_appendix_Table_G7:

- This file creates Table G7, which replicates Table 5 in the main text, except using an alternative "placebo" instrument based on the adoption shares of a brand other than station j's brand.

Stata_7_appendix_Table_G8:

- This file creates Table G8, which replicates Table 5 in the main text, except using E10 fuel prices and wholesale prices (and adoption based on E10 fuel structural breaks).

Stata_7_appendix_Table_G9:

- This file creates Table G9, which replicates Table 5 in the main text, except using diesel fuel prices and wholesale prices (and adoption based on diesel fuel structural breaks).

Stata_7_appendix_Table_G10:

- This file creates Table G10, which tests alternative standard error clustering approaches (e.g., clustering at the NUTS3 level, rather than the market level).

Stata_7_appendix_Table_G11:

- This file creates Table G11, which replicates the results of Table 5 in the main text, except it pools the monopolist and non-monopolist sample together, rather than estimating separate regressions for monopolist and non-monopolist stations.

List of Tables and Code:

In the main text, there are 10 tables (Tables 1-10) and 4 figures (Figures 1-3 and A1). In the online appendix, there are 16 tables and 10 figures.

The list of tables and figures, and where they are produced, is below.

Table/ Figure #	Program	Line Number
Table 1	Stata_5_main_text_results.do	11-18
Table 2	Stata_5_main_text_results.do	34-76
Table 3	Stata_5_main_text_results.do	90
Table 4	Stata_5_main_text_reuslts.do	102-127
Table 5	Stata_5_main_text_results.do	136-159
Table 6	Stata_5_main_text_results.do	173-186
Table 7	Stata_5_main_text_results.do	194-221
Table 8	Stata_5_main_text_results.do	242-262 (for top panel), and 263-283 (for bottom panel)
Table 9	Stata_5_main_text_results.do	293-331
Table 10	Stata_5_main_text_results.do	338-349
Figure 1	Stata_6_main_text_figures.do	7-8
Figure 2	Stata_6_main_text_figures.do	68-86 (panel a), 141-160 (panel b), 206-224 (panel d), 279-298 (panel c)
Figure 3	Stata_6_main_text_figures.do	389-400 (panel a), 427-430 (panel b)
Figure A1	Stata_6_main_text_figures.do	445 (panel a), 451 (panel b), 458 (panel c), 464 (panel d)
Figure B1	Stata_7_appendix_Figure_B1_B2.do	90
Figure B2	Stata_7_appendix_Figure_B1_B2.do	188
Figure C1	Stata_7_appendix_Figure_C1.do	45
Figure D1	Stata_7_appendix_Figure_D1.do	36
Figure D2	Stata_7_appendix_Figure_D2.do	298
Figure D3	Stata_7_appendix_Figure_D3.do	114

Figure D4	Stata_7_appendix_Figure_D4.do	31
Figure D5	Stata_7_appendix_Figure_D5.do	12
Figure D6	Stata_7_appendix_Figure_D6.do	18
Figure F1	Stata_7_appendix_Figure_F1.do	23
Table D1	Stata_7_appendix_Table_D1.do	31
Table E1	Stata_7_appendix_Table_E1.do	54-73
Table E2	Stata_7_appendix_Table_E2.do	76-82
Table E3	Stata_7_appendix_Table_E3.do	51-57
Table E4	Stata_7_appendix_Table_E4.do	18-38
Table G1	Stata_7_appendix_Table_G1.do	14-35, and 107
Table G2	Stata_7_appendix_Table_G2.do	47-76
Table G3	Stata_7_appendix_Table_G3.do	97-102
Table G4	Stata_7_appendix_Table_G4.do	148-269, and 427-432
Table G5	Stata_7_appendix_Table_G5.do	98-106
Table G6	Stata_7_appendix_Table_G6.do	6-27
Table G7	Stata_7_appendix_Table_G7.do	11-19
Table G8	Stata_7_appendix_Table_G8.do	225-260
Table G9	Stata_7_appendix_Table_G9.do	225-260
Table G10	Stata_7_appendix_Table_G10.do	102-118
Table G11	Stata_7_appendix_Table_G11.do	17-22

Acknowledgements:

None.