

Data for Problem Set 3 (Effects of Horizontal Merger)

The data set used for this problem set consists of two Matlab files: [ps2.mat](#) and [iv.mat](#) (both Matlab 5+ files). If you plan to use Matlab use these files directly. If not, either use the Matlab "load" and "save" commands to create ASCII files or download the Excel spreadsheets that contain the data ([cereal_ps3.xls](#) and [demog_ps3.xls](#)). The data are (semi-fabricated) data on 24 brands of the only REAL product (ready-to-eat cereal, what else did you think?), for 94 markets (47 US cities for the first 2 quarters of 1988). These variables are defined and were treated as described in Nevo (2000), "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand," *Journal of Economics & Management Strategy* 9(4): 513-548. Note: these data should **NOT** be used to make any real inference about the demand for RTE cereal (or any other product).

The file **ps2.mat** contains the following variables:

id - an id variable in the format bbbbccyyq, where bbbb is a unique 4 digit identifier for each brand (the first digit is company and last 3 are brand, i.e., 1006 is K Raisin Bran and 3006 is Post Raisin Bran), cc is a city code, yy is year (=88 for all observations in this data set) and q is quarter. All the other variables are sorted by date city brand.

id_demo - an id variable for the random draws and the demographic variables, of the format ccyyq. Since these variables do not vary by brand they are not repeated. The first observation here corresponds to the first market, the second to the next 24 and so forth.

s_{jt} - the market shares of brand *j* in market *t*. Each row corresponds to the equivalent row in **id**.

x1 - the variables that enter the linear part of the estimation. Here this consists of a price variable (first column) and 24 brand dummy variables. Each row corresponds to the equivalent row in **id**. This matrix is saved as a sparse matrix.

x2 - the variables that enter the non-linear part of the estimation. Here this consists of a constant, price, sugar content and a mushy dummy, respectively. Each row corresponds to the equivalent row in **id**.

v - random draws given for the estimation. For each market 80 iid normal draws are provided. They correspond to 20 "individuals", where for each individual there is a different draw for each column of **x2**. The ordering is given by **id_demo**.

demogr - draws of demographic variables from the CPS for 20 individuals in each market. The first 20 columns give the income, the next 20 columns the income squared, columns 41 through 60 are age and 61 through 80 are a child dummy variable (=1 if age ≤ 16). Each of the variables has been demeaned (i.e. the mean of each set of 20 columns over the 94 rows is 0). The ordering is given by **id_demo**.

The file **iv.mat** contains the variable **iv** which consists of an id column (see the **id** variable above) and 20 columns of IV's for the price variable. The variable is sorted in the same order as the variables in **ps2.mat**.

Excel spreadsheets:

cereal contains 2256 observations on id, brand, firm, city, quarter, share, price, sugar content, mushiness, and the 20 instruments in **iv**, called **z1-z20**.

demog contains the demographic draws for each market. There are 94 observations (47 cities by 2 quarters) and 80 variables (20 individuals X 4 variables).