

# Details for Replication

---

This folder contains data and code for the replication of “Competition in Pricing Algorithms,” authored by Zach Y. Brown and Alexander MacKay and published in the American Economic Journal: Microeconomics.

The file `/analysis/scripts/master_00.do` is the master script that can be used to replicate the results. The master script is set up for a replication from pre-processed data files. Code files that pertain to importing and cleaning the data (01-06) have been commented out of the master script. The rest of the analysis (07-17) will run. A log of the partial run is available in `/analysis/logs`. Logs from the R and Matlab scripts are generated as separate files.

The folder contains `/analysis` (data, logs, packages, simulation, and scripts) and `/paper` (tables and figures).

For additional comments, see below.

## Data and Code Availability

---

### 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.

### License for Data

The data are licensed under a Creative Commons Non-Commercial (CC-BY-NC) license. See [LICENSE.txt](#)

([/C:/Users/Alex/Dropbox/Papers/016%20Pricing%20Algorithms/replication/LICENSE.txt](#)) for details.

### 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

## High Frequency Price Data

The high-frequency data was collected (scraped hourly) from retailer websites and processed by the authors. It contains a panel of prices for specific products on different websites. Product characteristics were constructed by hand from the product descriptions.

As the specific identity of the retailers is not necessary for the analysis, and, as per guidance by our university counsel, we have removed variables that directly identify the retailers for the replication package, replacing them with placeholders. To avoid directly identifying the retailers, the raw scraped data cannot be made available, nor can the manually constructed dataset of product details. Certain product details are unique to specific retailers. This file (`product_list/product_list_20191004.xlsx`) is called in code file 06. We have included cleaned versions of the code files 01-06, but we have not provided the data necessary to run these files.

The cleaning steps from the raw data include loading from the raw html, selecting the relevant brands and products for each brand, and imputing missing prices for gaps of less than 6 hours. The processed data has a flag (equal to 1) when prices are imputed. Researchers that wish to use different imputation procedures may drop these observations and replace with an alternative.

Details: see Code.

File provided:

- `/analysis/data/analysis_data.dta`

Details on some specific variables:

- `website` is a placeholder for the actual website name
- `website_id` provides the website-specific code a particular item
- `period_id` provides the specific hour in the data
- `date` provides the date
- `hour` provides the hour
- `price` records the observed retail price
- `drug` records the specific drug (chemical)
- `brand` records the brand
- `form` records the form in which the drug is ingested

- `variant` records the specific version of the drug-brand-form (by decongestant-childrens-is12hour)
- `multipack` records the number of product packages bundled together
- `size` records the number of total units (= `multipack` \* `count`)
- `count` records the number of units in a product package
- `childrens` indicates whether it is a children's version of the drug
- `is24hour` indicates it is a 24-hour drug
- `is12hour` indicates it is a 12-hour drug
- `decongestant` indicates that it is blended with a decongestant
- `product_id` records the specific product (by drug-brand-form-variant-multipack-size)
- `product_website_id` records the specific product for a specific retailer (by `product_id`-website)
- `flag_missing_price` indicates that the price was not captured in that specific period
- `flag_imputed_price` indicates whether the price was imputed according to the procedure described in the online Appendix
- `is_observed` records whether or not price was observed in that period (and not imputed)
- `price_change` indicates whether or not we identify a price change for that product relative to the price observed in the previous hour

Amazon. 2019. "Price data for various products, 2018-2019". <https://amazon.com> (<https://amazon.com>).

CVS. 2019. "Price data for various products, 2018-2019". <https://cvs.com> (<https://cvs.com>).

Target. 2019. "Price data for various products, 2018-2019". <https://target.com> (<https://target.com>).

Walgreens. 2019. "Price data for various products, 2018-2019". <https://walgreens.com> (<https://walgreens.com>).

Walmart. 2019. "Price data for various products, 2018-2019". <https://walmart.com> (<https://walmart.com>).

## Google Search Data

We collect the average share of searches for each retailer based on Google Trends data (<https://trends.google.com/trends/> (<https://trends.google.com/trends/>)) over the sample period, specifically, weeks beginning 4/15/2018 through 9/29/2019 inclusive.

File provided:

- `/analysis/data/google_search_shares.csv`

The first column is a search for the retailer name, the second is a search for the retailer name + “allergy”, and the third column provides an average of the two. The third column is used in the calibration exercise. Similar to the price data, the retailer name was replaced with a placeholder.

Google. 2020. “Google Trends.” <https://trends.google.com/trends/?geo=US> (<https://trends.google.com/trends/?geo=US>). Accessed February 25, 2020.

## E-commerce Revenue Data

E-commerce revenue data for the five retailers was obtained from a subscription to ecommerceDB.

File provided:

- `/analysis/data/ecommerce_revenue_data.xlsx`

The included file has 2018 revenues, revenue growth, and estimated revenues for 2019 in columns 2 through 4. In addition to the revenue figures, the included file also has the relative price index, which is calculated from specification (1) of Table 3 in the manuscript. The price index is used to adjust revenues to quantities and calculate market shares. These are provided in columns 5 through 7. Similar to the price data, retailer names were replaced with placeholders.

ecommerceDB. 2019. “Top 100 Online Stores - U.S. - Personal Care - 2018.” <https://ecommercedb.com/> (<https://ecommercedb.com/>).

## Software

- The replication was run using Stata version 16.1, R version 3.6.3, and Matlab R2020b.
- Stata packages are pre-installed in the replication folder: carryforward, estout, ftools, gr0034, reghdfe, regsave, st0085\_2, tabout, texsave, tsspell
- R packages need to be installed before replication: foreign, data.table.
- Matlab packages need to be installed before replication: optimization toolbox, parallel computing toolbox.

## Code

- 
- The code used to run the analysis is stored in /analysis/scripts.
  - Running 00\_master.do will run the analysis scripts in order by the two-digit prefix. Please see the comments in 00\_master.do for a list of the code files.
  - Code files 01-06 have been commented out, as they require raw data that is not provided as per recommendation from university counsel. In the included scripts, any references to variables or files that would directly identify specific retailers have been modified or removed.
  - The Matlab files run in a self-contained folder (/analysis/simulation) that is executed from the 00\_master.do file.
  - Before running, it is necessary to set the global variables \$BASE to the project directory, \$R\_PATH to the system path for R, and \$M\_PATH to the system path for Matlab in 00\_master.do. See the script for more details.
  - Matlab scripts will automatically start a parallel pool consistent with the settings for the parallel computing toolbox.

## Tables and figures

---

- paper/tables and paper/figures contain the tables and figures of the manuscript and online appendix after running 00\_master.do.
- All tables are in .tex format and all figures are in .pdf format.
- Please note that some versions of Matlab may yield slightly different numbers for Table 6.
- Full list of files corresponding to the main article tables and figures, with the relevant script indicated in square brackets (by its two-digit prefix):
  - Table 1: daily\_stats [07]
  - Table 2: price\_response\_72\_a, price\_response\_72\_b [09]
  - Table 3: price\_diff [08]
  - Table 4: elast\_pricing\_alg [13]
  - Table 5: stub\_welfare [13]
  - Table 6: stub\_mergers\_with\_C\_panel\_a, stub\_mergers\_with\_C\_panel\_b [14]
  - Figure 1: xyzal\_tablet\_80, claritin\_tablet\_70 [07]
  - Figure 2: price\_change\_fraction\_hourofweek\_A, price\_change\_fraction\_hourofweek\_B, price\_change\_fraction\_hourofweek\_C, price\_change\_fraction\_hourofweek\_D, price\_change\_fraction\_hourofweek\_E [08]
  - Figure 3: pr\_price\_response\_D\_1, pr\_price\_response\_D\_2 [09]
  - Figure 4: plot\_technology\_premium [08]

- Figure 5: (tikz figure in LaTeX file)
  - Figure 6: (tikz figure in LaTeX file)
  - Figure 7: price\_3firm, price\_3firm\_merger [15]
  - Figure A1: (tikz figure in LaTeX file)
  - Figure B1: equilibrium\_selection\_asymmetric, equilibrium\_selection\_symmetric [12]
- Full list of files corresponding to the Online Appendix tables and figures.
    - Table C1: website\_brand [07]
    - Table D1: tab\_shipping\_costs [17]
    - Table E1: table\_3firm\_bn, table\_3firm\_pa [15]
    - Table G1: retailer\_shares (Table created by hand, by combining the data in google\_search\_shares.csv with proprietary data on revenues from ecommerceDB.com, as described in the text)
    - Table G2: stub\_calibrated\_mass [13]
    - Figure C1: count\_products\_weekly [07]
    - Figure E1: plot\_sim\_technology\_premium, price\_3firm\_post\_merger [15]
    - Figure F1: (tikz figure in LaTeX file)
    - Figure G1: pr\_price\_response\_E\_1, pr\_price\_response\_E\_2 [09]
    - Figure G2: plot\_sim\_prices\_by\_tech, plot\_sim\_shares\_by\_tech [16]