CHAPTER 6

Dynamic Structural Models of Industrial Organization

1. Introduction

Dynamics in demand and/or supply can be important aspects of competition in oligopoly markets. In many markets demand is dynamic in the sense that (a) consumers current decisions affect their future utility, and (b) consumers' current decisions depend on expectations about the evolution of future prices (states). Some sources of dynamics in demand are consumer switching costs, habit formation, brand loyalty, learning, and storable or durable products. On the supply side, most firm investment decisions have implications on future profits. Some examples are market entry, investment in capacity, inventories, or equipment, or choice of product characteristics. Firms' production decisions have also dynamic implications if there is learning by doing. Similarly, the existence of menu costs, or other forms of price adjustment costs, imply that pricing decisions have dynamic effects.

Identifying the factors governing the dynamics is important to understanding competition and the evolution of market structure, and for the evaluation of public policy. To identify and understand these factors, we specify and estimate dynamic structural models of demand and supply in oligopoly industries. A dynamic structural model is a model of individual behavior where agents are forward looking and maximize expected intertemporal payoffs. The parameters are structural in the sense that they describe preferences and technological and institutional constraints. Under the *principle of revealed preference*, these parameters are estimated using longitudinal micro data on individuals' choices and outcomes over time.

I start with some examples and a brief discussion of applications of dynamic structural models of Industrial Organization. These examples illustrate why taking into account forward-looking behavior and dynamics in demand and supply is important for the empirical analysis of competition in oligopoly industries.

1.1. Example 1: Demand of storable goods. For a storable product, purchases in a given period (week, month) are not equal to consumption. When the price is low, consumers have incentives to buy a large amount to store the product and consume it in the future. When the price is high, or the household has a large inventory of the product, consumers do not buy an consume from his inventory. Dynamics arise because consumers' past purchases and consumption decisions impact their current inventory and therefore the benefits of purchasing today. Furthermore, consumers expectations about future prices also impact the perceived trade-offs of buying today versus in the future.

What are the implications of ignoring consumer dynamic behavior when we estimate the demand of differentiated storable products? An important implication is that we can get serious biases in the estimates of price demand elasticities. In particular, we can interpret a short-run intertemporal substitution as a long-run substitution between brands (or stores). To illustrate this issue, it is useful to consider an specific example. The following figure presents weekly times series data of prices and sales of canned tuna in a supermarket store. The time series of prices is characterized by "High-Low" pricing, what is quite common in many supermarkets. 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. Sales of this storable product respond to this type of dynamics in prices. As we can see in figure 6.1, most sales are concentrated at the very few days with low prices. Apparently, the short-run response of sales to these temporary price reductions is very large: the typical discount of a sales promotion is between 10% and 20%, and the increase in sales are around 300%.

Figure 6.1: Price promotions and sales of a storable good

In a static demand model, this type of respond would suggest that the price elasticity of demand of the product is very large. In particular, with these data the estimation of a static demand model provides estimates of own-price elasticities greater than 8. The static model interprets the large response of sales to a price reduction in terms of consumers substitution between brands (and to some extend between supermarkets too). Based on this estimates of demand elasticities, our model of competition would imply that price-cost margins are very small and firms (both supermarmets and brand manufacturers) have very little market power. A large degree of substitution between brands implies that product differentiation is small and market power is low.

This interpretation that ignores dynamics in consumer purchasing decision can be seriously wrong. Most of the short-run response of sales to a temporary price reduction is not substitution between brands or stores but intertemporal substitution in households' purchases. The temporary price reduction induces consumers to buy for storage today and to buy less in the future. The long-run substitution effect is much smaller, and it is this long-run effect what is relevant to measure firms' market power.

In order to distinguish between short-run and long-run responses to price changes, we have to specify and estimate a dynamic model of demand of differentiated products. In this type of models consumers are forward looking and take into account their expectations about future prices as well as storage costs.

1.2. Example 2: Demand of a new durable product. Melnikov (2000), Esteban and Shum (*RAND*, 2007), Carranza (2006), Gowrisankaran and Rysman (2009).

The price of new durable products typically declines over time during the months after the introduction of the product. Figure 6.2 illustrates this point for the case of *****. Different factors may explain this price decline, e.g., intertemporal price discrimination, increasing competition, exogenous cost decline, or endogenous cost decline due to learning by doing. As in the case of the "high-low" pricing of storable goods, explaining this pricing dynamics also requires one to take into account dynamics in supply. For the moment, we concentrate here in the demand. If consumers are forward looking, they expect the price will be lower in the future and this generates an incentive to wait and buying the good in the future.

Figure 6.2: Price decline of new durable products

A static model that ignores dynamics in demand of durable goods can introduce two different type of biases in the estimates of the distribution consumers willingness to pay and therefore of demand. The first source of bias comes from the failure to recognize that each period the potential market size is changing. Each period the demand curve is changing because some high willingness-to-pay consumers have already bought the product and left the market. A second source of bias comes from ignoring consumer forward-looking behavior. In the static model, consumers willingness-to-pay can is contaminated by consumers' willingness to wait because the expectation of future lower prices.

To illustrate the first source of bias, consider a market with an initial mass of 100 consumers and a uniform distribution of willingness to pay over the the unit interval. To concentrate on the first source of bias, consider that consumers are myopic and buy the product if the price is below their willingness to pay. Once consumers buy the product they are out of the market forever. Time is discrete and indexed by $t \in \{1, 2, ...\}$. Every period t, the aggregate demand is $Q_t = H_t \Pr(v_t \geq P_t) = H_t \left[1 - F_t\left(P_t\right)\right]$, where Q_t and P_t are quantity and price, respectively, H_t is the mass of consumers who remain in market at period t, and F_t is the distribution function of willingness to pay for consumers who remain in the market at period t. Suppose that we observe a sequence of prices equal to $P_1 = 0.9$, $P_2 = 0.8$, $P_3 = 0.7$, etc. Given this price sequence, it is easy to show that the demand curve at period $t = 1$ is $Q_1 = 100(1 - P_1)$, at period $t = 2$ the demand is $Q_2 = 90(\frac{0.9-P_2}{0.9}) = 100(0.9 - P_2)$, at period $t = 3$ it is $Q_3 = 80(\frac{0.8-P_3}{0.8}) = 100(0.8 - P_3)$, and so on. Therefore, the sequence of quantities is constant over time: $Q_1 = Q_2 = Q_3 = ... = 10$. A static demand model lead the researcher to conclude that consumers are not sensitive to price, since the same quantity is sold as prices decline. The estimate of the price elasticity would be zero. This example but it illustrates how ignoring dynamics in demand of durable goods can lead to serious biases in the estimates of the price sensitivity of demand.

1.3. Example 3: Product repositioning in differentiated product markets. A common assumption in many static (and dynamic) demand models of differentiated products is that product characteristics, other than prices, are exogenous. However, in many industries, product characteristics are very important strategic variables.

Ignoring the endogeneity of product characteristics has several implications. First, it can biases in the estimated demand parameters. A dynamic game that acknowledges the endogeneity of some product characteristics and exploits the dynamic structure of the model to generate valid moment conditions can deal with this problem.

A second important limitation of a static model of firm behavior is that it cannot recover the costs of repositioning product characteristics. As a result, the static model cannot address important empirical questions such as the effect of a merger on product repositioning. That is, the evaluation of the effects of a merger using a static model should assume that the product characteristics (other than prices) of the new merging firm would remain the same as before the merger. This is at odds both with the predictions of theoretical models and with informal empirical evidence. Theoretical models of horizontal mergers show that product repositioning is a potentially very important source of value for a merging firm, and informal empirical evidence shows that soon after a merger firms implement significant changes in their product portfolio.

Sweeting (2007) and Aguirregabiria and Ho (2009) are two examples of empirical applications that endogenize product attributes using a dynamic game of competition in a differentiated products industry. Sweeting estimates a dynamic game of oligopoly competition in the US commercial radio industry. The model endogenizes the choice of radio stations format (genre), and estimates product repositioning costs. Aguirregabiria and Ho (2009) propose and estimate a dynamic game of airline network competition where the number of direct connections that an airline has in an airport is an endogenous product characteristic.

1.4. Example 4: Dynamics of market structure. Ryan (2006) and Kasahara (JBES, 2010) provide excellent examples of how ignoring supply-side dynamics and firms' forward looking behavior can lead to misleading results.

Ryan (2006) studies the effects of the 1990 Amendments to the Clean Air Act on the US cement industry. This environmental regulation added new categories of regulated emissions, and introduced the requirement of an environmental certification that cement plants have to pass before starting their operation. Ryan estimates a dynamic game of competition where the sources of dynamics are sunk entry costs and adjustment costs associated with changes in installed capacity. The estimated model shows that the new regulation had negligible effects on variable production costs but it increased significantly the sunk cost of opening a new cement plant. A static analysis, that ignores the effects of the policy on firms' entry-exit decisions, would conclude that the regulation had negligible effects on firms profits and consumer welfare. In contrast, the dynamic analysis shows that the increase in sunk-entry costs caused a reduction in the number of plants that in turn implied higher markups and a decline in consumer welfare.

Kasahara (2010) proposes and estimates a dynamic model of firm investment in equipment and it uses the model to evaluate the effect of an important increase in import tariffs in Chile during the 1980s. The increase in tariffs had a substantial effect of the price of imported equipment and it may have a significant effect on firms' investment. An important feature of this policy is that the government announced that it was a temporary increase and that tariffs would go back to their original levels after few years. Kasahara shows that the temporary aspect of this policy exacerbated its negative effects on firm investment. Given that firms anticipated the future decline in import tariffs and the price of capital, a significant fraction of firms decided not invest and waiting until the reduction of tariffs. This waiting and inaction would not appear if the policy change were perceived as permanent. Kasahara shows that the Chilean economy would have recovered faster from the economic crisis of 1982-83 if the increase in tariffs would have been perceived as permanent.

1.5. Example 5: Dynamics of prices in a retail market. The significant cross-sectional dispersion of prices is a well-known stylized fact in retail markets. Retailing firms selling the same product, and operating in the same (narrowly defined) geographic market and at the same period of time, do charge prices that differ by significant amounts, e.g., 10% price differentials or even larger. This empirical evidence has been well established for gas stations and supermarkets, among other retail industries. Interestingly, the price differentials between firms, and the ranking of firms in terms prices, have very low persistence over time. A gas station that charges a price 5% below the average in a given week may be charging a price 5% above the average the next week. Using a more graphical description we can say that a firm's price follows a cyclical pattern, and the price cycles of the different firms in the market are not synchronized. Understanding price dispersion and the dynamics of price dispersion is very important to understand not only competition and market power but also for the construction of price indexes.

Different explanations have been suggested to explain this empirical evidence. Some explanations have to do with dynamic pricing behavior or "state dependence" in prices.

For instance, an explanation is based on the relationship between firm inventory and optimal price. In many retail industries with storable products, we observe that firms' orders to suppliers are infrequent. For instance, for products such as laundry detergent, a supermarket ordering frequency can be lower than one order per month. A simple and plausible explanation of this infrequency is that there are fixed or lump-sum costs of making an order that do not depend on the size of the order, or at least they do not increase proportionally with the size of the order. Then, inventories follow a so called (S,s) cycle: the increase by a large amount up to a maximum when a place is order and then they decline slowly up a minimum value where a new order is placed. Given this dynamics of inventories, it is simple to show that optimal price of the firm should also follow a cycle. The price drops to a minimum when a new order is placed and then increases over time up to a maximum just before the next order when the price drops again. Aguirregabiria (*REStud*, 1999) shows this joint pattern of prices and inventories for many products in a supermarket chain. I show that this type of inventory-depedence price dynamics can explain more than 20% of the time series variability of prices in the data.

CHAPTER 7

Single-Agent Models of Firm Investment

1. Model and Assumptions

To present some common features of dynamic structural models, we start with a simple model of firm investment that we can represent as a machine replacement model.

Suppose that we have panel data of N plants operating in the same industry with information on output, investment, and capital stock over T periods of time.

$$\text{Data} = \{\ Y_{it},\ I_{it},\ K_{it} : i = 1, 2, ..., N \text{ and } t = 1, 2, ..., T\ \}$$

Suppose that the investment data is characterized by infrequent and lumpy investments. That is, I_{it} contains a large proportion of zeroes (no investment), and when investment is positive the investment-to-capital ratio I_{it}/K_{it} is quite large. For instance, for some industries and samples we can find that the proportion of zeroes is above 60% (even with annual data!) and the average investment-to-capital ratio conditional on positive investment is above 50%.

A possible explanation for this type of dynamics in firms' investment is that there are significant individibilities in the purchases of new capital, or/and fixed or lump-sum costs associated with purchasing and installing new capital. Machine replacement models are models of investment that emphasize the existence of these indivisibilities and lump-sum costs of investment.

This type of investment models have been applied before in papers by Rust (Ectca, 1987), Das (REStud, 1991), Kennet (RAND, 1994), Rust and Rothwell (JAE, 1995), Cooper, Haltiwanger and Power (AER, 1999), Cooper and Haltiwanger (REStud 2006), and Kasahara (JBES, 2010), among others. In Rust (1987) the firm is a bus company (in Madison, Wisconsin), a plant is a bus, and a machine is a bus engine. Das (1991) considers cement firms and a plant is a cement kiln. In Kennet (1994) studies airline companies and the machine is an aircraft engine. Rust and Rothwell (1995) consider nuclear power plants. Cooper, Haltiwanger and Power (1999), Cooper and Haltiwanger (2006), and Kasahara (2010) consider manufacturing firms and investment in equipment in general.

We index plants by i and time by t. A plant's profit function is:

$$\Pi_{it} = Y_{it} - C_t\ I_{it} - RC_{it}$$

Y_{it} is the revenue of market value of the output produced by plant i at period t. I_{it} is the amount of investment at period t. C_t is the price of new capital. And RC_{it} represents investment costs other than the cost of purchasing the new capital, i.e., costs of replacing the old equipment (machine) by the new equipment.

Let K_{it} be the capital stock of plant i at the beginning of period t. As usual, capital depreciates exogenously and it increases when new investments are made. This transition rule of the capital stock is:

$$K_{it+1} = (1-\delta)\ (K_{it} + I_{it})$$

Following the key feature in models of machine replacement, we assume that there is an indivisibility in the investment decision. In the standard machine replacement model, the firm decides between zero investment ($I_{it} = 0$) or the replacement of the old capital by a "new machine" that implies a fixed amount of capital K^*. Therefore,

$$I_{it} \in \{0\ ,\ K^* - K_{it}\ \}$$

Therefore,

$$K_{it+1} = \begin{cases} (1-\delta)\ K_{it} & \text{if} \quad I_{it} = 0 \\ \\ (1-\delta)\ K^* & \text{if} \quad I_{it} > 0 \end{cases}$$

or

$$K_{it+1} = (1-\delta)\ \ [(1-a_{it})\ K_{it} + a_{it}\ K^*]$$

where a_{it} is the indicator of positive investment, i.e., $a_{it} \equiv 1\{I_{it} > 0\}$.

This implies that the possible values of the capital stock are $(1-\delta)K^*$, $(1-\delta)^2 K^*$, etc. Let T_{it} be the number of periods since the last machine replacement, i.e., time duration since the last time that investment was positive. There is a one-to-one relationship between capital K_{it} and the time duration T_{it}:

$$K_{it} = (1-\delta)^{T_{it}}\ K^*$$

or in logarithms, $k_{it} = k^* - d\ T_{it}$, where $k^* \equiv \log K^*$ and $d \equiv -\log(1-\delta) > 0$.

These assumptions on the values of investment and capital seem natural in applications where the investment decision is actually a machine replacement decision, as in the papers by Rust (1987), Das (1991), Kennet (1994), or Rust and Rothwell (1995), among others. However, this framework may be restrictive when we look at less specific investment decisions, such as investment in equipment as in the papers by Cooper, Haltiwanger and Power (1999), Cooper and Haltiwanger (2006), and Kasahara (2010). In these other papers, investment in the data is very lumpy, which is a prediction of a model of machine replacement, but firms in the sample have very different sizes (average over long periods of time) and their capital stocks in those periods with positive investment are very different. These papers consider that investment is either zero or a constant proportion of the installed capital, i.e.,

$I_{it} \in \{0 \;,\; q\; K_{it}\}$ where q is a constant, e.g., $q = 25\%$. Here I maintained the most standard assumption of machine replacement models.

The production function (actually, revenue function) is:

$$Y_{it} = \exp\left\{\; \alpha_0 + \eta_i^Y \right\} \;\; [(1 - a_{it})\; K_{it} + a_{it}\; K^*]^{\alpha_1}$$

where α_0 and α_1 are parameters, and η_i^Y captures productivity differences between firms that are time-invariant. The specification of the replacement cost function is:

$$RC_{it} = a_{it}\;\left(\; r(K_{it}) + \eta_i^C + \varepsilon_{it}\;\right)$$

$r(K)$ is a function that is increasing in K, and η_i^C and ε_{it} are zero mean random variables that captures firm heterogeneity in replacement costs. Therefore, the profit function is:

$$\Pi_{it} = \begin{cases} \exp\left\{\; \alpha_0 + \eta_i^Y \right\}\;\; K_{it}^{\alpha_1} & \text{if} \quad a_{it} = 0 \\ \\ \exp\left\{\; \alpha_0 + \eta_i^Y \right\}\;\; K^{*\alpha_1} - C_t\; I^* - r(K_{it}) - \eta_i^C - \varepsilon_{it} & \text{if} \quad a_{it} = 1 \end{cases}$$

Every period t, the firm observes the state variables K_{it}, C_t, and ε_{it} and then it decides its investment in order to maximize its expected value:

$$E_t\left(\sum\nolimits_{j=0}^{\infty} \beta^j\; \Pi_{i,t+j}\right)$$

where $\beta \in (0, 1)$ is the discount factor. The main trade-off in this machine replacement decision is simple. On the one hand, the productivity/efficiency of a machine declines over time and therefore the firm prefers younger machines. However, using younger machines requires frequent replacement and replacing a machine is costly.

The firm has uncertainty about future realizations of C_t and ε_{it}. To complete the model we have to specify the stochastic processes of these variables. We assume that C_t follows a Markov process with transition probability $f_C(C_{t+1}|C_t)$. For the shock in replacement costs ε_{it} we consider that it is i.i.d. with a logistic distribution with dispersion parameter σ_ε. The individual effects (η_i^Y, η_i^C) have a finite mixture distribution, i.e., (η_i^Y, η_i^C) is a pair of random variables from a distribution with discrete and finite support F_η.

Let $S_{it} = (K_{it},\, C_t,\, \varepsilon_{it})$ be the vector of state variables in the decision problem of a plant and let $V_i(S_{it})$ be the value function. This value function is the solution to the **Bellman equation**:

$$V_i(S_{it}) = \max_{a_{it} \in \{0,1\}} \left\{\; \Pi_i(a_{it}, S_{it}) + \beta \int V_i(S_{it+1})\; f_S(S_{it+1}|a_{it}, S_{it})\; dS_{it+1} \right\}$$

where $f_S(S_{it+1}|a_{it}, S_{it})$ is the (conditional choice) transition probability of the state variables:

$$f_S(S_{it+1}|a_{it}, S_{it}) = 1\left\{K_{it+1} = (1 - \delta)\;\; [(1 - a_{it})\; K_{it} + a_{it}\; K^*]\right\}\;\; f_C(C_{t+1}|C_t)\;\; f_\varepsilon(\varepsilon_{it})$$

where $1\{.\}$ is the indicator function, and f_ε is the density function of ε_{it}.

We can also represent the Bellman equation as:

$$V_i(S_{it}) = \max\{\ v_i(0; K_{it}, C_t)\ ;\ v_i(1; K_{it}, C_t) - \varepsilon_{it}\ \}$$

where $v_i(0; K_{it}, C_t)$ and $v_i(1; K_{it}, C_t)$ are the choice-specific value functions:

$$v_i(0; K_{it}, C_t) \equiv \exp\left\{\alpha_0 + \eta_i^Y\right\}\ K_{it}^{\alpha_1} + \beta \int V_i((1-\delta)K_{it}, C_{t+1}, \varepsilon_{it+1})\ f_C(C_{t+1}|C_t)\ df_\varepsilon(\varepsilon_{it})$$

$$v_i(1; K_{it}, C_t) \equiv \begin{array}{l} \exp\left\{\alpha_0 + \eta_i^Y\right\}\ K_{it}^{\alpha_1} - C_t\ I^* - r(K_{it}) - \eta_i^C \\ +\beta \int V_i((1-\delta)K^*, C_{t+1}, \varepsilon_{it+1})\ f_C(C_{t+1}|C_t)\ df_\varepsilon(\varepsilon_{it}) \end{array}$$

2. Solving the dynamic programming (DP) problem

For given values of structural parameters and functions, $\{\alpha_0,\ \alpha_1,\ r(.),\ f_C(.),\ \sigma_\varepsilon\}$, and of the individual effects η_i^Y and η_i^C, we can solve the DP problem of firm i by simply using successive approximations to the value function, i.e., iterations in the Bellman equation.

In models where some of the state variables are not serially correlated, it is computationally very convenient (and also convenient for the estimation of the model) to define versions of the value function and the Bellman equation that are integrated over the non-serially correlated variables. In our model, ε is not serially correlated state variables. The *integrated value function* of firm i is:

$$\bar{V}_i(K_{it}, C_t) \equiv \int V_i(K_{it}, C_t, \varepsilon_{it})\ df_\varepsilon(\varepsilon_{it})$$

And the integrated Bellman equation is:

$$\bar{V}_i(K_{it}, C_t) = \int \max\{\ v_i(0; K_{it}, C_t)\ ;\ v_i(1; K_{it}, C_t) - \varepsilon_{it}\ \}\, df_\varepsilon(\varepsilon_{it})$$

The main advantage of using the integrated value function is that it has a lower dimensionality than the original value function.

Given the extreme value distribution of ε_{it}, the integrated Bellman equation is:

$$\bar{V}_i(K_{it}, C_t) = \sigma_\varepsilon\ \ln\left[\exp\left\{\frac{v_i(0; K_{it}, C_t)}{\sigma_\varepsilon}\right\} + \exp\left\{\frac{v_i(1; K_{it}, C_t)}{\sigma_\varepsilon}\right\}\right]$$

where

$$v_i(0; K_{it}, C_t) \equiv \exp\left\{\alpha_0 + \eta_i^Y\right\}\ K_{it}^{\alpha_1} + \beta \int \bar{V}_i((1-\delta)K_{it}, C_{t+1})\ f_C(C_{t+1}|C_t)$$

$$v_i(1; K_{it}, C_t) \equiv \exp\left\{\alpha_0 + \eta_i^Y\right\}\ K_{it}^{\alpha_1} - C_t\ I^* - r(K_{it}) - \eta_i^C + \beta \int \bar{V}_i((1-\delta)K^*, C_{t+1})\ f_C(C_{t+1}|C_t)$$

The optimal decision rule of this dynamic programming (DP) problem is:

$$a_{it} = 1\{\ \varepsilon_{it} \leq v_i(1; K_{it}, C_t) - v_i(0; K_{it}, C_t)\ \}$$

Suppose that the price of new capital, C_t, has a discrete a finite range of variation: $C_t \in \{c^1, c^2, ..., c^L\}$. Then, the value function $\bar{V}_i$ can be represented as a $M \times 1$ vector in the Euclidean space, where $M = T * L$ and the T is the number of possible values for the capital stock. Let $\mathbf{V}_i$ be that vector. The integrated Bellman equation in matrix form is:

$$\mathbf{V}_i = \sigma_\varepsilon \ln \left(\exp\left\{ \frac{\mathbf{\Pi}_i(0) + \beta \ \mathbf{F}(0) \ \mathbf{V}_i}{\sigma_\varepsilon} \right\} + \exp\left\{ \frac{\mathbf{\Pi}_i(1) + \beta \ \mathbf{F}(1) \ \mathbf{V}_i}{\sigma_\varepsilon} \right\} \right)$$

where $\mathbf{\Pi}_i(0)$ and $\mathbf{\Pi}_i(1)$ are the $M \times 1$ vectors of one-period profits when $a_{it} = 0$ and $a_{it} = 1$, respectively. $\mathbf{F}(0)$ and $\mathbf{F}(0)$ are $M \times M$ transition probability matrices of (K_{it}, C_t) conditional on $a_{it} = 0$ and $a_{it} = 1$, respectively.

Given this equation, the vector $\mathbf{V}_i$ can be obtained by using value function iterations in the Bellman equation. Let $\mathbf{V}_i^0$ be an arbitrary initial value for the vector $\mathbf{V}_i$. For instance, $\mathbf{V}_i^0$ could be a $M \times 1$ vector of zeroes. Then, at iteration $k = 1, 2, ...$ we obtain:

$$\mathbf{V}_i^k = \sigma_\varepsilon \ln \left(\exp\left\{ \frac{\mathbf{\Pi}_i(0) + \beta \ \mathbf{F}(0) \ \mathbf{V}_i^{k-1}}{\sigma_\varepsilon} \right\} + \exp\left\{ \frac{\mathbf{\Pi}_i(1) + \beta \ \mathbf{F}(1) \ \mathbf{V}_i^{k-1}}{\sigma_\varepsilon} \right\} \right)$$

Since the (integrated) Bellman equation is a contraction mapping, this algorithm always converges (regardless the initial $\mathbf{V}_i^0$) and it converges to the unique fixed point. Exact convergence requires infinite iterations. Therefore, we stop the algorithm when the distance (e.g., Euclidean distance) between $\mathbf{V}_i^k$ and $\mathbf{V}_i^{k-1}$ is smaller than some small constant, e.g., 10^{-6}.

An alternative algorithm to solve the DP problem is the **Policy Iteration algorithm**. Define the *Conditional Choice Probability (CCP) function* $P_i(K_{it}, C_t)$ as:

$$\begin{aligned} P_i(K_{it}, C_t) \ &\equiv \ \Pr(\ \varepsilon_{it} \leq v_i(1; K_{it}, C_t) - v_i(0; K_{it}, C_t)\ ) \\ &= \frac{\exp\left\{ \dfrac{v_i(1; K_{it}, C_t) - v_i(0; K_{it}, C_t)}{\sigma_\varepsilon} \right\}}{1 + \exp\left\{ \dfrac{v_i(1; K_{it}, C_t) - v_i(0; K_{it}, C_t)}{\sigma_\varepsilon} \right\}} \end{aligned}$$

Given that (K_{it}, C_t) are discrete variables, we can describe the CCP function $P_i(.)$ as a $M \times 1$ vector of probabilities $\mathbf{P}_i$. The expression for the CCP in vector form is:

$$\mathbf{P}_i \ = \ \frac{\exp\left\{ \dfrac{\mathbf{\Pi}_i(1) - \mathbf{\Pi}_i(0) + \beta \ [\mathbf{F}(1) - \mathbf{F}(0)] \ \mathbf{V}_i}{\sigma_\varepsilon} \right\}}{1 + \exp\left\{ \dfrac{\mathbf{\Pi}_i(1) - \mathbf{\Pi}_i(0) + \beta \ [\mathbf{F}(1) - \mathbf{F}(0)] \ \mathbf{V}_i}{\sigma_\varepsilon} \right\}}$$

Suppose that the firm behaves according to the probs in $\mathbf{P}_i$. Let $\mathbf{V}_i^{\mathbf{P}}$ the vector of values if the firm behaves according to $\mathbf{P}$. That is $\mathbf{V}_i^{\mathbf{P}}$ is the expected discounted sum of current and future profits if the firm behaves according to $\mathbf{P}_i$. Ignoring for the moment the expected future $\varepsilon's$, we have that:

$$\mathbf{V}_i^{\mathbf{P}} = (1 - \mathbf{P}_i) * \left[\mathbf{\Pi}_i(0) + \beta \ \mathbf{F}(0)\mathbf{V}_i^{\mathbf{P}}\right] + \mathbf{P}_i * \left[\mathbf{\Pi}_i(1) + \beta \ \mathbf{F}(1)\mathbf{V}_i^{\mathbf{P}}\right]$$

And solving for $\mathbf{V}_i^{\mathbf{P}}$:

$$\mathbf{V}_i^{\mathbf{P}} = \left(I - \beta \ \mathbf{F}_i^{\mathbf{P}}\right)^{-1} \left((1 - \mathbf{P}_i) * \mathbf{\Pi}_i(0) + \mathbf{P}_i * \mathbf{\Pi}_i(1)\right)$$

where $\mathbf{F}_i^{\mathbf{P}} = (1 - \mathbf{P}_i) * \mathbf{F}(0) + \mathbf{P}_i * \mathbf{F}(1)$.

Taking into account this expression for $\mathbf{V}_i^{\mathbf{P}}$, we have that the optimal CCP $\mathbf{P}_i$ is such that:

$$\mathbf{P}_i = \frac{\exp\left\{\dfrac{\tilde{\mathbf{\Pi}}_i + \beta \ \tilde{\mathbf{F}} \ \left(I - \beta \ \mathbf{F}_i^{\mathbf{P}}\right)^{-1} \left((1 - \mathbf{P}_i) * \mathbf{\Pi}_i(0) + \mathbf{P}_i * \mathbf{\Pi}_i(1)\right)}{\sigma_\varepsilon}\right\}}{1 + \exp\left\{\dfrac{\tilde{\mathbf{\Pi}}_i + \beta \ \tilde{\mathbf{F}} \ \left(I - \beta \ \mathbf{F}_i^{\mathbf{P}}\right)^{-1} \left((1 - \mathbf{P}_i) * \mathbf{\Pi}_i(0) + \mathbf{P}_i * \mathbf{\Pi}_i(1)\right)}{\sigma_\varepsilon}\right\}}$$

where $\tilde{\mathbf{\Pi}}_i \equiv \mathbf{\Pi_i}(1) - \mathbf{\Pi}_i(0)$, and $\tilde{\mathbf{F}} \equiv \mathbf{F}(1) - \mathbf{F}(0)$. This equation defines a fixed point mapping in $\mathbf{P}_i$. This fixed point mapping is called the Policy Iteration mapping. This is also a contraction mapping. Optimal $\mathbf{P}_i$ is its unique fixed point.

Therefore we compute $\mathbf{P}_i$ by iterating in this mapping. Let $\mathbf{P}_i^0$ be an arbitrary initial value for the vector $\mathbf{P}_i$. For instance, $\mathbf{P}_i^0$ could be a $M \times 1$ vector of zeroes. Then, at each iteration $k = 1, 2, ...$ we do "two things":

Valuation step:

$$\mathbf{V}_i^k = \left(I - \beta \ \mathbf{F}^{\mathbf{P}_i^{k-1}}\right)^{-1} \left((1 - \mathbf{P}_i^{k-1}) * \mathbf{\Pi}_i(0) + \mathbf{P}_i^{k-1} * \mathbf{\Pi}_i(1)\right)$$

Policy step:

$$\mathbf{P}_i^k = \frac{\exp\left\{\dfrac{\tilde{\mathbf{\Pi}}_i + \beta \ \tilde{\mathbf{F}} \ \mathbf{V}_i^k}{\sigma_\varepsilon}\right\}}{1 + \exp\left\{\dfrac{\tilde{\mathbf{\Pi}}_i + \beta \ \tilde{\mathbf{F}} \ \mathbf{V}_i^k}{\sigma_\varepsilon}\right\}}$$

Policy iterations are more costly than Value function iterations (especially because the matrix inversion in the valuation step). However, the policy iteration algorithm requires a much lower number of iterations, especially with β is close to one. Rust (1987, 1994) proposes an hybrid algorithm: start with a few value function iterations and then switch to policy iterations.

3. Estimation

The primitives of the model are: (a) The parameters in the production function; (b) the replacement costs function $r(.)$; (c) the probability distribution of firm heterogeneity $F_\eta(.)$; (d) the dispersion parameter σ_ε; and (e) the discount factor β. Let θ represent the vector of structural parameters. We are interested in the estimation of θ.

Here I describe the Maximum Likelihood estimation of these parameters. Conditional on the observe history of price of capital and on the initial condition for the capital stock, we have that:

$$\Pr(Data \mid C,\, K_{i1},\, \theta) \;=\; \prod_{i=1}^{N} \Pr(a_{i1}, Y_{i1}, ...,\; a_{iT}, Y_{iT} \mid C,\, K_{i1},\, \theta)$$

The probability $\Pr(a_{i1}, Y_{i1}, ...,\; a_{iT}, Y_{iT} \mid C,\, K_{i1},\, \theta)$ is the contribution of firm i to the likelihood function. Conditional on the individual heterogeneity, $\eta_i \equiv (\eta_i^Y, \eta_i^C)$, we have that:

$$\begin{aligned}
\Pr(a_{i1}, Y_{i1}, ...,\; a_{iT}, Y_{iT} \mid C,\, K_{i1},\, \eta_i,\, \theta) &= \prod_{t=1}^{T} \Pr(a_{it}, Y_{it} \mid C_t,\, K_{it},\, \eta_i,\, \theta) \\
&= \prod_{t=1}^{T} \Pr(Y_{it} \mid a_{it},\, C_t,\, K_{it},\, \eta_i,\, \theta) \Pr(a_{it} \mid C_t,\, K_{it},\, \eta_i,\, \theta)
\end{aligned}$$

where $\Pr(a_{it} \mid C_t,\, K_{it},\, \eta_i,\, \theta)$ is the CCP function:

$$\Pr(a_{it} \mid C_t,\, K_{it},\, \eta_i,\, \theta) \;=\; P_i(K_{it}, C_t,\, \theta)^{a_{it}} \;\; [1 - P_i(K_{it}, C_t,\, \theta)]^{1-a_{it}}$$

and $\Pr(Y_{it} \mid a_{it},\, C_t,\, K_{it},\, \eta_i,\, \theta)$ comes from the production function, $Y_{it} = \exp\{\, \alpha_0 + \eta_i^Y \}$ $[(1 - a_{it})\; K_{it} + a_{it}\; K^*]^{\alpha_1}$. In logarithms, the production function is:

$$\ln Y_{it} = \alpha_0 + \alpha_1\; (1 - a_{it}) \ln K_{it} + \kappa\; a_{it} + \eta_i^Y + e_{it}$$

where κ is a parameter that represents $\alpha_1 \ln K^*$, and e_{it} is a measurement error in output, that we assume i.i.d. $N(0, \sigma_e^2)$ and independent of ε_{it}. Therefore,

$$\Pr(Y_{it} \mid a_{it},\, C_t,\, K_{it},\, \eta_i,\, \theta) \;=\; \phi\left(\frac{\ln Y_{it} - \alpha_0 - \alpha_1\; (1 - a_{it}) \ln K_{it} - \kappa\; a_{it} - \eta_i^Y}{\sigma_e} \right)$$

where $\phi(.)$ is the PDF of the standard normal distribution.

Putting all these pieces together, we have that the log-likelihood function of the model is $\ell(\theta) = \sum_{i=1}^{N} \ln \mathcal{L}_i(\theta)$ where $\mathcal{L}_i(\theta) \equiv \Pr(a_{i1}, Y_{i1}, ...,\; a_{iT}, Y_{iT} \mid C,\, K_{i1},\, \theta)$ and:

$$\mathcal{L}_i(\theta) = \sum_{\eta \in \Omega} F_\eta(\eta) \left[\begin{array}{c} \displaystyle\prod_{t=1}^{T} \phi\left(\frac{\ln Y_{it} - \alpha_0 - \alpha_1\; (1 - a_{it}) \ln K_{it} - \kappa\; a_{it} - \eta^Y}{\sigma_e} \right) \\ P_i(K_{it}, C_t,\, \eta,\, \theta)^{a_{it}} \;\; [1 - P_i(K_{it}, C_t,\, \eta,\, \theta)]^{1-a_{it}} \end{array} \right]$$

Given this likelihood, we can estimate by Maximum Likelihood (ML)

The NFXP algorithm is a gradient iterative search method to obtain the MLE of the structural parameters.

This algorithm nests a BHHH method (outer algorithm), that searches for a root of the likelihood equations, with a value function or policy iteration method (inner algorithm), that solves the DP problem for each trial value of the structural parameters. The algorithm is initialized with an arbitrary vector $\hat{\boldsymbol{\theta}}_0$.

A BHHH iteration is defined as:

$$\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k + \left(\sum_{i=1}^{N} \nabla l_i(\hat{\boldsymbol{\theta}}_k) \nabla l_i(\hat{\boldsymbol{\theta}}_k)' \right)^{-1} \left(\sum_{i=1}^{N} \nabla l_i(\hat{\boldsymbol{\theta}}_k) \right)$$

where $\nabla l_i(\boldsymbol{\theta})$ is the gradient in $\boldsymbol{\theta}$ of the log-likelihood function for individual i. In a partial likelihood context, the score $\nabla l_i(\boldsymbol{\theta})$ is:

$$\nabla l_i(\boldsymbol{\theta}) = \sum_{t=1}^{T_i} \nabla \log P(a_{it}|x_{it}, \boldsymbol{\theta})$$

To obtain this score we have to solve the DP problem.

In our machine replacement model:

$$l(\boldsymbol{\theta}) \;\; = \;\; \sum_{i=1}^{N} \sum_{t=1}^{T_i} a_{it} \;\; \log P(x_{it}, \theta) + (1 - a_{it}) \;\; \log(1 - P(x_{it}, \theta))$$

with:

$$\mathbf{P}(\boldsymbol{\theta}) \;\; = \;\; F_{\tilde{\varepsilon}} \left(\begin{array}{c} [\theta_{Y0} + \theta_{Y1}\mathbf{X} + \beta \; \mathbf{F}_x(0)\mathbf{V}(\boldsymbol{\theta})] \\ - [\theta_{Y0} - \theta_{R0} - \theta_{Y1}\mathbf{X} + \beta \; \mathbf{F}_x(1)\mathbf{V}(\boldsymbol{\theta})] \end{array} \right)$$

The NFXP algorithm works as follows. At each iteration we can distinguish three main tasks or steps.

Step 1: Inner iteration: DP solution. Given $\hat{\boldsymbol{\theta}}_0$, we obtain the vector $\bar{\mathbf{V}}(\hat{\boldsymbol{\theta}}_0)$ by using either successive iterations or policy iterations.

Step 2: Construction of scores. Then, given $\hat{\boldsymbol{\theta}}_0$ and $\bar{\mathbf{V}}(\hat{\boldsymbol{\theta}}_0)$ we construct the choice probabilities

$$\mathbf{P}(\hat{\boldsymbol{\theta}}_0) \;\; = \;\; F_{\tilde{\varepsilon}} \left(\begin{array}{c} \left[\theta_{Y0} + \theta_{Y1}\mathbf{X} + \beta \; \mathbf{F}_x(0)\mathbf{V}(\hat{\boldsymbol{\theta}}_0)\right] \\ - \left[\theta_{Y0} - \theta_{R0} - \theta_{Y1}\mathbf{X} + \beta \; \mathbf{F}_x(1)\mathbf{V}(\hat{\boldsymbol{\theta}}_0)\right] \end{array} \right)$$

the Jacobian $\dfrac{\partial \bar{\mathbf{V}}(\hat{\boldsymbol{\theta}}_0)'}{\partial \theta}$and the scores $\nabla l_i(\hat{\boldsymbol{\theta}}_0)$

Step 3: BHHH iteration. We we use the scores $\nabla l_i(\hat{\boldsymbol{\theta}}_0)$ to make a new BHHH iteration to obtain $\hat{\boldsymbol{\theta}}_1$.

$$\hat{\boldsymbol{\theta}}_1 = \hat{\boldsymbol{\theta}}_0 + \left(\sum_{i=1}^{N} \nabla l_i(\hat{\boldsymbol{\theta}}_0) \nabla l_i(\hat{\boldsymbol{\theta}}_0)' \right) \left(\sum_{i=1}^{N} \nabla l_i(\hat{\boldsymbol{\theta}}_0) \right)$$

Then, we replace $\hat{\boldsymbol{\theta}}_0$ by $\hat{\boldsymbol{\theta}}_1$ and go back to step 1.

* We repeat stesp 1 to 3 until convergence: i.e., until the distance between $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_0$ is smaller than a pre-specified convergence constant.

The main advantages of the NFXP algorithm are its conceptual simplicity and, more importantly, that it provides the MLE which is the most efficient estimator asymptotically under the assumptions of the model.

The main limitation of this algorithm is its computational cost. In particular, the DP problem should be solved for each trial value of the structural parameters.

4. Patent Renewal Models

• **What is the value of a patent? How to measure it?**

• The valuation of patents is very important for: merger & acquisition decisions; using patents as collateral for loans; value of innovations; value of patent protection.

• Very few patents are traded, and there is substantial selection. We cannot use an "hedonic" approach.

• The number of citations of a patent is a very imperfect measure of patent value.

• Multiple patents are used in the production of multiple products, and in generating new patents. A "production function approach" seems also unfeasible.

• Pakes (1986) proposes using information on patent renewal fees together with a *Reveal Preference approach* to estimate the value of a patent.

• Every year, a patent holder should pay a renewal fee to keep her patent.

• If the patent holder decides to renew, it is because her expected value of holding the patent is greater than the renewal fee (that is publicly known).

• Therefore, observed decisions on patent renewal / non renewal contain information on the value of a patent.

Model: Basic Framework

• Consider a patent holder who has to decide whether to renew her patent or not. We index patents by i.

• This decision should be taken at ages $t = 1, 2, ..., T$ where $T < \infty$ is the regulated term of a patent (e.g., 20 years in US, Europe, or Canada).

• Patent regulation also establishes a sequence of **Maintenance Fees** $\{c_t : t = 1, 2, ..., T\}$. This sequence of renewal fees is deterministic such that a patent owner knows with certainty future renewal fees.

• The schedule $\{c_t : t = 1, 2, ..., T\}$ is typically increasing in patent age t and it may god from a few hundred dollars to a few thousand dollars.

• A patent generates a sequence of profits $\{\pi_{it} : t = 1, 2, ..., T\}$.
• At age t, a patent holder knows current profit π_{it} but has uncertainty about future profits $\pi_{i,t+1}$, $\pi_{i,t+2}$, ...
• The evolution of profits depends on the following factors:
(1) the initial "quality" of the idea/patent;
(2) innovations (new patents) which are substitutes of the patent and therefore, depreciate its value or even make it obsolete;
(3) innovations (new patents) which are complements of the patent and therefore, increase its value.

Stochastic process of patent profits
• Pakes proposes the following stochastic process for profits, that tries to capture the three forces mentioned above.
• A patent profit at the first period is a random draw from a log-normal distribution with parameters μ_1 and σ_1:

$$\ln(\pi_{i1}) \sim N(\mu_1, \sigma_1^2)$$

• After the first year, profit evolves according to the following formula:

$$\pi_{i,t+1} = \tau_{i,t+1} \ \max\left\{ \ \delta \ \pi_{it} \ ; \ \xi_{i,t+1} \right\}$$

• $\delta \in (0, 1)$ is the depreciation rate. In the absence of unexpected shocks, the value of the patent depreciates according to the rule: $\pi_{i,t+1} = \delta \ \pi_{it}$.
• $\tau_{i,t+1} \in \{0, 1\}$ is a binary variable that represents that the patent becomes obsolete (i.e., zero value) due to competing innovations. The probability of this event is a decreasing function of profit at previous year:

$$\Pr(\tau_{i,t+1} = 0 \mid \pi_{it}, t) = \exp\{-\lambda \ \pi_{it}\}$$

• The largest is the profit of the patent at age t, the smallest is the probability that it becomes obsolete.
• Variable $\xi_{i,t+1}$ represents innovations which are complements of the patent and increase its profitability.
• $\xi_{i,t+1}$ has an exponential distribution with mean γ and standard deviation $\phi^t\sigma$:

$$p(\xi_{i,t+1} \mid \pi_{it}, t) = \frac{1}{\phi^t\sigma} \exp\left\{-\frac{\gamma + \xi_{i,t+1}}{\phi^t\sigma}\right\}$$

• If $\phi < 1$, the variance of $\xi_{i,t+1}$ declines over time (and the $E(\max\left\{ \ x \ ; \ \xi_{i,t+1}\right\})$ value declines as well).

• If $\phi > 1$, the variance of $\xi_{i,t+1}$ increases over time (and the $E(\max\{\ x\ ;\ \xi_{i,t+1}\})$ value increases as well).

• Under this specification, profits $\{\pi_{it}\}$ follow a non-homogeneous Markov process with initial density $\pi_{i1} \sim \ln N(\mu_1, \sigma_1^2)$, and transition density function:

$$f_\varepsilon\left(\pi_{it+1}|\pi_{it}, t\right) = \begin{cases} \exp\{-\lambda\ \pi_{it}\} & \text{if} \quad \pi_{it+1} = 0 \\ \Pr\left(\xi_{it+1} < \delta\pi_{it}\ |\ \pi_{it}, t\right) & \text{if} \quad \pi_{it+1} = \delta\pi_{it} \\ \dfrac{1}{\phi^t\sigma}\exp\left\{-\dfrac{\gamma + \pi_{it+1}}{\phi^t\sigma}\right\} & \text{if} \quad \pi_{it+1} > \delta\pi_{it} \end{cases}$$

• The vector of structural parameters is $\theta = (\lambda, \delta, \gamma, \phi, \sigma, \mu_1, \sigma_1)$.

Model: Dynamic Decision Model

• $V_t(\pi)$ is the value of an active patent of age t and current profit π.

• Let $a_{it} \in \{0, 1\}$ be the decision variable that represents the event "the patent owner decides to renew the patent at age t".

• The value function is implicitly defined by the Bellman equation:

$$V_t(\pi_{it}) = \max\left\{\ 0\ ;\ \pi_{it} - c_t + \beta\int V_{t+1}(\pi_{i,t+1})\ f_\varepsilon(d\pi_{i,t+1}\ |\ \pi_{it}, t)\ \right\}$$

with $V_t(\pi_{it}) = 0$ for any $t \geq T + 1$.

• The value of not renewal ($a_{it} = 0$) is zero. The value of renewal ($a_{it} = 1$) is the current profit $\pi_{it} - c_t$ plus the expected and discounted future value.

Model: Solution (Backwards induction)

• We can use backwards induction to solve for the sequence of value functions $\{V_t\}$ and optimal decision rules $\{\alpha_t\}$:

• Starting at age $t = T$, for any profit π:

$$V_T(\pi) = \max\{\ 0\ ;\ \pi - c_T\}$$

and

$$\alpha_T(\pi) = 1\{\ \pi - c_T \geq 0\ \}$$

• Then, for age $t < T$, and for any profit π:

$$V_t(\pi) = \max\left\{\ 0\ ;\ \pi - c_t + \beta\int V_{t+1}(\pi')\ f_\varepsilon(d\pi'|\pi, t)\ \right\}$$

and

$$\alpha_t(\pi) = 1\left\{\ \pi - c_t + \beta\int V_{t+1}(\pi')\ f_\varepsilon(d\pi'|\pi, t) \geq 0\ \right\}$$

Solution - A useful result

• Given the form of $f_{\varepsilon}(\pi'|\pi,t)$, the future and discounted expected value, $\beta \int V_{t+1}(\pi')$ $f_{\varepsilon}(d\pi'|\pi,t)$, is increasing in current π.
• This implies that the solution of the DP problem can be described as a **sequence of threshold values for profits** $\{\pi_t^* : t = 1, 2, ..., T\}$ such that the optimal decision rule is:

$$\alpha_t(\pi) = 1\{\ \pi \geq \pi_t^*\ \}$$

• π_t^* is the level of current profits that leaves the owner indifferent between renewing the patent or not: $V_t(\pi_t^*) = 0$.
• These threshold values are obtained using backwards induction:
• At period $t = T$:

$$\pi_T^* = c_T$$

• At period $t < T$, π_t^* is the unique solution to the equation:

$$\pi_t^* - c_t + E\left(\sum_{s=t+1}^{T} \beta^{s-t}\ \max\{\ 0\ ;\ \pi_{t+1} - \pi_{t+1}^*\ \} \mid \pi_t = \pi_t^* \right) = 0$$

• Solving for a sequence of threshold values is much simpler that solving for a sequence of value functions.

Data

• Sample of N patents with complete (uncensored) durations $\{d_i : i = 1, 2, ...N\}$, where $d_i \in \{1, 2, ..., T+1\}$ is patent i's duration or age at its last renewal period.
• The information in this sample can be summarized by the empirical distribution of $\{d_i\}$:

$$\widehat{p}(t) = \frac{1}{N}\sum_{i=1}^{N} 1\{d_i = t\}$$

Estimation: Likelihood

• The log-likelihood function of this model and data is:

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{N}\sum_{t=1}^{T+1} 1\{d_i = t\}\ \ln \Pr(d_i = t|\theta) \\ &= N\sum_{t=1}^{T+1} \widehat{p}(t)\ \ln P(t|\theta) \end{aligned}$$

where:

$$\begin{aligned} P(t|\theta) &= \Pr\left(\pi_s \geq \pi_s^* \text{ for } s \leq t-1, \text{and } \pi_t < \pi_t^* \mid \theta\right) \\ &= \int_{\pi_1^*}^{\infty} ... \int_{\pi_{t-1}^*}^{\infty} \int_{0}^{\pi_t^*} dF(\pi_1, ..., \pi_{t-1}, \pi_t) \end{aligned}$$

• Computing $P(t|\theta)$ involves solving an integral of dimension t. For t greater than 4 or 5, it is computationally very costly to obtain the exact value of these probabilities. Instead, we approximate these probabilities using Monte Carlo simulation.

Estimation: Simulation of Probabilities

• For a given value of θ, let $\{\pi_t^{sim}(\theta) : t = 1, 2, ..., T\}$ be a simulated history of profits for patent i.

• Suppose that, for a given value of θ, we simulate R **independent** profit histories. Let $\{\pi_{rt}^{sim}(\theta) : t = 1, 2, ..., T;\ r = 1, 2, ..., R\}$ be these histories.

• Then, we can approximate the probability $P(t|\theta)$ using the following simulator:

$$\tilde{P}_R(t|\theta) = \frac{1}{R} \sum_{r=1}^{R} 1\{\pi_{rs}^{sim}(\theta) \geq \pi_s^* \text{ for } s \leq t-1, \text{and } \pi_{rt}^{sim} < \pi_t^*\}$$

• $\tilde{P}_R(t|\theta)$ is a *raw frequency simulator*. It has the following properties (Note that these are properties of a simulator, not of an estimator. $\tilde{P}_R(t|\theta)$ does not depend on the data).

(1) Unbiased: $E\left(\tilde{P}_R(t|\theta)\right) = P(t|\theta)$

(2) $Var(\tilde{P}_R(t|\theta)) = P(t|\theta)(1 - P(t|\theta))/R$

(3) Consistent as $R \to \infty$.

• It is possible to obtain better simulators (with lower variance) by using importance-sampling simulation. This is relevant because the bias and variance of simulated-based estimators depend on the variance (and bias) of the simulator.

• Furthermore, when $P(t|\theta)$ is small, the simulator $\tilde{P}_R(t|\theta)$ can be zero even when R is large, and this creates problems for ML estimation.

• A simple solution to this problem is to consider the following simulator which is based on the raw-frequency simulated probabilities $\tilde{P}_R(1|\theta)$, $\tilde{P}_R(2|\theta)$, ... $\tilde{P}_R(T+1|\theta)$:

$$P_R^*(t|\theta) = \frac{\exp\left\{\dfrac{\tilde{P}_R(t|\theta)}{\eta}\right\}}{\sum_{s=1}^{T+1} \exp\left\{\dfrac{\tilde{P}_R(s|\theta)}{\eta}\right\}}$$

where $\eta > 0$ is an smoothing parameter.

• The simulator P_R^* is biased. However, if $\eta \to 0$ as $R \to \infty$, then P_R^* is consistent, it has lower variance than $\tilde{P}_R$, and it is always strictly positive.

Simulation-Based Estimation

• The estimator of θ (Simulated Method of Moments estimator) is the value that solves the system of T equations: for $t = 1, 2, ...T$:

$$\frac{1}{N}\sum_{i=1}^{N}\left[1\{d_i = t\} - \tilde{P}_{R,i}(t|\theta)\right] = 0$$

where the subindex i in the simulator $\tilde{P}_{R,i}(t|\theta)$ indicates that for each patent i in the sample we draw R independent histories and compute independent simulators.

• **Effect of simulation error.** Note that $\tilde{P}_{R,i}(t|\theta)$ is unbiased such that $\tilde{P}_{R,i}(t|\theta) = P(t|\theta)+ e_i(t,\theta)$, where $e_i(t,\theta)$ is the simulation error. Since the simulation errors are independent random draws:

$$\frac{1}{N}\sum_{i=1}^{N} e_i(t,\theta) \to_p 0 \quad \text{and} \quad \frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_i(t,\theta) \to_d N(0, V_R)$$

The estimator is consistent an asymptotically normal for any R. The variance of the estimator declines with R.

Identification

• Since there are only 20 different values for the renewal fees $\{c_t\}$ we can at most identify 20 different points in the probability distribution of patent values.

• The estimated distribution at other points is the result of interpolation or extrapolation based on the functional form assumptions on the stochastic process for profits.

• It is important to note that the identification of the distribution of patent values is NOT up to scale but in dollar values.

• For a given patent of with age t, all what we can say is that: if $a_{it} = 0$, then $V_{it} < V(\pi_t^*)$; and if $a_{it} = 1$, then $V_{it} \geq V(\pi_t^*)$.

Empirical Questions

• The estimated model can be used to address important empirical questions.

• **Valuation of the stock of patents.** Pakes uses the estimated model to obtain the value of the stock of patents in a country.

• According to the estimated model, the value of the stock of patents in 1963 was \$315 million in France, \$385 million in UK, and \$511 in Germany.

• Combining these figures with data on R&D investments in these countries, Pakes calculates rates of return of 15.6%, 11.0% and 13.8%, which look like quite reasonable.

Empirical Questions

• **Factual policies.** The estimated model shows that a very important part of the observed between-country differences in patent renewal can be explained by differences in policy parameters (i.e., renewal fees and maximum length).

• **Counterfactual policy experiments.** The estimated model can be used to evaluate the effects of policy changes (in renewal fees and/or in maximum length) which are not observed in the data.

5. Dynamic structural models of temporary sales and inventories

Recent empirical papers show that temporary sales account for approximately half of all price changes of retail products in US (Hosken and Reiffen, 2004, Nakamura and Steinsson, 2008, Midrigan, 2011). Understanding the determinants of temporary sales is important to understand price stickiness and price dispersion, and it has important implications on the effects of monetary policy. It has also important implications in the study of firms' market power and competition.

Here I describe three different models of sales promotions based on the papers by Slade (1998), Aguirregabiria (1999), Pesendorfer (2002), and Kano (2013).

5.1. Slade (1998).

5.2. Aguirregabiria (1999). • The significant cross-sectional dispersion of prices is a well-known stylized fact in retail markets. Retailing firms selling the same product, and operating in the same (narrowly defined) geographic market and at the same period of time, do charge prices that differ by significant amounts, e.g., 10% price differentials or even larger. This empirical evidence has been well established for gas stations and supermarkets, among other retail industries. Interestingly, the price differentials between firms, and the ranking of firms in terms prices, have very low persistence over time. A gas station that charges a price 5% below the average in a given week may be charging a price 5% above the average the next week. Using a more graphical description we can say that a firm's price follows a cyclical pattern, and the price cycles of the different firms in the market are not synchronized. Understanding price dispersion and the dynamics of price dispersion is very important to understand not only competition and market power but also for the construction of price indexes.

• Different explanations have been suggested to explain this empirical evidence. Some explanations have to do with dynamic pricing behavior or "state dependence" in prices.

• For instance, an explanation is based on the relationship between firm inventory and optimal price. In many retail industries with storable products, we observe that firms' orders to suppliers are infrequent. For instance, for products such as laundry detergent, a supermarket ordering frequency can be lower than one order per month. A simple and plausible explanation of this infrequency is that there are fixed or lump-sum costs of making an order that do not depend on the size of the order, or at least they do not increase proportionally with the size of the order. Then, inventories follow a so called (S,s) cycle: the increase by a large amount up to a maximum when a place is order and then they decline slowly up a minimum value where a new order is placed. Given this dynamics of inventories, it is simple to show that optimal price of the firm should also follow a cycle. The price drops to a minimum when a new order is placed and then increases over time up to a maximum just before the next order when the price drops again. Aguirregabiria (*REStud*, 1999) shows this joint pattern of prices and inventories for many products in a supermarket chain. I show that this type of inventory-depedence price dynamics can explain more than 20% of the time series variability of prices in the data.

Temporary sales and inventories

• Recent empirical papers show that temporary sales account for approximately half of all price changes of retail products in US: Hosken and Reiffen (RAND, 2004); Nakamura and Steinsson (QJE, 2008); Midrigan (Econometrica, 2011).
• Understanding the determinants of temporary sales is important to understand price stickiness and price dispersion, and it has important implications on the effects of monetary policy.
• It has also important implications in the study of firms' market power and competition.
• Different empirical models of sales promotions: Slade (1998) [Endogenous consumer loyalty], Aguirregabiria (1999) [Inventories], Pesendorfer (2002) [Intertemporal price discrimination], and Kano (2013).

• This paper studies how retail inventories, and in particular (S,s) inventory behavior, can explain both price dispersion and sales promotions in retail markets.
• Three factors are key for the explanation provided in this paper:

(1) Fixed (lump-sum) ordering costs, that generates (S,s) inventory behavior.

(2) Demand uncertainty.

(3) Sticky prices (Menu costs) that, together with demand uncertainty, creates a positive probability of excess demand (stockout).

Model: Basic framework

• Consider a retail firm selling a product. We index products by i.

• Every period (month) t the firm decides the retail price and the quantity of the product to order to manufacturers/wholesalers

• **Monthly sales** are the minimum of supply an demand:

$$y_{it} = \min\{\ d_{it}\ ;\ s_{it} + q_{it}\ \}$$

- y_{it} = sales in physical units
- d_{it} = demand
- s_{it} = inventories at the beginning of month t
- q_{it} = orders (and deliveries) during month t

Demand and Expected sales

• The firm has uncertainty about **current demand**:

$$d_{it} = d_{it}^e\ \exp(\xi_{it})$$

- d_{it}^e = expected demand
- ξ_{it} = zero mean demand shock unknown to the firm at t.

• Therefore, **expected sales** are:

$$y_{it}^e = \int \min\{d_{it}^e\ \exp(\xi)\ \ ;\ s_{it} + q_{it}\ \}\, dF_\xi(\xi)$$

• Assume monopolistic competition. **Expected Demand** depends on the own price, p_{it}, and a demand shock ω_{it}. The functional form is isoelastic:

$$d_{it}^e = \exp\{\gamma_0 - \gamma_1\ \ln(p_{it}) + \omega_{it}\ \}$$

where γ_0 and $\gamma_1 > 0$ are parameters.

Price elasticity of expected sales

• **Demand uncertainty** has important implications for the relationship between prices and inventories.

• The price elasticity of expected sales is a function of the **supply-to-expected-demand ratio** $(s_{it} + q_{it})/d_{it}^e$:

$$\begin{aligned}\eta_{y^e|p} \equiv \frac{-\partial y^e}{\partial p}\frac{p}{y^e} &= -\left[\int I\{d^e \exp(\xi)\ \ ;\ s+q\ \}\, dF_\xi(\xi)\right] \frac{\partial d^e}{\partial p}\frac{p}{y^e} \\ &= \gamma_1\ F_\xi\left(\log\left[\frac{s+q}{d^e}\right]\right)\frac{d^e}{y^e}\end{aligned}$$

• And we have that:

$$\eta_{y^e|p} \longrightarrow \begin{cases} \gamma_1 & \text{as} \quad (s+q)/d^e \longrightarrow \infty \\ 0 & \text{as} \quad (s+q)/d^e \longrightarrow 0 \end{cases}$$

Price elasticity of expected sales

$$\eta_{y^e|p} = \gamma_1 \; F_{\xi}\left(\log\left[\frac{s+q}{d^e}\right]\right)\frac{d^e}{y^e}$$

[FIGURE: $\eta_{y^e|p}$ increasing in $\dfrac{s+q}{d^e}$, with asymptote at γ_1]

• When the supply-to-expected-demand ratio is large, the probability of stockout is very small and $y^e \simeq d^e$, so the elasticity of expected sales is just the elasticity of demand.

• However, when the supply-to-expected-demand ratio is small, the probability of stockout is large and the elasticity of expected sales can be much lower than the elasticity of demand.

Markup and inventories (myopic case)

• This has potentially important implications for the optimal price of an oligopolistic firm.

• To give some intuition, consider the pricing decision of the monopolistic firm without forward-looking behavior. That optimal price is:

$$\frac{p-c}{p} = \frac{1}{\eta_{y^e|p}}$$

OR

$$\frac{p-c}{c} = \frac{1}{\eta_{y^e|p}-1}$$

• Variability over time in the supply-to-expected-demand ratio can generate significant fluctuations in price-cost margins. It can also explain temporary sales promotions.

• That can be the case under (S, s) inventory behavior.

Evolution of inventories and price without menu cost

Evolution of inventories and price with menu cost

Empirical Application

• The paper investigates this hypothesis using a data from a supermarket chain, with rich information on prices, sales, inventories, orders, and wholesale prices for many different products.

• Reduced form estimations present evidence that supports the hypothesis:

(1) Prices depend negatively and very significantly on the level of inventories.

(2) Inventories of many products follow (S,s) cycles.

(3) Price cost margins increase at the beginning of an (S,s) cycle, and decline monotonically during the cycle.

• I estimate the parameters in the profit function (demand parameters, ordering costs, inventory holding costs) and use the estimated model to analyze how much of price variation and temporary sales promotions can be explained by firm inventories.

Profit function

• **Expected current profits** are equal to expected revenue, minus ordering costs, inventory holding costs and price adjustment costs:

$$\pi_{it} = p_{it}\ y_{it}^{e} - OC_{it} - IC_{it} - PAC_{it}$$

- OC_{it} = ordering costs
- IC_{it} = inventory holding costs
- PAC_{it} = price adjustment (menu) costs

• **Ordering costs:**

$$OC_{it} = \begin{cases} 0 & \text{if} \quad q_{it} = 0 \\ F_{oc} + \varepsilon_{it}^{oc} - c_{it}\ q_{it} & \text{if} \quad q_{it} > 0 \end{cases}$$

- F_{oc} = fixed (lump-sum) ordering cost. Parameter.
- ε_{it}^{oc} = zero mean shock in the fixed ordering cost.
- c_{it} = wholesale price

• **Inventory holding costs:**

$$IC_{it} = \alpha\ s_{it}$$

• **Menu costs:**

$$PAC_{it} = \begin{cases} 0 & \text{if} \quad p_{it} = p_{i,t-1} \\ F_{mc}^{(+)} + \varepsilon_{it}^{mc(+)} & \text{if} \quad p_{it} > p_{i,t-1} \\ F_{mc}^{(-)} + \varepsilon_{it}^{mc(-)} & \text{if} \quad p_{it} < p_{i,t-1} \end{cases}$$

- $F_{mc}^{(+)}$ and $F_{mc}^{(-)}$ are price adjustment cost parameters
- $\varepsilon_{it}^{mc(+)}$ and $\varepsilon_{it}^{mc(-)}$ are zero mean shocks in menu costs

State variables

• The state variables of this DP problem are:

$$\left\{ \underbrace{s_{it},\ c_{it},\ p_{i,t-1},\ \omega_{it}}_{x_{it}},\ \underbrace{\varepsilon_{it}^{oc},\ \varepsilon_{it}^{mc(+)},\ \varepsilon_{it}^{mc(+)}}_{\varepsilon_{it}} \right\}$$

• The decision variables are q_{it} and $\Delta p_{it} \equiv p_{it} - p_{i,t-1}$. We use a_{it} to denote $(q_{it},\Delta p_{it})$.

• Let $V(x_{it}, \varepsilon_{it})$ be the value of the firm associated with product i. This value function solves the Bellman equation:

$$V(x_{it}, \varepsilon_{it}) = \max_{a_{it}} \left\{ \begin{array}{c} \pi(a_{it}, x_{it}, \varepsilon_{it}) \\ \\ +\beta \int V(x_{i,t+1}, \varepsilon_{i,t+1}) \; dF(x_{i,t+1}, \varepsilon_{i,t+1} | a_{it}, x_{it}, \varepsilon_{it}) \end{array} \right\}$$

Discrete Decision variables

• Most of the variability of q_{it} and Δp_{it} in the data is discrete. For simplicity, we assume that these variables have a discrete support.

$$\begin{array}{rcl} q_{it} & \in & \{0 \ , \ \kappa_i\} \\ \\ \Delta p_{it} & \in & \{0 \ , \ \delta_i^{(+)}, \ \delta_i^{(-)}\} \end{array}$$

where $\kappa_i > 0$, $\delta_i^{(+)} > 0$, and $\delta_i^{(-)} < 0$ are parameters.

• Therefore, the set of choice alternatives at every period t is:

$$a_{it} \in A = \left\{ \ (0,0), \ (0, \delta_i^{(+)}), \ (0, \delta_i^{(-)}), \ (\kappa_i, 0), \ (\kappa_i, \delta_i^{(+)}), \ (\kappa_i, \delta_i^{(-)}) \ \right\}$$

• The transition rules for the state variables are:

$$\begin{array}{rcl} s_{i,t+1} & = & s_{it} + q_{it} - y_{it} \\ p_{it} & = & p_{i,t-1} + \Delta p_{it} \\ c_{i,t+1} & \sim & AR(1) \\ \omega_{i,t+1} & \sim & AR(1) \\ \varepsilon_{it} & \sim & i.i.d. \end{array}$$

(Integrated) Bellman Equation

• The components of ε_{it} are independently and extreme value distributed with dispersion parameter σ_ε.

• Therefore, as in Rust (1987), the integrated value function $\bar{V}(x_{it})$ is the unique fixed point of the integrated Bellman equation:

$$\bar{V}(x_{it}) = \sigma_\varepsilon \ln \left(\sum_{a \in A} \exp \left\{ \frac{v(a, x_{it})}{\sigma_\varepsilon} \right\} \right)$$

where:

$$v(a, x_{it}) = \bar{\pi}(a, x_{it}) + \beta \sum_{x_{i,t+1}} \bar{V}(x_{i,t+1}) \; f_x(x_{i,t+1} | a, x_{it})$$

Discrete choice profit function

• $\bar{\pi}(a, x_{it})$ is the part of current profit which does not depend on ε_{it}:

$$\bar{\pi}(a, x_{it}) = \begin{cases} R_{it}(0,0) - \alpha\ s_{it} & \text{if} \quad a = (0,0) \\ R_{it}(0,\delta_i^{(+)}) - \alpha\ s_{it} - F_{mc}^{(+)} & \text{if} \quad a = (0,\delta_i^{(+)}) \\ R_{it}(0,\delta_i^{(-)}) - \alpha\ s_{it} - F_{mc}^{(-)} & \text{if} \quad a = (0,\delta_i^{(-)}) \\ R_{it}(\kappa_i,0) - \alpha\ s_{it} - F_{oc} - c_{it}\kappa_i & \text{if} \quad a = (\kappa_i,0) \\ R_{it}(\kappa_i,\delta_i^{(+)}) - \alpha\ s_{it} - F_{oc} - c_{it}\kappa_i - F_{mc}^{(+)} & \text{if} \quad a = (\kappa_i,\delta_i^{(+)}) \\ R_{it}(\kappa_i,\delta_i^{(-)}) - \alpha\ s_{it} - F_{oc} - c_{it}\kappa_i - F_{mc}^{(-)} & \text{if} \quad a = (\kappa_i,\delta_i^{(-)}) \end{cases}$$

where $R_{it}(.,.)$ is the expected revenue function.

Some predictions of the model

• Fixed ordering cost F_{oc} generate infrequent orders: $(\mathbf{S},\mathbf{s})$ **inventory policy**.
• (S, s) inventory behavior, together demand uncertainty (i.e., optimal prices depend on the supply-to-expected demand ratio) generate a cyclical pattern in the price elasticity of sales.
• Prices decline significantly when an order is placed (sales promotion).
• This price decline and the consequently inventory reduction generate a price increase.
• Then, as inventories decline between two orders, prices tend to increase.

Data

• Data from the central warehouse of a supermarket chain in the Basque Country (Spain).
• Monthly data: period January 1990 to May 1992.

Data: Products

Descriptive Statistics

Reduced Form estimation of decision rules

Evolution of markup between two orders

Estimation of Structural Parameters

Counterfactual Experiments

5.3. Pesendorfer (2002).

5.4. Kano (2013).

CHAPTER 8

Structural Models of Dynamic Demand of Differentiated Products

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 the price change is. For instance, if a price reduction is perceived by consumers as very transitory (e.g., 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 estimated 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.

Pesendorfer (*Journal of Business*, 2002)

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 implies also a biased 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 (Econometrica, 2006) and Erdem, Keane and Imai (QME, 2003). These papers extend microeconometric discrete choice models of product differentiation to a dynamic setting, and contains useful methodological contributions. Their empirical results show that ignoring the dynamics of demand can lead to serious biases. Also the papers 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.

2. Data and descriptive evidence

We assume that 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, if a store was visited then which one, 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 set covers the period June 1991 to June 1993. Purchases and price information is available in real (continuous) time but for the analysis in the paper it is aggregated at 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 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 (e.g., laundry detergent, milk, etc). The paper 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 (RAND, 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 (e.g., house size) are negatively correlated with households' propensity to buy on sale.

3. Model

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. I explain below the computational advantages of this assumption. Of course, the assumption imposes some restrictions on the intertemporal substitution between brands, and I will discuss this point too. Erdem, Imai, and 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 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} is 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:

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

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

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

$j \in \{1, 2, ..., J\}$ is the brand index. $x \in \{0, 1, 2, ..., X\}$ 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, i.e., for some brands and weeks p_{jxt} and $x * p_{j1t}$ can take different values. This is potentially important because the price data shows 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, e.g., 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. That is the approach in Hendel and Nevo's paper.

Finally, ε_{hjxt} is a consumer idiosyncratic shock that is indepenendetly 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} vary over individuals, ξ_{jxt} do 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 his 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 his consumption of the storable product, c_{ht}, and how much to purchase and which product, $d_{ht} = \{d_{hjxt}\}$. The household makes this decision to maximize his expected and discounted stream of current and future utilities,

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

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 processes. 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:

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

where the expectation $E\left(.\mid\mathbf{s}_{ht},\ c_{ht},\ d_{ht}\right)$ is 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: $c_{ht}=c_h^*\left(\mathbf{s}_{ht}\right)$ and $d_{ht}=d_h^*\left(\mathbf{s}_{ht}\right)$ where $c_h^*\left(.\right)$ 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 $c_h^*\left(.\right)$ 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 $c_{ht}=c_h^*\left(\mathbf{s}_{ht}\right)$ and $d_{ht}=d_h^*\left(\mathbf{s}_{ht}\right)$ and our data on (some) decision 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 of 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, Imai, and Keane (2003).

And 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 of a dynamic demand model. However, dealing with this problem may not be independent of the other issues mentioned above.

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 there 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 that there is only one inventory, the aggregate inventory of all the products, and not one inventory for each brand, $\{i_{hjt}\}$, has already reduced 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 its 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 its 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 captures 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:

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

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 expression that describes the optimal brand choice, $j_{ht} = \arg\max_{j\in\{1,2,...,J\}}$ $\{\beta_h\ a_{jxt} - \alpha_h\ p_{jxt} + \xi_{jxt} + \varepsilon_{hjxt}\}$ 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 $\boldsymbol{\omega}_h(\mathbf{p}_t)$ be the vector with the inclusive values for every possible size $\{\omega_h(x, \mathbf{p}_t) : x = 1, 2, ..., X\}$.

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

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

In words, the vector $\boldsymbol{\omega}_h(\mathbf{p}_t)$ contains all the relevant information in $\mathbf{p}_t$ to obtain the probability distribution of $\boldsymbol{\omega}_h(\mathbf{p}_{t+1})$ 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 $\boldsymbol{\omega}_h(\mathbf{p}_t)$.

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 $\boldsymbol{\omega}_h(\mathbf{p}_t)$ 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 $\bar{V}_h(i_{ht}, \boldsymbol{\omega}_{ht})$ that includes only $X + 1$ variables, instead of $3 * J * X + 1$ state variables.

4. Estimation

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. First, 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) facilities 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, e.g., 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. And 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:

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

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, to 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 $\sigma_{jxt} \equiv a_{jxt}\ \sigma_\beta - p_{jxt}\ \sigma_\alpha + \xi_{jxt}\sigma_\lambda$.

4.1.1. *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 he 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 in sample 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, i.e., $H(x, t) = \sum_{h=1}^{H} 1\{x_{ht} = x\}$. 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, i.e., $H(t) = \sum_{h=1}^{H} 1\{x_{ht} > 0\}$. We need that $H(t) > J(K+1)$ which is a much weaker condition.

Given estimates of the incidental parameters, $\{\hat{\delta}_{jxt}, \hat{\sigma}_{jxt}\}$, now we can 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", i.e., 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:

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

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 get also consistent estimates of the marginal utilities β_h and α_h as:

$$\hat{\beta}_h = \hat{\beta}_0 + (z_h - \bar{z})\hat{\sigma}_\beta$$

$$\hat{\alpha}_h = \hat{\alpha}_0 + (z_h - \bar{z})\hat{\sigma}_\alpha$$

Finally, we can get estimates of the inclusive values:

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

4.1.2. *Control function approach.* The previous approach, though simple, has the limitation that we need to have, for every week in the sample, a large enough number of households making positive purchases. That requirement is not needed for identification of the parameters. It is only needed 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 other 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 (Journal of Econometrics, 1988), Blundell and Powell (REStud, 2004), and in the specific context of demand of differentiated products, Petrin and Train (Journal of Marketing Research, 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:

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

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 $J * X$, that is a pretty 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,

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

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 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:

$$g_{jx}^{\xi}(e_t) \equiv E(\xi_{jxt} \mid e_t)$$

Then, we can write ξ_{jxt} as the sum of two components, $\xi_{jxt} = g_{jx}^{\xi}(e_t) + v_{jxt}$. By construction, the error term v_{jxt} is mean independent of e_t. But then, 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}). Then, we can write the utility of product (j, x) as:

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

The term $g_{jx}^{\xi}(e_t)$ 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:

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

where the control functions $\{g_{jx}^{\xi}(e_t)\}$ consists of a brand dummies and 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}.

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 of demand of a storable good. Also, this is not a "standard" unobservable variable in the sense that it follows a stochastic process that is endogenous. That is, not only inventories affect purchasing decision, but also purchasing decisions affect the evolution of inventories.

The approach used by Erdem, Imai, and Keane (2003) to deal with this problem is to assume that household inventories is a (deterministic) function of "number of weeks (duration) since last purchase", T_{ht}, and the quantity purchased in the last purchase, x_{ht}^{last}:

$$i_{ht} = f_h(x_{ht}^{last}, T_{ht})$$

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,

$$i_{ht+1} = \max \left\{ \ 0 \ ; \ x_{ht}^{last} - c_h \ T_{ht} \right\}$$

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

$$i_{ht+1} = \max \left\{ \ 0 \ ; \ (1 - \lambda_h)^{T_{ht}} \ x_{ht}^{last} \right\}$$

Using this approach, the state variable i_{ht} should be replaced by the state variables (x_{ht}^{last}, 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 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): $\Pr(x_{ht} | x_{ht}^{last}, T_{ht}, \mathbf{p}_t)$. Suppose that we see that this probability is not a function of past behavior (x_{ht}^{last}, 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.

4.2.1. *Maximum Likelihood estimation (with proxies for inventories).* To Be Completed

4.2.2. *Hotz-Miller estimation (with proxies for inventories.* To Be Completed

4.2.3. *Maximum Likelihood estimation (without proxies for inventories).* To Be Completed

5. Empirical Results

To Be Completed

6. Dynamic Demand of Differentiated Durable Products

- Gowrisankaran and Rysman (2009)

TBW