

“Quantifying the Sources of Firm Heterogeneity”

Code Documentation

November 13, 2015

Throughout we use Stata version 12.1 .do and .dta files (except the Monte Carlo, which is done in Matlab version 2015b). You will also need to have Perl (available by default on Linux or Mac, and as a free download on Windows¹) to properly compile the table output. Our codes make occasional use of user-written commands. These can be added to your copy of Stata with the command `ssc install XXX`, where XXX is the name of the user-written package. You will need `egenmore`, `tabout`, `estout`, `ivreg2`, `labutil`, `unique`, `ranktest`, `parmest`, `roman`, `lmnwhitenl`. If you are having trouble running the code, with Stata informing you that something is an “unrecognized command”, the most likely explanation is that you are missing one of these packages.

The codes expect a certain directory structure, which can be saved to your machine by unzipping the package `MPF_files_QJE.zip`. The archive will map out the required directories starting from a specified location (probably where you save the .zip file). This base location should be given to the Stata code through the global `TERMINAL` in the first line of each do file. If you want to run straight through everything, you should be able to send the file `do_all.do` directly into Stata *after* you have updated the `TERMINAL` global in *each* do file. The following sections describe in more detail how you can replicate individual results.² When .do files are ordered (whether with numbers or letters in the file name), they must be run in that order. Please be aware that the replication will require a substantial amount of computing time, on the order of 3–5 days. It is especially time consuming to (1) import the raw data, (2) run the counterfactual experiments for Table 12, and (3) run the grid search for Online Supplement Figures S1–S3.

1 DATA CLEANING AND ESTIMATION

There are three sources of raw data used in this project. The main results rely on data from AC Nielsen, via the Kilts Marketing Center at the University of Chicago’s Booth School of Business. These files should be placed in a directory corresponding to their observation year; for example, the 2001 data are placed in `data/KILTS/nielsen_extracts/HMS/2001/Annual_Files/panelists_2001.tsv`. You must have the following annual files: `panelists_YYYY.tsv`, `purchases_YYYY.tsv`, `trips_YYYY.tsv`. You also need the Nielsen file `data/KILTS/nielsen_extracts/HMS/Master_Files/`

¹ <http://strawberryperl.com/>

² Our code was run in a Linux environment. If you are working in Linux with multiple users, be sure that everyone has the needed read/write/execute permissions for the files in the MPF directory. Information on the relevant `chmod` command can be found here: <http://ss64.com/bash/chmod.html>.

Latest/products.tsv. The second source of raw data is the GS1 database. You must place the file gs1.dta in the directory data/GS1.³

To assemble the datasets used for the descriptive statistics and estimation, run the following files in order (so, for example, 1_RawTSV_to_Stata.do before 2_Kilts_Aggregation.do):

1. 1_RawTSV_to_Stata.do – Loads raw TSV files (as downloaded from the Kilts Marketing Center portal) and saves them in Stata’s .dta format. Also constructs the file unmatchedupc.dta from the Kilts and GS1 data.
2. 2_Kilts_Aggregation.do – Aggregates the raw annual files from household–city–week to the national–quarterly level, then combines into a single dataset.
3. 3_Kilts_Construction.do – Cleans the combined data
4. 4_GMM_SigU_Prep.do – Prepares data for estimating σ^U , the elasticity of substitution between UPCs
5. 5_GMM_SigU.do – Estimates σ^U using a nonlinear least squares representation of the Broda and Weinstein (2006) methodology, as discussed in Section 5.2 of the paper.
6. 6_GMM_SigF_Prep.do – Prepares data for estimating σ^F , the elasticity of substitution between firms
7. 7_GMM_SigF_2SLS.do – Estimates σ^F using instrumental variables, as discussed in Section 5.3 of the paper.
8. 8_MarginalCosts.do – Collects estimation outcomes and prepares data for tables and figures used in paper

2 PAPER TABLES

To generate the tables of results reported in the paper, run the following files (all in the directory codes/PostEstimation). The numbers below correspond to the number of the table in the main text. If there are multiple .do files under a given number, they must be run in the order (so, for example, SampleStats1_A.do before SampleStats1_B.do).

All output files are saved in the directory results/tables with the file name given in square brackets. To view these tables, please use the included file MPF_Replication_Tables.tex to compile the L^AT_EX sources. If you are only interested in (or have only reproduced) some of the tables, simply comment (by adding a “%” to the beginning of the relevant line) the input statements for the ones you do not care about. Please see inside the file MPF_Replication_Tables.tex for more detailed explanation of this process.

1. (a) SampleStats1_A.do – Prepare sample statistics
- (b) SampleStats1_B.do – Table of sample statistics [1_sample_stat_Scol.tex]

³ The third source is only needed for the Chile results in the Supplementary Appendix (Tables S2–S3). See section 5 below.

2. (a) RankTables_A.do – Prepare by-decile size distribution (with type loop set to Decile)
(b) DecileRank_B.do – Size distribution by decile [2_size_distro_decile2_Scol.tex]
3. (a) RankTables_A.do – Prepare by-decile size distribution (with type loop set to Firm)
(b) FirmRank_B.do – Size distribution by firm size rank [3_size_distro_firm_rk_Scol.tex]
4. (a) SampleStat2_A.do – Prepares size distribution by number of UPCs
(b) SampleStat2_B.do – Table of size distribution by number of UPCs [4_sample_stat2_Scol.tex]
5. GMM_Distro.do – Distribution of GMM estimates [5_gmm_distro_Scol.tex]
6. VD_Correlations.do – UPC-level sales component correlations [6_vd_corrs_upc_noscale.tex]
7. VD_Correlations.do [same as 6.] – Firm-level sales component correlations [7_vd_corrs_noscale.tex]
8. (a) MarkupTables_A.do – Prepare markup distribution
(b) MarkupTable_B.do – Output table for distribution of markups [8_markup_distro_Scol.tex]
9. (a) MarkupTables_A.do – Prepare markup distribution (same as Table 8)
(b) RelativeMarkupTable_B.do – Distribution of large-firm markups relative to product group [9_relative_markups_Scol.tex]
10. (a) VD_XSection.do, VD_Growth.do – Conduct variance decomposition of sales components
(b) VD_Tables.do – Report and tabulate results [10_var_decomp3_v5_Scol.tex]
11. (a) NestTest_1.do – Compute results under different models of nesting structure
(b) NestTest_2.do – Run within-nest regressions
(c) NestTest_3.do – Collect regression results
(d) NestTest_4_Table.do – Report within-nest R^2 [11_nest_test_Scol.tex]
12. (a) Counterfactual_I_B.do, Counterfactual_I_C.do, Counterfactual_II_B.do, Counterfactual_II_C.do – Conduct counterfactual exercises
(b) CounterfactualTable.do – Report counterfactual results [12_counterfactual_profit_v7.tex]

3 FIGURES

To generate the figures reported in the paper, run the following files (all in the directory codes/PostEstimation). The numbers below correspond to the number of the figure in the main text. All output files are saved in the directory results/figures with the file name given in square brackets.

1. VD_Graphs.do – Sales decomposition by firm rank [Fig1_top50_share_v5.eps]

- Need to run `VD_XSection.do` and `VD_Growth.do` first
- 2. `AggBiasOutput_Figure_A.do` – Bias in conventional real output share measure [`Fig2_agg_bias_qje.eps`]

4 MISCELLANEOUS RESULTS IN THE TEXT OF THE PAPER

To generate miscellaneous results reported in the paper, run the following files (all in the directory `codes/Misc`). All output files are saved in the directory `results/tables` with the file name given in square brackets.

- The basic summary statistics on the data described in
 - “around 20,000 firms that supply goods with bar codes in the Nielsen HomeScan Database in a typical quarter.”
 - “Our database covers approximately 1.6 million goods purchased at some point by households in our sample.”
 - “On average, 440,000 different UPCs were sold each quarter. ”

are compiled by `FirmUPC_counts.do` [`M1_counts.tex`]

- The cannibalization statistics given in
 - “On average across product groups, we find a cannibalization rate for the median firm of around 0.50”
 - “However, we find an average cannibalization rate across all deciles of the firm size distribution that ranges between 0.50 and 0.51, which reflects the fact that even in the upper decile of the size distribution most firms have trivial market shares. In contrast, when the largest firm within a sector (which has an average market share of 22 percent) introduces a new product, we estimate that 62 percent of the sales of that product comes from from the sales of its existing products. ”

are generated by the sequence:

1. `Markup-Cannibalization_A`
2. `Markup-Cannibalization_Decile_B` [`M2_markup_cannib_decile_Scol.tex`]
3. `Markup-Cannibalization_Rank_B` [`M3_markup_cannib_rank_Scol.tex`]

- Footnote 13 –

1. `MarketDefinition_1_DataCity.do`
2. `MarketDefinition_2_Result.do`
3. `MarketDefinition_3_Table.do` [`M4_MarketDefinition.tex`]

- “Ninety-nine percent of firms in our sample have aggregate market shares (across all product groups) of less than 0.1 percent of total bar-code sales. Even the largest firm only sells 3 percent of total bar-code sales in our sample.” – `FirmMarketShareAggregate.do` [`M5_FirmShare.tex`]

- Footnote 23 – WhiteQualityTime.do [M6_white.tex]
- “Moreover, most of the elasticities are precisely estimated—in 82 percent of the cases, we can statistically reject the hypothesis that $\sigma_g^U = \sigma_g^F$ at the 5% level” – SigFSigU_Difference.do [M7_sigmaDiff.tex]
- Footnote 28 (and corresponding passage in main text) – QualityTime.do [M8_qualitytime.tex]
- The difference in relative variance statistics described in
 - “The variance of relative appeal is 16 percent larger than the variance of relative marginal cost.”
 - “In the firm-level results, the stronger correlation between sales and appeal is driven by the fact that the variance of firm appeal is 60 percent larger than that of firm average marginal cost.”

are calculated by RelativeVariances.do. You must have run VD_Correlations.do (responsible for Paper Tables 6–7) first. [M9_RelativeVariance.tex]

- “we regressed the importance of product upgrading for firm growth against how much product turnover occurs in a product group.”
 1. Turnover_A.do
 2. Turnover_B.do
 3. Turnover_C.do – [M10_VDturn_reg_small.tex]
- Footnote 36 –
 1. SigmaU_BrandDifference_A.do
 2. SigmaF_BrandDifference_B.do
 3. Sigmas_BrandDifference_C.do – [M11_BrandDiffMeds.tex]
- Footnote 37 –
 1. Sigmas_LongDifference_A.do
 2. Sigmas_LongDifference_B.do [M12_LongDiffMeds.tex]
- “If one regresses $Q_{fgt}^F / \sum_{f \in \Omega_{gt}^F} Q_{fgt}^F$ on $Q_{fgt}^F / \sum_{f \in \Omega_{gt}^F} Q_{fgt}^F$, one only obtains an R^2 of 0.77 with a coefficient of 0.65 (s.e. 0.004).” – AggBiasOutput_Table_B You must have run AggBiasOutput_Figure_A.do (which produces Figure 2) first. [M13_aggbias_reg.tex]
- *From the Online Supplement:* “Our estimate of the cannibalization rate for a firm in the carbonated beverages sector ($\sigma^F = 5.18$ and $\sigma^U = 7.82$) ranges from 39 percent to 51 percent assuming the firm has a market share of under 20 percent and introduces a product that has a market share equal to the typical product of that firm.” – Cannib_CarbBev.do [SM14_can_carbBev.tex]

5 ONLINE SUPPLEMENT TABLES

To generate the tables of results reported in the online supplement, run the following files (all in the directory codes/Appendix). All output files are saved in the directory results/tables with the file name given in square brackets.

To replicate Tables S2–S3, you must have access to the Chilean data through Datamyne. If you have this access, we will provide you with the relevant data queried from Datamyne in the text file `chile_final_brands.csv`. This file must be in the Chile directory.

- S1. `MPF_montecarlo.m` – Conducts a Monte Carlo exercise for parameter values and generates the results in Table S1 [`S1_MonteCarloResults.log`]. You must set the `TERMINAL` variable in line 12 to the same root directory you use for the Stata `TERMINAL` global. This `.m` file needs to be in the same directory as the files `epsfirm.m`, `estimsigmaF.m`, `estimsigmaU.m`, `expshare.m`, `markup.m`, `pfirm.m`, and `pgroup.m`.
- S2. (a) `Chile_Estimation.do` – Estimates model on Chilean data up through the cross-sectional decomposition, gives distribution of GMM results.
(b) `Chile_Estimation_Tables.do` – Outputs table [`S2_est_distro_chile_Scol.tex`]
- S3. (a) `Chile_VD.do` – Variance decomposition for Chilean data (must have run the file `Chile_Estimation.do` first)
(b) `Chile_VD_Tables.do` – Outputs table [`S3_var_decomp_chile_Scol.tex`]
- S4. `LitComparisonTable.do` – Produces table of previous literature elasticity estimates, from the manually formatted CSV file `lit_est.csv` (included in the .zip replication archive)[`S4_lit_comps.tex`]
- S5. (a) `VD_PG_XSection.do` – Performs variance decomposition by product group
(b) `VD_PG_Top10_Table.do` – Reports VD results for top 10 selling product groups [`S5_PGVD_table_top10.tex`]
- S6. `PG_List.do` – Lists product groups [`S6_pgroups_descr.tex`]

6 ONLINE SUPPLEMENT FIGURES

To generate the figures reported in the online supplement, run the following files (all in the directory codes/Appendix). All output files are saved in the directory results/figures with the file name given in square brackets.

- S1. (a) `GridSearchDecomp.do` – Grid searches elasticity parameters and computes resulting sales decompositions
(b) `GridSearchHistograms.do` – Produces histograms [`FigS1_gridsearch_firmdemand.eps`]
- S2. Same as S1 [`FigS2_gridsearch_scope.eps`]
- S3. Same as S1 [`FigS3_gridsearch_resid.eps`]

- S4. `MPF_montecarlo.m` – Histogram of Monte Carlo estimates [`montecarlosigU.pdf`]. The section of code that prints figures requires your computer system to have the required graphics card and drivers. If you are having trouble, you may also wish to look into the user-written program `export_fig` available at https://github.com/altmany/export_fig.
- S5. Same as S4 [`montecarlodelta.pdf`]
- S6. Same as S4 [`montecarlosigF.pdf`]