README for Reimers and Waldfogel, "Digitization and Pre-Purchase Information…"

Data Availability Statement

Our paper relies on two explicit datasets that we can share, one confidential dataset, and industry figures collected by hand.

1. Shareable dataset 1. Keepa – for asin-level price and sales rank, etc. We include the extract we used in the replication package. We have permission to share the extract we have provided. Researchers can obtain additional data from keepa.com, using their api.

2. Dataset 2. We use the Ni et al data on individual-level Amazon ratings for books. We used the dataset Books.csv – the ratings-only file for books – which we obtained from https://nijianmo.github.io/amazon/index.html. This file is listed under the heading, "Small" subsets for experimentation. The creators of the dataset would like users to obtain the data directly from them at the site noted. The file needed to replicate the results in the paper is called Books.csv (and is available at http://deepyeti.ucsd.edu/jianmo/amazon/categoryFilesSmall/Books.csv).

3. Confidential data source. Some of our programs read from confidential Nielsen data. The files the programs refer to, but which we do not include, are master_nielsen.dta, nielsen_A_B_data, nielsen_top100.dta. **We do include the parameters produced by these files, which allow other programs to run.** The files containing the relevant estimates are A_B_params.dta, A_B_params_bs.dta, and sigma.dta. The programs that read confidential data are: create_sigma.do, create_A_B.do, create_A_B_bs.do, and figure_4_nyt_over_time_effect.do. Researchers interested in obtaining these data should contact Nielsen/NPD for their Bookscan product. See the section below on Confidential Nielsen Data for detailed instructions for creating replication datasets from raw data.

4. Industry aggregates collected by hand. We also obtain reported in Publishers Weekly Magazine for weekly aggregate US book sales. The figures are reported in a weekly feature called "Weekly Scorecard," which is available in the print edition of Publishers Weekly. Each issue reports a recent week and the week one year before. We turned these figures into a dataset by manually inputting them, and we are making this file available.

To produce the tables and figures, first get the Books.csv file (see point 2 above). Then run the file _main.do, which runs each of the underlying Stata programs in an order that allows them all to function. Using the program run order table below, a user can see what has to be run before running particular programs.

All programs were run using Stata 15 and 16. The do files are indexed and run from a master file, main.do. Each constituent program produces output containing results from tables or figures, or mentioned in the text. The output takes the form of tables, figures, and log files.

Programs required (amoeba, distinct, etc) are installed in the setup.do file, which also creates subdirectories for output. If you have trouble installing amoeba, try typing the following into Stata's command line: `net describe sg71, from(http://www.stata.com/stb/stb38) .`

With the exception of the bootstrapping programs (which have the suffix bs), the programs all run within a few minutes on either a Dell Precision 5820 desktop machine with an i9_990X CPU and 32 Gb of RAM, or on a Dell OptiPlex 7050 with i7_770 CPU with 64Gb of RAM.

Confidential Nielsen Data

Instructions for creating `master_nielsen.dta`

1. Start with our "spine" (`nielsen_notables_spine.dta`) file containing ISBNs, review dates and Notable years.
2. Using your Nielsen subscription, download files for one book at a time, using the ISBN13s. For each ISBN, Nielsen delivers an XLSX file with multiple sheets. One of the sheets ("136.0 Search Results") contains the ISBN. Another sheet ("135.0 Title Sales and Rank ...") contains sales history.
3. The provided do file (create_nielsen_notables_data.do) will read from the XLSX files you download and will merge them with the spine we have provided.
4. The resulting file (`master_nielsen.dta`) can be used in the program `figure_4_nyt_over_time_effect.do`).

Instructions for making the Nielsen Top 100 (nielsen_top100.dta)

1. Using your Nielsen subscription, you can download the top 100 ISBNs and their associated weekly sales for each week as XLSX files. Do this for each of the weeks from 2015 - 2018.
2. The provided do file (create_nielsen_top100_data.do) will read from the XLSX files you download and will merge them to create the weekly top 100, 2015-2018 (nielsen_top100.dta).
3. The resulting file (nielsen_top100.dta) is used in the program create_sigma.do.

Instructions for making nielsen_A_B_data.dta

1. Append the Nielsen_top100.dta to master_nielsen.dta
2. Create a variable that allows the linking of 13-character ISBN13 to 9-character asins:

```
gen match = substr(isbn13,4,9)
```

3. Keep only the Notables from 2018 (notable_year==2018).
4. Save this as nielsen_A_B_data.dta. The resulting file is used by create_A_B.do and create_A_B_bs.do.

PROGRAM RUN ORDER	input files	what needs to run before	creates output
setup.do			
******			
* PARAMETER CREATION			
******			
create_A_B.do	main_amazon_sales.dta, confidential_data\nielsen_A_B_data.dta		data\A_B_params.dta
create_A_B_bs.do	main_amazon_sales.dta, confidential_data\nielsen_A_B_data.dta		data\A_B_params_bs.dta
			data\sigma.dta
create_sigma.do	confidential_data\nielsen_top100.dta		
******			
* TABLES - pt 1			
******			
table_1_sample_stats.do	main_amazon_sales.dta		
table_2_regressions.do	main_amazon_sales.dta		
******			
* INTERMEDIATE DATA CREATION			
******			
create_quantity_us_simulations.do	main_amazon_sales.dta, A_B_params.dta, sigma.dta, 3coregint.ster	table_2_regressions.do	
create_quantity_us_simulations_bs.do	quantity_us_simulations.dta, main_amazon_sales.dta, A_B_params_bs.dta, sigma.dta, 3coregint.ster	table_2_regressions.do, create_quantity_us_simulations.do	
create_quantity_us_simulations_50_quantiles.do	main_amazon_sales.dta, A_B_params.dta, sigma.dta, review_quantiles_50.ster	figure_3_causal_effects_stars_50_quantiles.do	

create_quantity_us_simulations_50_quantiles_bs.do	quantity_us_simulations.dta, main_amazon_sales.dta, A_B_params_bs.dta, sigma.dta, review_quantiles_50.ster	figure_3_causal_effects_stars_50_quantiles.do, create_quantity_us_simulations.do	
******************************			
* TABLES - pt 2			
******************************			
table_3_quantity_effects.do	main_amazon_sales.dta, A_B_params.dta, estimates\3coregint.ster, tempfiles\quantity_us_simulations.dta	create_quantity_us_simulations.do, table_2_regressions.do	
table_3_quantity_effects_bs.do	main_amazon_sales.dta, A_B_params_bs.dta, estimates\3coregint.ster, tempfiles\quantity_us_simulations_bs_X.dta (X=1,…,500)	create_quantity_us_simulations_bs.do, table_2_regressions.do	
table_4_welfare_effects_baseline.do	tempfiles\quantity_us_simulations.dta	create_quantity_us_simulations.do	table_4_baseline_log.txt
table_4_welfare_effects_baseline_bs.do	tempfiles\quantity_us_simulations_bs_X.dta (X=1,…,500)	create_quantity_us_simulations_bs.do	table_4_baseline_bs_log.txt
table_4_welfare_effects_50_quantiles.do	tempfiles\quantity_us_simulations_50_quantiles.dta	create_quantity_us_simulations_50_quantiles.do	table_4_50_quantiles_log.txt
table_4_welfare_effects_50_quantiles_bs.do	tempfiles\quantity_us_simulations_50_quantiles_bs_X.dta (X=1,…,500)	create_quantity_us_simulations_50_quantiles_bs.do	table_4_50_quantiles_bs_log.txt
table_4_welfare_effects_reviewed_books.do	tempfiles\quantity_us_simulations.dta	create_quantity_us_simulations.do	table_4_reviewed_books_log.txt
table_4_welfare_effects_reviewed_books_bs.do	tempfiles\quantity_us_simulations_bs_X.dta (X=1,…,500)	create_quantity_us_simulations_bs.do	table_4_reviewed_books_bs_log.txt
table_4_welfare_effects_varying_sigma.do	tempfiles\quantity_us_simulations.dta	create_quantity_us_simulations.do	table_4_varying_sigma_log.txt
table_4_welfare_effects_varying_sigma_bs.do	tempfiles\quantity_us_simulations_bs_X.dta (X=1,…,500)	create_quantity_us_simulations_bs.do	table_4_varying_sigma_bs_log.txt
table_4_welfare_effects_marshallian.do	tempfiles\quantity_us_simulations.dta	create_quantity_us_simulations.do	table_4_marshallian_log.txt

table_4_welfare_effects_marshallian_bs.do	tempfiles\quantity_us_simulations_bs_X.dta (X=1,…,500)	create_quantity_us_simulations_bs.do	table_4_marshallian_bs_log.txt
******************************			
* FIGURES			
******************************			
figure_2_daily_review_effects.do	daily_param.ster	table_2_regressions.do	results\nyt_effect_daily.pdf,results\non_nyt_effect_daily.pdf
figure_3_causal_effects_stars_50.do	main_amazon_sales.dta		review_quantiles_50.ster
figure_4_nyt_over_time_effect.do	confidential_data\master_nielsen.dta		results\nyt_notables_effect.pdf
******************************			
* Numbers in main text			
******************************			
text_numbers_main_sections1_5.do	main_amazon_sales.dta, tempfiles\quantity_us_simulations.dta, A_B_params.dta, A_B_params_bs.dta, data\Books.csv, and some confidential data files		text_section_1_through_5_log.txt
text_welfare_effects_varying_R_squared.do	main_amazon_sales.dta, A_B_params.dta, sigma.dta, 3coregint.ster, tempfiles\quantity_us_simulations.dta	table_2_regressions.do	text_welfare_effects_varying_R_squared_log.txt, results\text_welfare_effects_varying_R_squared.pdf
text_welfare_effects_WOM.do	tempfiles\quantity_us_simulations.dta, data\Books.csv	create_quantity_us_simulations.do	text_welfare_effects_WOM_log.txt
text_numbers_section7.do	main_amazon_sales.dta		text_numbers_section7_log.txt
******************************			
* Numbers in appendix text			
******************************			

appendix_welfare_effects_double_B.do	main_amazon_sales.dta, A_B_params.dta, sigma.dta, estimates\3coregint.ster	table_2_regressions.do	appendix_welfare_effects_double_B_log.txt
appendix_welfare_effects_2level_nl.do	tempfiles\quantity_us_simulations.dta	create_quantity_us_simulations.do	appendix_welfare_effects_2level_nl_log.txt
appendix_welfare_effects_representativeness.do	tempfiles\quantity_us_simulations.dta	create_quantity_us_simulations.do	appendix_welfare_effects_representativeness_log.txt
**********			
* Appendix figures			
**********			
appendix_figure_1_sales_rank_illustration.do	main_amazon_sales.dta		results\appendix_sales_rank_illustration.pdf
appendix_figure_2_genre_distribution.do	main_amazon_sales.dta, A_B_params.dta		results\appendix_genre_distribution.pdf