

This document describes the data and programs for AEJApp-2012-0138, “Demand Spillovers, Combative Advertising, and Celebrity Endorsements.”

The main data for this study come from Nielsen BookScan (hereafter BookScan). BookScan is the premier source for title level data on book sales in the United States. As described in the paper, these data are assembled from a large panel of booksellers representing the majority of the United States market. The primary use of these data is for marketing and planning purposes by publishers, booksellers and authors. As a result of their potential commercial applications, these data are proprietary and a condition of the data use agreement is that these data cannot be shared with any party and must be turned back over to BookScan after the completion of the project. Access to the BookScan data was obtained by contacting Jonathan Stolper at Nielsen BookScan. His contact information is below:

Jonathan Stolper  
VP, General Manager  
Nielsen BookScan  
The Nielsen Company  
770 Broadway, 8th Floor  
New York, NY 10003  
Office: 646-654-4781  
Cell: 917-568-2190  
[jonathan.stolper@nielsen.com](mailto:jonathan.stolper@nielsen.com)

Access to the Bookscan data involves the use of their proprietary website. Unfortunately, this website is not intended to be used for mass downloads of data, and therefore it is a bit cumbersome to assemble a large dataset of title level sales. The terms of service with BookScan for accessing the data explicitly do not allow for any automated scraping of the data. However, the data can be searched by author, title or ISBN number. Each ISBN number is unique to a particular edition of the book, and therefore to gather a complete record of sales you must collect data for each ISBN. For some books this is relatively trivial, i.e. they have 2 ISBN numbers for the pre and post Oprah Club editions of the title. For many of the books, particularly the older selections there are a large number of ISBNs that have been for sale over time (this is particularly true for Dickens and Tolstoy). On the BookScan website, data can be output to a CSV file for 8 weeks at a time for each ISBN number. These can be combined together for a complete sales record for each title over time. This CSV file can obviously then be read into any statistical software for analysis (the accompanying code is for Stata). In addition, the BookScan website contains national aggregate sales data by genre. These data can be downloaded in the same manner, however there is only one series of data for each genre as well as for adult fiction, fiction, etc. The attached programs from generating the final results are intended to be used on a version of the data that follows the BookScan format.

I have included the program AEJApp-2012-0138\_Program.do. This State Do File provides the appropriate coding for the raw BookScan data to indicate the weeks in which there were endorsements of texts by Winfrey or the Today Show. This is true for the panel of book data as well as the aggregate sales data. The code then generates the main figures and tables for the results of the paper.