Econ 512
Fall 2011

Estimation of Nevo's Random Coefficients Demand Model

1 The Model

1.1 Utility Function

Utility consumer i gets from product j in market t:

$$U_{ijt} = X_j\beta_i + \alpha_i P_{jt} + \xi_j + \Delta\xi_{jt} + \epsilon_{ijt}$$

There are $t = 1, ...T$ markets, $j = 1, ...J$ products in each market, $i = 1, ...I_t$ consumers in market t. In the empirical application we will work with there are T=94 markets (47 cities and 2 time periods), J=24 brands, and $I_t = 20$ consumers.

X_j is 3 observable brand characteristics [constant, sugar content, mushy dummy]. They are the same across markets.

P_{jt} is brand price. It can vary across markets.

ξ_j is an unobserved (to us) product characteristic that is the same across markets

$\Delta\xi_{jt}$ is an unobserved (to us) product characteristic that varies across markets- deviation from ξ_j. This represents across-market variation in the utility from a brand.

ϵ_{ijt} is an unobserved consumer characteristic. Varies across individuals

1.2 Random Coefficients - consumer heterogeneity

For consumer i: $$\begin{bmatrix} \beta_i^1 \\ \beta_i^2 \\ \beta_i^3 \\ \alpha_i \end{bmatrix} = \begin{bmatrix} \beta^1 \\ \beta^2 \\ \beta^3 \\ \alpha \end{bmatrix} + \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} \\ \pi_{31} & \pi_{32} & \pi_{33} & \pi_{34} \\ \pi_{41} & \pi_{42} & \pi_{43} & \pi_{44} \end{bmatrix} \begin{bmatrix} D_i^1 = inc \\ D_i^2 = inc^2 \\ D_i^3 = age \\ D_i^4 = child \end{bmatrix} + \begin{bmatrix} \sigma_{11} & 0 & 0 & 0 \\ 0 & \sigma_{22} & 0 & 0 \\ 0 & 0 & \sigma_{33} & 0 \\ 0 & 0 & 0 & \sigma_{44} \end{bmatrix} \begin{bmatrix} \nu_i^1 \\ \nu_i^2 \\ \nu_i^3 \\ \nu_i^4 \end{bmatrix}$$

$$\begin{bmatrix} \beta_i^1 \\ \beta_i^2 \\ \beta_i^3 \\ \alpha_i \end{bmatrix} = \begin{bmatrix} \beta \\ \alpha \end{bmatrix} + [\Pi D_i + \Sigma\nu_i]$$

The D_i are observable consumer characteristics (their distribution across consumers can be measured) and the ν_i are unobserved consumer characteristics. In the empirical work, there are 20 consumers in each market.

Assume that ϵ_{ijt} are iid random variables with type 1 extreme value distribution,

Assume that $\nu_i \sim N(0, I)$ and assume that ϵ_{ijt}, ν_i, and D_i are independently distributed.

1.3 Market Shares

Utility can now be rewritten as:

$$
\begin{aligned}
U_{ijt} &= \begin{bmatrix} X_j & P_{jt} \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} + \xi_j + \Delta\xi_{jt} + \begin{bmatrix} X_j & P_{jt} \end{bmatrix} (\Pi D_i + \Sigma \nu_i) + \epsilon_{ijt.} \\
&= \delta_{jt}(X_{j,} P_{jt}, \xi_j, \Delta\xi_{jt}; \alpha, \beta) + \mu_{ijt}(X_{j,} P_{jt}, D_i, \nu_i; \Pi, \Sigma) + \epsilon_{ijt.}
\end{aligned}
$$

The first term δ_{jt} is the mean utility of brand j in market t. It is composed of two terms $X_j\beta + \xi_j$ that only vary across brands and two terms $P_{jt}\alpha + \Delta\xi_{jt}$ that vary across markets and brands. The brand-specific terms can be combined into a single brand-level coefficient γ_j (we will see how to recover the β and ξ_j parameters from the coefficients on product dummy variables later).

$$
\begin{aligned}
\delta_{jt} &= X_j\beta + \xi_j + P_{jt}\alpha + \Delta\xi_{jt} \\
&= \gamma_j + P_{jt}\alpha + \Delta\xi_{jt}
\end{aligned}
$$

Each individual consumer is described by a set of 8 characteristics, D_i and ν_i, and a set of 25 product shocks $\epsilon_{i0t}, \epsilon_{i1t}, ...\epsilon_{i24t.}$ The market share for product j in market t is constructed by adding over the consumers who prefer product j to any of the other choices. This is done using both a distributional assumption on ϵ_{ijt}, that provides an analytical solution for part of the integration, and simulation for the other consumer characteristics.

1.4 Estimation

The coefficients to be estimated are $\theta_1 = (\gamma_{1,}\gamma_2, ...\gamma_J, \alpha)$ which enters the mean utility for each brand and $\theta_2 = (\Pi, \Sigma)$ which accounts for the consumer heterogeneity in preferences. We will use a GMM estimator that minimizes the correlation between a set of M instrumental variables $Z = (z_{1,}...z_M)$ and a structural error term - $\Delta\xi_{jt}$ the shocks to the mean utility of each market/brand. Notice that the error term only enters the mean utility of each brand/market.

The moment conditions are $E(z_m\Delta\xi) = 0$ for $m = 1, ..., M$. There will usually be more moment conditions than parameters to estimate (in our case there will be 44 instruments, prices in 20 other time periods and 24 product dummies. In the case where there are observable demographic variables there will be 29 parameters in θ_1 and θ_2. If there are no observable demographic variables ($\Pi = 0$) then there will be 29 parameters in θ_1 and θ_2). so you cannot set the sample analogs exactly equal to zero. The GMM objective is a quadratic form:

$$
\min_{\theta_1, \theta_2} \Delta\xi' Z(Z'Z)^{-1} Z' \Delta\xi
$$

2 A Description of the Matlab Data Sets for Cereal Brands

Based on the readme file on Aviv Nevo's web site

The data set consists of two Matlab files: ps2.mat and iv.mat. The data are (semi-fabricated) data on J=24 brands of ready-to-eat cereal, for T=94 markets (47 US cities for the first 2 quarters of 1988). Total observations = 24 brands * 94 markets = 2256. (This is not the full set of data used in his Econometrica paper. There he had 1124 markets for each brand).

The file **ps2.mat** contains the following matrices/vectors:

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 is this data set) and q is quarter(=1 or 2 in this data set). All the other variables are sorted by date city brand. Dimension is 2256 x 1

s_jt- the market shares of brand j in market t. Each row corresponds to the equivalent row in id. Dimension is 2256 x 1.

X_1 - 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. Dimension is 2256 x 25.

X_2 - 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. Dimension is 2256 x 4.

id_demo - an id variable for each of the 94 markets with the format ccyyq. This will be used with the "consumer " data matrices (v and demogr described next) which do not vary by brand. Dimension is 94 x 1

v- random draws of consumers for each market. 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. Dimension is 94 x 80.

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. Dimension is 94 x 80.

The file **iv.mat** contains the matrix IV which consists of an id column (see the id variable above) and 20 columns of IV's for the price variable. Each column is the average price in the region in which the city is located (not including the city in question) in one quarter. The 20 quarters are from 1988:1 to 1992:4. The variable is sorted in the same order as the variables in the ps2.mat. Dimension is 2256 x 21.

3 A Summary of the Programming Steps

3.1 Overview

main.m

This is the main program.

- Inputs the data sets

- Constructs the simple logit estimates of the mean utilities δ_{jt} to use as starting values for the estimation (the variable MVAOLD is $\exp(\delta_{jt})$

- Inputs starting values for θ_2

- Calls the optimization algorithm that is used to minimize the GMM objective function w.r.t. θ_1 and θ_2. This is done in two steps. The subroutine that calculates the GMM objective function **(gmmobjg.m)** returns the value minimized w.r.t. $\theta_1$for a given θ_2 and the optimization routine (fminunc or fminsearch) minimizes w.r.t. θ_2.

- Constructs standard errors for the final estimates in θ_1 and θ_2.

- Constructs the estimates of $\beta^1, \beta^2, \beta^3$ using a WLS regression of the vector of brand dummies γ on a constant and the two brand characteristics (sugar and mushy).

3.2 Integration over Consumers

One important part of the program constructs predicted market shares from the model parameters. There are a set of three subroutines that take δ_{jt} and θ_2 and return the predicted market share s_{jt} for each observation. They use Monte-Carlo integration to aggregate over the observable (demographic) and unobservable (stochastic) dimensions of consumer heterogeneity. The subroutines are called in the following order:

mufunc.m

This takes the data that interacts in μ_{ijt} (X_2 =constant, price, sugar, mushy), the consumer data matrices D_i, ν_i (called DEMOGR, DFULL, V, and VFULL), and θ_2 and returns μ_{ijt}

ind_sh.m

This takes δ_{jt} and μ_{ijt} and constructs the logit market share for consumer i.

$$s_{ijt} = \frac{e^{\delta_{jt}+\mu_{ijt}}}{1 + \sum\limits_{m=1}^{J} e^{\delta_{mt}+\mu_{imt}}}$$

mktsh.m

This takes the logit market shares for each random-chosen consumer and averages them to construct the brand-level market share

$$s_{jt} = \frac{1}{I_t} \sum_i s_{ijt}$$

3.3 Solving Equation System for the Mean Utilities

meanval.m

A second step in the program is to update the estimates of the mean utilities based on a comparison of the predicted s_{jt} and actual s_{jt}^n market shares. Given a vector of predicted brand-level market shares, set them equal to the actual observed market shares and you have a system of JT equations in JT unknowns, the mean utilities δ_{jt}. This subroutine iterates on the equation system:

$$\exp(\delta_{jt}^{H+1}) = \exp(\delta_{jt}^{H})(s_{jt}^n / s_{jt})$$

until the mean utilities converge. Notice that the equation system can be solved market by market. In each market there are 24 market shares and the mean utilities of each brand in market t do not affect the predicted market shares in any other market t', so the equation system can be solved separately for each market.

3.4 Estimation of parameters

gmmobjg.m

This subroutine constructs the value of θ_1 that minimizes the GMM objective for a given value of θ_2 and returns the value of the objective function.

Input: θ_2 and δ_{jt}
Output: Value of the GMM objective function for the paramter vector θ_2
This subroutine first estimates $\theta_1 = (\gamma_{1,}\gamma_2, ...\gamma_J, \alpha)$ using a linear IV regression:

$$\delta_{jt} = \gamma_j + P_{jt}\alpha + \Delta\xi_{jt} = X_1\theta_1 + \Delta\xi_{jt}$$

where the γ_j are brand-level dummy variables and $\Delta\xi_{jt}$ is the error term. The IV estimator of θ_1 is

$$\hat{\theta}_1 = \left[(X_1^{'}Z)(Z^{'}Z)^{-1}Z^{'}X_1\right]^{-1}\left[(X_1^{'}Z)(Z^{'}Z)^{-1}Z^{'}\delta\right]$$

The vector of structural errors (called gmmresid in the program) is

$$\Delta\xi = \delta - X_1\hat{\theta}_1$$

The value of the GMM objective function (called f1 in the program) is

$$f1 = \Delta\xi^{'}Z(Z^{'}Z)^{-1}Z^{'}\Delta\xi$$

Finally, the subroutine analytically calculates the gradient of the objective function w.r.t. θ_2. It is named df in the program and uses the subroutine **jacob.m.** Analytical gradients can speed up the optimization greatly if we choose an optimization algorithm that requires first derivatives, such as quasi-Newton methods.

All of the calculations in the gmmobjg subroutine are for a given value of θ_2.

main.m

This uses either fminunc (quasi-Newton method) or fminsearch (Nelder-Mead method) to minimize the GMM objective function $f1$ with respect to the nonlinear parameters in the model θ_2.