7. Dynamic Consumer Demand

7.1 Introduction

Consumers can stockpile a storable good when prices are low and use the stock for future consumption. This stockpiling behavior can introduce significant differences between short-run and long-run responses of demand to price changes. Also, the response of demand to a price change depends on consumers' expectations/beliefs about how permanent that price change is. For instance, if a price reduction is perceived by consumers as very transitory (for instance, a sales promotion), then a significant proportion of consumers may choose to increase purchases today, stockpile the product and reduce their purchases during future periods when the price will be higher. If the price reduction is perceived as permanent, this intertemporal substitution of consumer purchases will be much lower or even zero.

Ignoring consumers' stockpiling and forward-looking behavior can introduce serious biases in our estimates of own- and cross- price demand elasticities. These biases can be particularly serious when the time series of prices is characterized by "High-Low" pricing. The price fluctuates between a (high) regular price and a (low) promotion price. The promotion price is infrequent and last only few days, after which the price returns to its "regular" level. Most sales are concentrated in the very few days of promotion prices.

Static demand models assume that all the substitution is either between brands or product expansion. They rule out intertemporal substitution. This can imply serious biases in the estimated demand elasticities. With High-Low pricing, we expect the static model to over-estimate the own-price elasticity. The bias in the estimated elasticities also implies a bias in the estimated Price Cost Margins (PCM). We expect PCMs to be underestimated. These biases have serious implications on policy analysis, such as merger analysis and antitrust cases.

Here we discuss two papers that have estimated dynamic structural models of demand of differentiated products using consumer level data (scanner data): Hendel and Nevo (2006) and **erdem_keane_2003** (**erdem_keane_2003**). These papers extend

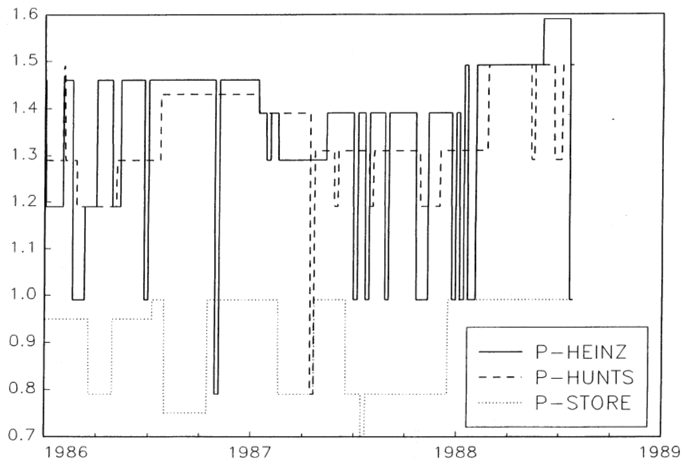

Figure 7.1: Pesendorfer (2002): Hi-Lo Pricing - Ketchup

microeconometric discrete choice models of product differentiation to a dynamic setting, and contain useful methodological contributions. Their empirical results show that ignoring the dynamics of demand can lead to serious biases. The papers also illustrate how the use of **micro level data on household choices** (in contrast to only aggregate data on market shares) is key for credible identification of the dynamics of differentiated product demand.

7.2 Data and descriptive evidence

In the following section, the researcher has access to consumer level data. Such data is widely available from several data collection companies and recently researchers in several countries have been able to gain access to such data for academic use. The data include the history of shopping behavior of a consumer over a period of one to three years. The researcher knows whether a store was visited, and what product (brand and size) was purchased and at what price. From the view point of the model, the key information that is not observed is consumer inventory and consumption decisions.

Hendel and Nevo use consumer-level scanner data from Dominicks, a supermarket chain that operates in the Chicago area. The dataset comes from 9 supermarket stores and it covers the period June 1991 to June 1993. Purchases and price information are available in real (continuous) time but for the analysis in the paper it is aggregated at the weekly frequency.

The dataset has two components: store-level and household-level data. **Store level data:** For each detailed product (brand–size) in each store and in each week, we observe the (average) price charged, (aggregate) quantity sold, and promotional activities. **Household level data:** For a sample of households, we observe the purchases of households at the 9 supermarket stores: supermarket visits and total expenditure in each visit; purchases (units and value) of detailed products (brand-size) in 24 different product categories (for instance, laundry detergent, milk, etc). The paper specifically studies demand of laundry detergent products.

Table I in the paper presents summary statistics on household demographics, purchases, and store visits.

Table II in the paper presents the market shares of the main brands of laundry detergent in the data. The market is significantly concentrated, especially the market for Powder laundry detergent where the concentration ratios are $CR1 = 40\%$, $CR2 = 55\%$, and $CR3 = 65\%$. For most brands, the proportion of sales under a promotion price is important. However, this proportion varies importantly between brands, showing that different brands have different patterns of prices.

Descriptive evidence. H&N present descriptive evidence which is consistent with household inventory holding. See also Hendel and Nevo (2006). Though household purchase histories are observable, household inventories and consumption are unobservable. Therefore, empirical evidence on the importance of household inventory holding is indirect.

(a) Time duration since previous sale promotion has a positive effect on the aggregate quantity purchased.

(b) Indirect measures of storage costs (for instance, house size) are negatively correlated with households' propensity to buy on sale.

7.3 Model

7.3.1 Basic Assumptions

Consider a differentiated product, laundry detergent, with J different brands. Every week a household has some level of inventories of the product (that may be zero) and chooses (a) how much to consume from its inventory; and (b) how much to purchase (if any) of the product, and the brand to purchase.

An important simplifying assumption in Hendel-Nevo model is that consumers care about brand choice when they purchase the product, but not when they consume or store it. We explain below the computational advantages of this assumption. Of course, the assumption imposes some restrictions on the intertemporal substitution between brands, and we will discuss this point too. **erdem_keane_2003** (**erdem_keane_2003**) do not impose that restriction.

The subindex t represents time, the subindex j represents a brand, and the subindex h represents a consumer or household. A household's current utility function is:

$$u_h(c_{ht}, v_{ht}) - C_h(i_{h,t+1}) + m_{ht}$$

$u_h(c_{ht}, v_{ht})$ is the utility from consumption of the storable product, with c_{ht} being consumption and v_{ht} being a shock in the utility of consumption:

$$u_h(c_{ht}, v_{ht}) = \gamma_h \ \ln(c_{ht} + v_{ht})$$

$C_h(i_{h,t+1})$ is the inventory holding cost, where $i_{h,t+1}$ is the level of inventory at the end of period t, after consumption and new purchases:

$$C_h(i_{h,t+1}) = \delta_{1h} \ i_{h,t+1} + \delta_{2h} \ i_{h,t+1}^2$$

m_{ht} is the indirect utility function from consumption of the composite good (outside good) plus the utility from brand choice (that is, the utility function in a static discrete model of differentiated products):

$$m_{ht} = \sum_{j=1}^{J} \sum_{x=0}^{X} d_{hjxt} \left(\beta_h \ a_{jxt} - \alpha_h \ p_{jxt} + \xi_{jxt} + \varepsilon_{hjxt}\right)$$

where $j \in \{1, 2, ..., J\}$ is the brand index, and $x \in \{0, 1, 2, ..., X\}$ is the index of quantity choice, where the maximum possible size is X units. In this application $X = 4$. Brands with different sizes are standardized such that the same measurement unit is used in x. The variable $d_{hjxt} \in \{0, 1\}$ is a binary indicator for the event "household purchases x units of brand j at week t". p_{jxt} is the price of x units of brand j at period t. Note that the models allows for nonlinear pricing, that is, for some brands and weeks p_{jxt} and $x * p_{j1t}$ can take different values. This is potentially important because the price data shows a significant degree of nonlinear pricing. a_{jxt} is a vector of product characteristics other than price that is observable to the researcher. In this application, the most important variables in a_{jxt} are those that represent store-level advertising, for instance, display of the product in the store, etc. The variable ξ_{jxt} is a random variable that is unobservable to the researcher and that represents all the product characteristics which are known to consumers but not in the set of observable variables in the data.

α_h and β_h represent the marginal utility of income and the marginal utility of product attributes in a_{jxt}, respectively. As it is well-known in the empirical literature of demand of differentiated products, it is important to allow for heterogeneity in these marginal utilities in order to have demand systems with flexible and realistic own and cross elasticities or substitution patterns. Allowing for this heterogeneity is much simpler with consumer level data on product choices than with aggregate level data on product market shares. In particular, micro level datasets can include information on a rich set of household socioeconomic characteristics such as income, family size, age, education, gender, occupation, house-type, etc, that can be included as observable variables that determine the marginal utilities α_h and β_h. This is the approach in Hendel and Nevo's paper.

Finally, ε_{hjxt} is a consumer idiosyncratic shock that is independently and identically distributed over (h, j, x, t) with an extreme value type 1 distribution. This is the typical logit error that is included in most discrete models of demand of differentiated products. Note that while ε_{hjxt} varies over individuals, ξ_{jxt} does not.

Let $\mathbf{p}_t$ be the vector of product characteristics, observable or unobservable, for all the brands and sizes at period t:

$$\mathbf{p}_t \equiv \left\{ p_{jxt}, a_{jxt}, \xi_{jxt} : j = 1,2,...,J \text{ and } x = 1,2,...,X \right\}$$

Every week t, the household knows her level of inventories, i_{ht}, observes product attributes $\mathbf{p}_t$, and its idiosyncratic shocks in preferences, v_{ht} and ε_{ht}. Given this information, the household decides her consumption of the storable product, c_{ht}, and how much to purchase of which product, $d_{ht} = \{d_{hjxt}\}$. The household makes this decision in order to maximize her expected and discounted stream of current and future utilities,

$$\mathbb{E}_t \left(\sum_{s=0}^{\infty} \delta^s \left[u_h(c_{ht+s}, v_{ht+s}) - C_h(i_{h,t+s+1}) + m_{ht+s} \right] \right)$$

where δ is the discount factor.

The vector of state variables of this DP problem is $\{i_{ht}, v_{ht}, \varepsilon_{ht}, \mathbf{p}_t\}$. The decision variables are c_{ht} and d_{ht}. To complete the model we need to make some assumptions on the stochastic processes of the state variables. The idiosyncratic shocks v_{ht} and ε_{ht} are assumed iid over time. The vector of product attributes $\mathbf{p}_t$ follows a Markov process. Finally, consumer inventories i_{ht} has the obvious transition rule:

$$i_{h,t+1} = i_{h,t+1} - c_{ht} + \left(\sum_{j=1}^{J} \sum_{x=0}^{X} d_{hjxt}\, x \right)$$

where $\sum_{j=1}^{J} \sum_{x=0}^{X} d_{hjxt}\, x$ represents the units of the product purchased by household h at period t.

Let $V_h(\mathbf{s}_{ht})$ be the value function of a household, where $\mathbf{s}_{ht}$ is the vector of state variables $(i_{ht}, v_{ht}, \varepsilon_{ht}, \mathbf{p}_t)$. A household decision problem can be represented using the Bellman equation:

$$V_h(\mathbf{s}_{ht}) = \max_{\{c_{ht}, d_{ht}\}} \left[u_h(c_{ht}, v_{ht}) - C_h(i_{h,t+1}) + m_{ht} + \delta\, \mathbb{E}\left(V_h(\mathbf{s}_{ht+1}) \mid \mathbf{s}_{ht},\ c_{ht},\ d_{ht} \right) \right]$$

where the expectation $\mathbb{E}(. \mid \mathbf{s}_{ht},\ c_{ht},\ d_{ht})$ is taken over the distribution of $\mathbf{s}_{ht+1}$ conditional on $(\mathbf{s}_{ht}, c_{ht}, d_{ht})$. The solution of this DP problem implies optimal decision rules for consumption and purchasing decisions: $c_{ht} = c_h^*(\mathbf{s}_{ht})$ and $d_{ht} = d_h^*(\mathbf{s}_{ht})$ where $c_h^*(.)$ and $d_h^*$are the decision rules. Note that they are household specific because there is time-invariant household heterogeneity in the marginal utility of product attributes (α_h and β_h), in the utility of consumption of the storable good u_h, and in inventory holding costs, C_h.

The optimal decision rules $c_h^*(.)$ and d_h^*depend also on the structural parameters of the model: the parameters in the utility function, and in the transition probabilities of the state variables. In principle, we could use the equations $c_{ht} = c_h^*(\mathbf{s}_{ht})$, $d_{ht} = d_h^*(\mathbf{s}_{ht})$, and our data on (some) decisions and state variables to estimate the parameters of the model. To apply this revealed preference approach, there are three main issues we have to deal with.

First, the dimension of the state space of $\mathbf{s}_{ht}$ is extremely large. In most applications of demand of differentiated products, there are dozens (or even more than a hundred) products. Therefore, the vector of product attributes $\mathbf{p}_t$ contains more than a hundred continuous state variables. Solving a DP problem with this state space, or even approximating the solution with enough accuracy using Monte Carlo simulation methods, is computationally very demanding even with the most sophisticated computer equipment. We will see how Hendel and Nevo propose and implement a method to reduce the dimension of the state space. The method is based on some assumptions that we discuss below.

Second, though we have good data on households' purchasing histories, information on households' consumption and inventories of storable goods is very rare. In this application, consumption and inventories, c_{ht} and i_{ht}, are unobservable to the researchers. Not observing inventories is particularly challenging. This is the key state variable in a dynamic demand model for the demand of a storable good. We will discuss below the approach used by Hendel and Nevo to deal with this issue, and also the approach used by **erdem_keane_2003** (**erdem_keane_2003**).

Third, as usual in the estimation of a model of demand, we should deal with the endogeneity of prices. Of course, this problem is not specific to a dynamic demand model. However, dealing with this problem may not be independent of the other issues mentioned above.

7.3.2 Reducing the dimension of the state space

Given that the state variables $(v_{ht},\ \varepsilon_{ht})$ are independently distributed over time, it is convenient to reduce the dimension of this DP problem by using a value function that is integrated over these iid random variables. The integrated value function is defined as:

$$\bar{V}_h(i_{ht},\mathbf{p}_t) \equiv \int V_h(\mathbf{s}_{ht})\, dF_\varepsilon(\varepsilon_{ht})\, dF_v(v_{ht})$$

where F_ε and F_v are the CDFs of ε_{ht} and v_{ht}, respectively. Associated with this integrated value function is an integrated Bellman equation. Given the distributional assumptions on the shocks ε_{ht} and v_{ht}, the integrated Bellman equation is:

$$\bar{V}_h(i_{ht},\mathbf{p}_t) = \max_{c_{ht},d_{ht}} \int \ln\left(\sum_{j=1}^{J} \exp\left\{\begin{array}{l} u_h(c_h,v_{ht}) - C_i(i_{ht+1}) + m_{ht} \\ \\ +\delta\, \mathbb{E}\left[\bar{V}_h(i_{ht+1},\mathbf{p}_{t+1}) \mid i_{ht},\mathbf{p}_t,c_{ht},d_{ht}\right] \end{array}\right\}\right) dF_v(v_{ht}).$$

This Bellman equation is also a contraction mapping in the value function. The main computational cost in the computation of the functions $\bar{V}_h$ comes from the dimension of the vector of product attributes $\mathbf{p}_t$. We now explore ways to reduce this cost.

First, note that the assumption of a single aggregate inventory for all products, instead of one inventory for each brand, $\{i_{hjt}\}$, already reduces importantly the dimension of the state space. This assumption not only reduces the state space but, as we see below, it also allows us to modify the dynamic problem, which can significantly aid in the estimation of the model.

Taken literally, this assumption implies that there is no differentiation in consumption: the product is homogenous in use. Note, that through ξ_{jxt} and ε_{ijxt} the model allows differentiation in purchase, as is standard in the IO literature. It is well known that this differentiation is needed to explain purchasing behavior. This seemingly creates a tension in the model: products are differentiated at purchase but not in consumption. Before explaining how this tension is resolved we note that the tension is not only in the model but potentially in reality as well. Many products seem to be highly differentiated at the time of purchase but it is hard to imagine that they are differentiated in consumption. For example, households tend to be extremely loyal to the laundry detergent brand they purchase – a typical household buys only 2-3 brands of detergent over a very long horizon – yet it is hard to imagine that the usage and consumption are very different for different brands.

A possible interpretation of the model that is consistent with product differentiation in consumption is that the variables ξ_{jxt} not only capture instantaneous utility at period t but also the discounted value of consuming the x units of brand j. This is a valid interpretation if brand-specific utility in consumption is additive such that it does not affect the marginal utility of consumption.

This assumption has some implications that simplify importantly the structure of the model. It implies that the optimal consumption does not depend on which brand is purchased, only on the size. And relatedly, it implies that the brand choice can be treated as a static decision problem.

We can distinguish two components in the choice d_{ht}: the quantity choice, x_{ht}, and the brand choice j_{ht}. Given $x_{ht} = x$, the optimal brand choice is:

$$j_{ht} = \arg \max_{j \in \{1,2,\ldots,J\}} \left\{ \beta_h \, a_{jxt} - \alpha_h \, p_{jxt} + \xi_{jxt} + \varepsilon_{hjxt} \right\}$$

Then, given our assumption about the distribution of ε_{hjxt}, the component m_{ht} of the utility function can be written as $m_{ht} = \sum_{x=0}^{X} \omega_h(x, \mathbf{p}_t) + e_{ht}$ where $\omega_{ht}(x, \mathbf{p}_t)$ is the inclusive value:

$$\begin{aligned}
\omega_h(x, \mathbf{p}_t) &\equiv \mathbb{E}\left(\max_{j \in \{1,2,\ldots,J\}} \left\{ \beta_h \, a_{jxt} - \alpha_h \, p_{jxt} + \xi_{jxt} + \varepsilon_{hjxt} \right\} \mid x_{ht} = x, \ \mathbf{p}_t \right) \\
&= \ln\left(\sum_{j=1}^{J} \exp\left\{ \beta_h \, a_{jxt} - \alpha_h \, p_{jxt} + \xi_{jxt} \right\} \right)
\end{aligned}$$

and e_{ht} does not depend on size x (or on inventories and consumption), and therefore we can ignore this variable for the dynamic decisions on size and consumption.

Therefore, the dynamic decision problem becomes:

$$\bar{V}_h(i_{ht}, \mathbf{p}_t) = \max_{c_{ht}, x_{ht}} \int \left\{ u_h(c_{ht}, v_{ht}) - C_i(i_{ht+1}) + \omega_h(x, \mathbf{p}_t) + \delta \, \mathbb{E}\left[\bar{V}_h(i_{ht+1}, \mathbf{p}_{t+1}) \mid i_{ht+1}, \mathbf{p}_t \right] \right\} dF_v(v_{ht})$$

In words, the problem can now be seen as a choice between sizes, each with a utility given by the size-specific inclusive value (and extreme value shock). The dimension of the state space is still large and includes all product attributes, because we need these attributes to compute the evolution of the inclusive value. However, in combination with additional assumptions, the modified problem is easier to estimate.

Note also that the expression describing the optimal brand choice, $j_{ht} = \arg\max_{j \in \{1,2,\ldots,J\}} \{\beta_h \, a_{jxt} - \alpha_h \, p_{jxt} + \xi_{jxt} + \varepsilon_{hjxt}\}$, is a "standard" multinomial logit model with the caveat that prices are endogenous explanatory variables because they depend on the unobserved attributes in ξ_{jxt}. We describe below how to deal with this endogeneity problem. With household level data, dealing with the endogeneity of prices is much simpler than with aggregate data on market shares. More specifically, we do not need to use Monte Carlo simulation techniques, or an iterative algorithm to compute the "average utilities" $\{\delta_{jxt}\}$.

To reduce the dimension of the state space, Hendel and Nevo (2006) introduce the following assumption. Let $\omega_h(\mathbf{p}_t)$ be the vector with the inclusive values for every possible size $\{\omega_h(x, \mathbf{p}_t) : x = 1, 2, ..., X\}$.

Assumption: The vector $\omega_h(\mathbf{p}_t)$ is a sufficient statistic of the information in $\mathbf{p}_t$ that is useful to predict $\omega_h(\mathbf{p}_{t+1})$:

$$\Pr(\omega_h(\mathbf{p}_{t+1}) \mid \mathbf{p}_t) = \Pr(\omega_h(\mathbf{p}_{t+1}) \mid \omega_h(\mathbf{p}_t))$$

In words, the vector $\omega_h(\mathbf{p}_t)$ contains all the relevant information in $\mathbf{p}_t$ needed to obtain the probability distribution of $\omega_h(\mathbf{p}_{t+1})$ conditional on $\mathbf{p}_t$. Instead of all the prices and attributes, we only need a single index for each size. Two vectors of prices that yield the same (vector of) current inclusive values imply the same distribution of future inclusive values. This assumption is violated if individual prices have predictive power above and beyond the predictive power of $\omega_h(\mathbf{p}_t)$.

The inclusive values can be estimated outside the dynamic demand model. Therefore, the assumption can be tested and somewhat relaxed by including additional statistics of prices in the state space. Note that $\omega_h(\mathbf{p}_t)$ is consumer specific: different consumers value a given set of products differently and therefore this assumption does not further restrict the distribution of heterogeneity.

Given this assumption, the integrated value function is $\bar{V}_h(i_{ht}, \omega_{ht})$, which includes only $X+1$ variables, instead of $3 * J * X + 1$ state variables.

7.4 Estimation

7.4.1 Estimation of brand choice

Let j_{ht} represent the brand choice of household h at period t. Under the assumption that there is product differentiation in purchasing but not in consumption or in the cost of inventory holding, a household brand choice is a static decision problem. Given $x_{ht} = x$, with $x > 0$, the optimal brand choice is:

$$j_{ht} = \arg \max_{j \in \{1,2,...,J\}} \left\{ \beta_h \, a_{jxt} - \alpha_h \, p_{jxt} + \xi_{jxt} + \varepsilon_{hjxt} \right\}$$

The estimation of demand models of differentiated products, either static or dynamic, should deal with two important issues. The first is the endogeneity of prices. The model implies that p_{jxt} depends on observed and unobserved products attributes, and therefore p_{jxt} and ξ_{jxt} are not independently distributed. The second issue is that the model should allow for rich heterogeneity in consumers' marginal utilities of product attributes, β_h and α_h. Using consumer-level data (instead of aggregate market share data) facilitates significantly the econometric solution of these issues.

Consumer-level scanner datasets contain rich information on household socioeconomic characteristics. Let z_h be a vector of observable socioeconomic characteristics that have a potential effect on demand, for instance, income, family size, age distribution of children and adults, education, occupation, type of housing, etc. We assume that β_h and α_h depend on this vector of household characteristics:

$$\begin{aligned} \beta_h &= \beta_0 + (z_h - \bar{z})\sigma_\beta \\ \alpha_h &= \alpha_0 + (z_h - \bar{z})\sigma_\alpha \end{aligned}$$

β_0 and α_0 are scalar parameters that represent the marginal utility of advertising and income, respectively, for the average household in the sample. $\bar{z}$ is the vector of household attributes of the average household in the sample. σ_β and σ_α are $K \times 1$ vectors of parameters that represent the effect of household attributes on marginal utilities. Therefore, the utility of purchasing can be written as:

$$\begin{aligned} & \left[\beta_0 + (z_h - \bar{z})\sigma_\beta\right] a_{jxt} - \left[\alpha_0 + (z_h - \bar{z})\sigma_\alpha\right] p_{jxt} + \xi_{jxt} + \varepsilon_{hjxt} \\ = \; & \left[\beta_0 \; a_{jxt} - \alpha_0 \; p_{jxt} + \xi_{jxt}\right] + (z_h - \bar{z}) \left[a_{jxt} \; \sigma_\beta - p_{jxt} \; \sigma_\alpha\right] + \varepsilon_{hjxt} \\ = \; & \delta_{jxt} + (z_h - \bar{z}) \; \sigma_{jxt} + \varepsilon_{hjxt} \end{aligned}$$

where $\delta_{jxt} \equiv \beta_0 \; a_{jxt} - \alpha_0 \; p_{jxt} + \xi_{jxt}$, and $\sigma_{jxt} \equiv a_{jxt} \; \sigma_\beta - p_{jxt} \; \sigma_\alpha$. δ_{jxt} is a scalar that represents the utility of product (j,x,t) for the average household in the sample. σ_{jxt} is a vector and each element in this vector represents the effect of a household attribute on the utility of product (j,x,t).

In fact, it is possible to allow also for interactions between the observable household attributes and the unobservable product attributes, such that we have a term $\lambda_h \xi_{jxt}$ where $\lambda_h = 1 + (z_h - \bar{z})\sigma_\lambda$. With this more general specification, we still have that $\delta_{jxt} \equiv \beta_0 \; a_{jxt} - \alpha_0 \; p_{jxt} + \xi_{jxt}$, but now $\sigma_{jxt} \equiv a_{jxt} \; \sigma_\beta - p_{jxt} \; \sigma_\alpha + \xi_{jxt}\sigma_\lambda$.

Dummy-Variables Maximum Likelihood + IV estimator

Given this representation of the brand choice model, the probability that a household with attributes z_h purchases brand j at period t given that it buys x units of the product is:

$$P_{hjxt} = \frac{\exp\left\{\delta_{jxt} + (z_h - \bar{z}) \; \sigma_{jxt}\right\}}{\sum_{k=1}^{J} \exp\left\{\delta_{kxt} + (z_h - \bar{z}) \; \sigma_{kxt}\right\}}$$

Given a sample with a large number of households, we can estimate δ_{jxt} and σ_{jxt} for every (j,x,t) in a multinomial logit model with probabilities $\{P_{hjxt}\}$. For instance, we can estimate these "incidental parameters" δ_{jxt} and σ_{jxt} separately for every value of (x,t). For $(t = 1, x = 1)$ we select the subsample of households who purchase $x = 1$ unit of the product at week $t = 1$. Using this subsample, we estimate the vector of $J(K+1)$ parameters $\{\delta_{j11}, \sigma_{j11} : j = 1,2,...,J\}$ by maximizing the multinomial log-likelihood function:

$$\sum_{h=1}^{H} 1\{x_{h1} = 1\} \sum_{j=1}^{J} 1\{j_{h1} = j\} \; \ln P_{hj11}$$

We can proceed in the same way to estimate all the parameters $\{\delta_{jxt}, \sigma_{jxt}\}$.

This estimator is consistent as H goes to infinity for fixed T, X, and J. For a given (finite) sample, there are some requirements on the number of observations in order to be able to estimate the incidental parameters. For every value of (x,t), the number of incidental parameters to estimate is $J(K+1)$, and the number of observations is equal to the number of households who purchase x units at week t, that is, $H(x,t) = \sum_{h=1}^{H} 1\{x_{ht} = x\}$. We need that $H(x,t) > J(K+1)$. For instance, with $J = 25$ products and $K = 4$ household attributes, we need $H(x,t) > 125$, for every week t and every size x. We may need a very large number of households H in the sample in order to satisfy these conditions. An assumption that may eliminate this problem is that the utility from brand choice is proportional to quantity: $x(\beta_h\, a_{jt} - \alpha_h\, p_{jt} + \xi_{jt} + \varepsilon_{hjt})$. Under this assumption, we have that for every week t, the number of incidental parameters to estimate is $J(K+1)$, but the number of observations is now equal to the number of households who purchase any quantity $x > 0$ at week t, that is, $H(t) = \sum_{h=1}^{H} 1\{x_{ht} > 0\}$. We need that $H(t) > J(K+1)$ which is a much weaker condition.

Given estimates of the incidental parameters, $\{\hat{\delta}_{jxt}, \hat{\sigma}_{jxt}\}$, we can now estimate the structural parameters β_0, α_0, σ_β, and σ_α using an IV (or GMM) method. For the estimation of β_0 and α_0, we have that:

$$\hat{\delta}_{jxt} = \beta_0\, a_{jxt} - \alpha_0\, p_{jxt} + \xi_{jxt} + e_{jxt}$$

where e_{jxt} represents the estimation error $(\hat{\delta}_{jxt} - \delta_{jxt})$. This is a linear regression where the regressor p_{jxt} is endogenous. We can estimate this equation by IV using the so-called "BLP instruments", that is, the characteristics other than price of products other than j, $\{a_{kxt} : k \neq j\}$. Of course, there are other approaches to deal with the endogeneity of prices in this equation. For instance, we could consider the following Error-Component structure in the endogenous part of the error term: $\xi_{jxt} = \xi_{jx}^{(1)} + \xi_{jxt}^{(2)}$ where $\xi_{jxt}^{(2)}$ is assumed not serially correlated. Then, we can control for $\xi_{jx}^{(1)}$ using product-size dummies, and use lagged values of prices and other product attributes to deal with the endogeneity of prices that comes from the correlation with the transitory shock $\xi_{jxt}^{(2)}$.

For the estimation of σ_β, and σ_α, we have the system of equations:

$$\hat{\sigma}_{jxt} \quad = \quad a_{jxt}\sigma_\beta - p_{jxt}\, \sigma_\alpha + \xi_{jxt}\sigma_\lambda + e_{jxt}$$

where e_{jxt} represents the estimation error $(\hat{\sigma}_{jxt} - \sigma_{jxt})$. We have one equation for each household attribute. We can estimate each of these equations using the same IV procedure as for the estimation of β_0 and α_0.

Once we have estimated (β_0, α_0, σ_β, σ_α), we can also obtain estimates of ξ_{jxt} as residuals from the estimated equation. We can also get consistent estimates of the marginal utilities β_h and α_h as:

$$\begin{aligned} \hat{\beta}_h &= \hat{\beta}_0 + (z_h - \bar{z})\hat{\sigma}_\beta \\ \hat{\alpha}_h &= \hat{\alpha}_0 + (z_h - \bar{z})\hat{\sigma}_\alpha \end{aligned}$$

Finally, we can get estimates of the inclusive values:

$$\hat{\omega}_{hxt} \quad = \quad \ln\left(\sum_{j=1}^{J} \exp\left\{ \hat{\beta}_h\, a_{jxt} - \hat{\alpha}_h\, p_{jxt} + \hat{\xi}_{jxt} \right\} \right)$$

Control function approach

The previous approach, though simple, has a certain limitation: we need to have, for every week in the sample, a large enough number of households making positive purchases. However, this requirement is not needed for identification of the parameters, only for the implementation of the simple two-step dummy variables approach to deal with the endogeneity of prices.

When our sample does not satisfy that requirement, there is another simple method that we can use. This method is a control function approach that is in the spirit of the methods proposed by Rivers and Vuong (1988), **blundell_powell_2004** (**blundell_powell_2004**), and in the specific context of demand of differentiated products, **petrin_train_2010** (**petrin_train_2010**).

If firms choose prices to maximize profits, we expect that prices depend on the own product characteristics and also on the characteristics of competing products: $p_{jxt} = f_{jxt}(a_t, \xi_t)$, where $a_t = \{a_{jxt}$:for any $j, x\}$, and $\xi_t = \{\xi_{jxt}$:for any $j, x\}$. Define the conditional mean function:

$$g^{p}_{jx}(a_t) \equiv \mathbb{E}(p_{jxt} \mid a_t) = \mathbb{E}(f_{jxt}(a_t, \xi_t) \mid a_t)$$

Then, we can write the regression equation:

$$p_{jxt} = g^{p}_{jx}(a_t) + e_{jxt}$$

where the error term e_{jxt} is by construction mean independent of a_t.

The first step of the control function method consists in the estimation of the conditional mean functions $g^{p}_{jx}$for every brand and size (j, x). Though we have a relatively large number of weeks in our dataset (more than 100 weeks in most scanner datasets), the number of variables in the vector a_t is $J * X$, which is a significantly large number. Therefore, we need to impose some restrictions on how the exogenous product characteristics in a_t affect prices. For instance, we may assume that,

$$g^{p}_{jx}(a_t) = g^{p}_{jx}\left(\, a_{jxt},\, \bar{a}_{j(-x)t},\, \bar{a}_{(-j)xt},\, \bar{a}_{(-jx)t}\,\right)$$

where $\bar{a}_{j(-x)t}$ is the sample mean of variable a at period t for all the products of brand j but with different size than x; $\bar{a}_{(-j)xt}$ is the sample mean for all the products with size x but with brand different than j; and $\bar{a}_{(-jx)t}$ is the sample mean for all the products with size different than x and brand different than j. Of course, we can consider more flexible specifications but still with a number of regressors much smaller than $J * X$.

The second step of the method is based on a decomposition of the error term ξ_{jxt} in two components: an endogenous component that is a deterministic function of the error terms in the first step, $e_t \equiv \{e_{jxt}$: for any j and $x\}$, and an "exogenous" component that is independent of the price p_{jxt} once we have controlled for e_{jxt}. Define the conditional mean function:

$$g^{\xi}_{jx}(e_t) \equiv \mathbb{E}(\xi_{jxt} \mid e_t)$$

Then, we can write ξ_{jxt} as the sum of two components, $\xi_{jxt} = g^{\xi}_{jx}(e_t) + v_{jxt}$. By construction, the error term v_{jxt} is mean independent of e_t. Additionally, v_{jxt} is mean independent of all the product prices because prices depend only on the exogenous product characteristics a_t (that by assumption are independent of ξ_{jxt}) and on the "residuals" e_t (that by construction are mean independent of v_{jxt}). Therefore, we can write the utility of product (j,x) as:

$$\beta_h\, a_{jxt} - \alpha_h\, p_{jxt} + g^{\xi}_{jx}(e_t) + (v_{jxt} + \varepsilon_{hjxt})$$

The term $g^{\xi}_{jx}(e_t)$ is the control function.

Under the assumption that $(v_{jxt} + \varepsilon_{hjxt})$ is iid extreme value type 1 distributed, we have that the brand choice probabilities conditional on $x_{ht} = x$ are:

$$P_{hjxt} = \frac{\exp\left\{\beta_0 a_{jxt} - \alpha_0 p_{jxt} + a_{jxt}(z_h - \bar{z})\sigma_\beta - p_{jxt}(z_h - \bar{z})\sigma_\alpha + g^{\xi}_{jx}(e_t)\right\}}{\sum_{k=1}^{J} \exp\left\{\beta_0 a_{kxt} - \alpha_0 p_{kxt} + a_{kxt}(z_h - \bar{z})\sigma_\beta - p_{kxt}(z_h - \bar{z})\sigma_\alpha + g^{\xi}_{kx}(e_t)\right\}}$$

where the control functions $\{g^{\xi}_{jx}(e_t)\}$ consist of a brand dummy and a polynomial in the residual variables $\{e_{jxt} : j = 1,2,...,J\}$. Then, we can estimate $(\beta_0,\, \alpha_0,\, \sigma_\beta,\, \sigma_\alpha)$ and the parameters of the control function by using Maximum Likelihood in this multinomial logit model. The log-likelihood function is:

$$\ell(\theta) = \sum_{h=1}^{H} \sum_{t=1}^{T} \sum_{x=1}^{X} \sum_{j=1}^{J} 1\{x_{ht} = x\ ,\ j_{ht} = j\}\ \ln P_{hjxt}$$

As in the previous method, once we have estimated these parameters, we can construct consistent estimates of the inclusive values ω_{hxt}.

7.4.2 Estimation of quantity choice

As mentioned above, the lack of data on household inventories is a challenging econometric problem because this is a key state variable in a dynamic demand model for the demand of a storable good. Also, this is not a "standard" unobservable variable, as it follows a stochastic process that is endogenous. That is, not only do inventories affect purchasing decisions, but purchasing decisions also affect the evolution of inventories.

The approach used by **erdem_keane_2003** (**erdem_keane_2003**) to deal with this problem is to assume that household inventories is a (deterministic) function of the "number of weeks (duration) since last purchase", T_{ht}, and the quantity purchased in the last purchase, $x^{\ell ast}_{ht}$:

$$i_{ht} = f_h(x^{\ell ast}_{ht}, T_{ht})$$

In general, this assumption holds under two conditions: (1) consumption is deterministic; and (2) when a new purchase is made, the existing inventory at the beginning of the week is consumed or scrapped. For instance, suppose that these conditions hold and that the level of consumption is constant $c_{ht} = c_h$. Then,

$$i_{ht+1} = \max\left\{\ 0\ ;\ x^{\ell ast}_{ht} - c_h\ T_{ht}\right\}$$

The constant consumption can be replaced by a consumption rate that depends on the level of inventories. For instance, $c_{ht} = \lambda_h i_{ht}$. Then:

$$i_{ht+1} = \max\left\{\ 0\ ;\ (1 - \lambda_h)^{T_{ht}}\ x^{\ell ast}_{ht}\right\}$$

Using this approach, the state variable i_{ht} should be replaced by the state variables $(x_{ht}^{\ell ast}, T_{ht})$, but the rest of the features of the model remain the same. The parameters c_h or λ_h can be estimated together with the rest of the parameters of the structural model. Also, we may not need to solve for the optimal consumption decision.

There is no doubt that using observable variables to measure inventories is very useful for the estimation of the model and for identification. It also provides a more intuitive interpretation of the identification of the model.

The individual level data provide the probability of purchase conditional on current prices, and past purchases of the consumer (amounts purchased and duration from previous purchases): $\Pr(x_{ht}|x_{ht}^{\ell ast}, T_{ht}, \mathbf{p}_t)$. Suppose that we observe that this probability is not a function of past behavior $(x_{ht}^{\ell ast}, T_{ht})$. We would then conclude that dynamics are not relevant, and that consumers are purchasing for immediate consumption and not for inventory. On the other hand, if we observe that the purchase probability is a function of past behavior, and we assume that preferences are stationary, then we conclude that there is dynamic behavior.

Regarding the identification of storage costs, consider the following example. Suppose we observe two consumers who face the same price process and purchase the same amount over a relatively long period. However, one of them purchases more frequently than the other. This variation leads us to conclude that this consumer has higher storage costs. Therefore, the storage costs are identified from the average duration between purchases.

Hendel and Nevo use a different approach, though the identification of their model is based on the same intuition.

7.5 Empirical Results

To Be Completed

7.6 Dynamic Demand of Differentiated Durable Products

- **gowrisankaran_rysman_2009** (**gowrisankaran_rysman_2009**). TBW