

Data Sources

Homescan

The Nielsen Homescan data consist of a panel of households who record their grocery purchases. The purchases can come from a wide variety of store types, including traditional food stores, supercenters and warehouse clubs, and online merchants. Interested consumers who are 18 or older register online to participate (at <http://www.homescan.com>) and are asked to supply demographic information. Based on this information, Nielsen contacts a subset of the registered consumers to become panel members. They are not paid in currency for participating in the program, but every week a panel member who scans at least one purchase receives a set amount of points. The points can be redeemed for merchandise. Panelists can earn additional points for answering surveys and by participating in sweepstakes that are open only to panel members.

The data used was for trips made during 2004. The original data consisted of two panels. Members of the larger set of households, the “61K panel,” recorded all UPC-coded food purchases. A subset of these households, the “15K panel,” also recorded non-UPC coded products including fresh food purchases and other random weight items, such as fruit and vegetables). In what follows the focus is on the larger panel.

Each participating household was provided with a scanner. As part of setting up the scanner, the households recorded the stores they usually visit. For each shopping trip, the panelist recorded the date and the store, ideally from one of the previously programmed outlets. They then scanned the barcodes of the products they purchased, and entered the quantity of each item, whether the item was purchased at the regular or promotional (“deal”) price, and the coupon amount (if used) associated with this purchase.

Nielsen then matches the barcode, or UPC, with detailed product characteristics. The recording of price will turn out to be particularly important for this study. If the household purchased products at a store covered in the Nielsen store-level data (“ScanTrack”)—and we think (but could not verify) that all stores operated by the retailer who provided us with the data are covered in the store-level data—Nielsen did not require the household to enter the price paid for each item, in an effort to make the scanning process less time-consuming for the household. Instead, Nielsen imputes the price from the store-level data. To construct this price, we understand that Nielsen uses the average weekly price paid at the store for the corresponding item (UPC). If the same item could be transacted at different prices within the same store during the same week, this imputation process can introduce errors into the price data. A common reason for such price variation across transactions (of the same item within a store-week) is loyalty-card discounts that are only applied to the subset of consumers who use cards. Unless all consumers always use the card, the imputed price is unlikely to be the exact price paid by the consumer. This imputation leads to frequent, sometimes large, price reporting errors.^{3,4}

³Coupon use itself does not cause this problem. If the consumer purchased the product using a coupon, this value is reported in Homescan and can be subtracted from the reported price to get the actual consumer cost.

⁴For stores that are not covered by the store-level data, Nielsen asks households to report the price paid and checks the prices entered by consumers by comparing them to a range of prices observed elsewhere for the same or similar item. (We think those stores are not in our sample.) If a price is considered out of range, the median regional price is used instead. As an additional validity check, Nielsen also manually reviews transactions with high quantities and households who are at the top of the expenditure distribution in each category.

Retailer's data

The second data set comes from a large national grocery chain, hereafter referred to as the retailer. This retailer records all the transactions in all its stores. For each transaction, the data record the exact time of the transaction, the cashier number, and the loyalty card number, if one was used. The data also list the UPCs purchased, the quantity purchased of each product, the price paid, and information regarding discounts (loyalty card discounts, coupons, etc.). The retailer also links loyalty cards that belong to members of the same household, primarily by matching the street addresses and telephone numbers individuals use when applying for a loyalty card. The retailer then assigns each household a unique identification number. Clearly, this definition of a household is more prone to errors compared to Homescan's definition, in which a household is simply associated with the house at which the scanner resides.