Data Appendix for "Exclusive Preferential Placement as Search Diversion: Evidence from Flight Search"

Benjamin Edelman

Harvard Business School
Boston, MA 02163
bedelman@hbs.edu

Zhenyu Lai

Department of Economics
Harvard University
Cambridge, MA 02138
zlai@fas.harvard.edu

April 28, 2013

1 Data Description

Our data originates from ComScore Search Planner, a commercial database product tracking online behavior of a panel of internet users. More information on the dataset is available at: `http://www.comscore.com/Products/Audience_Analytics/Search_Planner`. ComScore Search Planner provides the destination URL and corresponding internet traffic volumes of outgoing US search engine users, grouped by their exact search queries used.

We compile our dataset using the following steps:

1.1 Identify Search Query Terms

First, we identify the search query terms that led to outgoing search engine traffic to 17 popular online travel agencies (OTA) and 6 US-based airline websites. These 23 websites are:

1. www.aa.com
2. www.airfarewatchdog.com
3. www.alaskaairlines.com
4. www.bookingbuddy.com
5. www.cheapflights.com
6. www.cheapoair.com
7. www.cheaptickets.com
8. www.delta.com
9. www.expedia.com
10. www.farecompare.com
11. www.farespotter.com
12. www.kayak.com
13. www.jetblue.com
14. www.lowfares.com
15. www.onetravel.com
16. www.orbitz.com
17. www.priceline.com
18. www.travelocity.com
19. www.travelzoo.com
20. www.tripadvisor.com
21. www.tripmama.com
22. www.southwest.com
23. www.united.com

We identify these search query terms for 4 bi-monthly time periods: May to June 2011, July to August 2011, January to February 2012, and March to April 2012. We intentionally omit the four months immediately surrounding the launch of GFS. The GFS result was displayed in a sidebar from mid-September to November before its full release with gradual phase-in. This made the appearance of GFS links unpredictable during the omitted time period.

The exact wording of all search queries that led to these websites for each of the given bi-monthly time periods was collected from the web interface under the "Site Profile" tab in

Comscore Search Planner. We keep only search queries for flights that mention a US destination. This is because Google began to offer its international flight search tool significantly later, specifically on March 15, 2012. We delete all search queries that either mention an international destination or airport code, or mention a phrase that disqualifies search intent for flights, for example, "flight attendant school". We also delete all search queries which directly specify a desired site, for example, "Orbitz" or "Jetblue".

1.2 Classification of Search Queries

Next, we classify these search queries into 12 groups. First, we classify each search query as either being "GFS-eligible" or "GFS-ineligible", based on whether the exact wording of the search query specified would trigger the exclusive preferential placement of GFS in a Google search after Decmber 2011. We verify this classification by conducting searches at `http://www.google.com`, both manually and by using automation.

Second, each search query expresses an intended flight destination, which we use to classify the search queries into six groups by geographic region. We divide the 50 US states into the following six geographic regions:

1. Midwest Region
    - Includes: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin
2. Mountain West Region
    - Includes: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming
3. Northeast Region
    - Includes: Conneticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont
4. Pacific West Region

- Includes: Alaska, California, Hawaii, Oregon, Washington

5. Southeast Region
   - Includes: Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia
6. Southwest Region
   - Includes: Arkansas, Lousiana, Oklahoma, Texas

Note that Washington DC is included into the Northeast region, and Puerto Rico is included into the Southeast region. We tested perturbations to our definitions of geographical aggregation, and results were robust to changes.

1.3 Outgoing Search Engine Traffic Volumes

For each of these classified groups of search queries, we obtain the outgoing non-paid algorithmic traffic volumes and paid advertising traffic volumes to each destination URL. We use the Search Planner interface to require that traffic originate from searches that correspond with an "exact" match of the specified search queries rather than a "broad" match, and that traffic occur during one of our specified bi-monthly time periods. We assume that each internet search concludes with a single click, and we infer the total number of searches as the sum of non-paid algorithmic and paid traffic volumes. For each of the 23 travel sites, we divide traffic volume by total searches to calculate the non-paid algorithmic and paid advertising click-through rates.

Table 1 (in the paper) describes the composition of searches in our sample.