Dynamic Discrete Choice Structural Models in Empirical IO

Lecture 1: Introduction

Victor Aguirregabiria (University of Toronto)

Carlos III, Madrid

June 26, 2017

Lecture 1: Introduction to Dynamic Discrete Choice Models in Empirical IO

- 1. Some Basic Ideas in IO
- 2. Dynamic Models in Empirical IO: Examples
- 3. Outline of this series of lectures
- 4. Introduction to DDC structural models
- 5. Some applications

Some Basic Ideas in IO (1)

- IO studies the behavior of firms in markets, their strategic interactions, and the implications on profits and consumer welfare.
- Some examples of type of firm decisions that we study in IO are:
  - Price and Quantity choice;
  - Investment in capacity, inventories, physical capital, ...;
  - R&D, patents;
  - Advertising;
  - Geographic location of plants and stores;
  - Product design;
  - Entry in new markets;
  - Adoption of new technologies;
  - Vertical relationships;

"New" Empirical IO

- Emphasizes the need to:
- **[1]** **Study competition separately for each industry**. Industries are very heterogeneous in their exogenous characteristics. There is not a common relationship between market power and concentration across industries.
- **[2]** Use **micro-level data** of individual firms, products, and markets, on prices, quantities, number of firms, and exogenous characteristics affecting demand or costs.
- **[3]** Estimate **structural models** of consumer and firm behavior.

Specification of a Structural Model in EIO (1)

- To study competition in an industry, EIO researchers propose and estimate **structural models of demand and supply**.
- **What is an structural model in empirical IO?**
- Models of consumer and firm behavior where consumers are utility maximizers and firms are profit maximizers.
- The **parameters are structural** in the sense that they describe **consumer preferences, production technology, and institutional constraints**.
- Under the principle of **revealed preference**, these parameters are estimated using micro data on consumers' and firms' choices and outcomes.

Specification: Typical Structure of IO Models

1. **Model of consumer behavior (Demand)**

   - Product differentiation?

2. **Model for firms' costs**

   - Economies of scale; Economies of scope? Entry costs? Investment costs?

3. **Equilibrium model of static competition**
   - Price (Bertrand), Quantity (Cournot).

4. **Equilibrium model of market Entry-Exit and dynamic competition**

- Investment, advertising, quality, product characteristics, stores, etc.

Specification: Example

- Example based on Ryan (Econometrica, 2012).
- **We start with an empirical question.**
- US cement industry. Evaluation of the effects in this industry of the 1990 Amendments to the Air Clean Act.
- The new law restricts the amount of emissions a cement plant can make.
- It requires the adoption of a "new" technology that implies lower marginal costs but larger fixed costs than the "old" technology.

Specification: Key Characteristics of the Industry

- The model here, though simple, incorporates some important features of the cement industry.

1. Homogeneous product. (We abstract from spatial differentiation).
2. Substantial fixed costs from operating a plant (cement furnace).
3. Variable costs increase in a convex way when output approaches full capacity.
4. Capacity investment is an important strategic variable.
5. Industry is very local (due to high transportation costs per dollar value). It can be characterized as a set of many "isolated" local markets.
6. Oligopolist industry. Small number of firms at a local market.

Dynamic Models in Empirical IO: Examples

- Dynamics in demand and/or supply are important aspects of competition in oligopoly markets.
- **Dynamics in demand:** consumer switching costs; habit formation; brand loyalty; learning; and storable or durable products.
- **Dynamics in supply:** almost every firm investment decision: market entry; investment in capacity, inventories, or equipment; choice of product characteristics; production if there is learning by doing; pricing if there are menu costs, or other forms of price adjustment costs.

Example 1: Demand of a storable good (1)

- For a storable product, purchases in a given period (week, month) are not equal to consumption.
- When the price is low consumers buy for storage and future consumption. When the price is high they do not purchase and consume from inventory.
- Dynamics arise because:
  (a) consumers' past purchases impact their current inventory and the benefits of purchasing today;

  (b) consumers' expectations about future prices impact the trade-offs of buying today versus in the future.

Example 1: Demand of a storable good (2)

- **What are the implications of ignoring these dynamics in a model of consumer demand?**

- Bias estimates of both long-run and short-run demand elasticities

- Bias estimates of firms' market power. Implications on merger analysis, anti-trust cases, etc.

Example 1: Demand of a storable good (3)

- The time series of prices of many supermarket products 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:** the typical discount of a sales promotion is between 10% and 20%, and the increase in sales can be 200% or even larger.

Example 1: Demand of a storable good (4)

- The estimation of a static demand model provides a large estimates of the own-price elasticity, e.g., > 8.
- Given this estimated demand elasticity, we may conclude that brand manufacturers have very low market power.
- However, the static model can be seriously wrong because it ignores that a substantial part of the temporary increase in sales comes from **consumer intertemporal substitution**.
- 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.

Example 2: Demand of a new durable good (1)

- The price of new durable products typically declines over time during the months after the introduction of the product.
- 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 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.

Example 2: Demand of a new durable good (2)

- A static model that ignores dynamics in demand of durable goods introduces two different type of biases:
- **[1] Endogenous selection in the number of potential buyers:** each period the demand curve is changing because some high willingness-to-pay consumers have already bought the product and left the market.
- **[2] Consumer forward-looking behavior:** consumers willingness to pay is downward biased because it is contaminated by the expectation of future price declines.

Example 2: Demand of a new durable good (3)

- To illustrate the first source of bias, consider a simple example. Market with an initial mass of H_1 consumers and a uniform distribution of willingness to pay over $[0,\$100]$.

- Demand at time $t \in \{1, 2, ...\}$ is:

$$Q_t = H_t \Pr(v_t \geq P_t) = H_t \left[1 - F_t \left(P_t\right)\right]$$

$H_t =$ Consumers still in the market at period t; i.e., $H_t = H_{t-1} - Q_{t-1}$.
$F_t =$ Distribution function of willingness to pay for consumers who remain in the market at period t.

Example 2: Demand of a new durable good (4)

- Suppose that the sequence of prices is $P_1 = \$90$, $P_2 = \$80$, $P_3 = \$70$, etc. Then,

$$
\begin{aligned}
Q_1 &= H_1 \left[1 - \frac{P_1}{100} \right] = 0.1\ H_1 \text{; and } H_2 = 0.9\ H_1 \\
Q_2 &= 0.9\ H_1 \left[1 - \frac{P_2}{90} \right] = 0.1\ H_1 \text{; and } H_3 = 0.8\ H_1 \\
Q_3 &= 0.8\ H_1 \left[1 - \frac{P_3}{80} \right] = 0.1\ H_1 \text{; and } H_4 = 0.7\ H_1
\end{aligned}
$$

- The sequence of quantities Q_1, Q_2, Q_3, ... is constant over time.
- A static demand model concludes that consumers are not sensitive to price, since price is declining and demand is constant. The estimate of the price elasticity would be zero.

Example 3: Dynamics of market structure (1)

- Consider a technological change (or a public policy) that reduces firms' Marginal Costs but increases Fixed Costs.
- Short-run and long-run effects of this change can be very different.
- To measure long-run effects we need to take into account that market structure is endogenous but its response is not instantaneous and adjust slowly over time.

Example 3: Dynamics of market structure (2)

- In the short-run, the number of firms in the market does not change. The reduction in MCs implies a reduction in prices, and increases in output and consumer surplus.
- Over time, the greater fixed cost may imply a reduction of the number of firms in the market (exits $>$ entries).
- The reduction in the number of firms over time implies a reduction in competition, and consequently that prices will increase over time, and output and consumer surplus will decline.
- Long-run effects may have the opposite sign than short-run effects.

Outline of this series of lectures

- Lecture 1: Introduction to the Econometrics of DDC structural models [single-agent]
- Lecture 2: Empirical dynamic games of oligopoly competition
- Lecture 3: Dynamic games when players have out-of-equilibrium beliefs.
- Lecture 4: Curse of dimensionality: Euler equations for the solution and estimation of DDC structural models
- Lecture 5: Unobserved heterogeneity: Consistent Fixed Effect Estimation of DDC structural models

Dynamic Discrete Choice Structural Models

• In dynamic structural models, agents are forward looking and maximize expected intertemporal payoffs.

• The parameters to be estimated are *structural* in the sense that they describe agents' preferences and technological and institutional constraints.

• Under the principle of **revealed preference**, these parameters are estimated using micro data on individuals' choices and outcomes.

Dynamic Discrete Choice Structural Models

• Econometric models in this class can be useful tools for the evaluation of new (**counterfactual**) policies in settings with important dynamic aspects.

• Another attractive feature is that structural parameters have a transparent interpretation within the theoretical model that frames the empirical investigation.

• Seminal papers in this literature include:

** Wolpin (JPE, 1984) on fertility and child mortality*
** Miller (JPE, 1984) on occupational choice*
** Pakes (Econometrica, 1986) on patent renewal*
** Rust (Econometrica, 1987) on machine replacement*

Empirical applications in IO:

- Demand models: Consumer switching costs; brand loyalty; storable products; durable products; adoption of a new product.
- Price competition when demand is dynamic.
- Inventories, capacity, capital investment.
- R&D investment; Innovation; Adoption of new technologies.
- Market entry-exit; Quality choice; Product design.

Dynamic Discrete Choice Structural Models

- Time is discrete: $t = 0, 1, ..., \infty$. Every period t an agent observes a vector of state variables s_t and makes a choice:

$$a_t \in A = \{0, 1, ..., J\}$$

- The agent maximizes expected intertemporal payoff:

$$\mathbb{E}_t \left[\sum_{j=0}^{T-t} \beta^j \; U_{t+j}(a_{t+j}, s_{t+j}) \right]$$

U_t is the one-period utility and $\beta \in (0, 1)$ is the discount factor.

- The agent knows s_t but has uncertainty about future state variables s_{t+1}, s_{t+2}, ... She has beliefs about uncertain future state variables that can be represented as a Markov transition probability

$$p_t(s_{t+1} | s_t, a_t)$$

Dynamic Discrete Choice Structural Models

- Agents are expected utility maximizers. An agent optimal decision rule is:

$$\alpha_t(s_t) = \arg\max_{a_t \in A} \; E\left[\sum_{j=0}^{T-t} \beta^j \; U_{t+j}(a_{t+j}, s_{t+j}) \mid s_t, a_t \right]$$

- Using Bellman's principle of optimality, we can represent this optimal decision rule as:

$$\alpha_t(s_t) = \arg\max_{a_t \in A} \; \left[U_t(a_t, s_t) + \beta \int V_t(s_{t+1}) \; p_t(s_{t+1} | s_t, a_t) \right]$$

- Where:

$$V_t(s_t) = \max_{a_t \in A} \left[U_t(a_t, s_t) + \beta \int V_{t+1}(s_{t+1}) \; p_t(s_{t+1} | s_t, a_t) \right]$$

Infinite Horizon - Stationary Case

- When $T = \infty$, $U_t(.) = U(.)$, and $p_t(.) = p(.)$, Blackwell Theorem establishes that the value function and the optimal decision rule are invariant over time.

- At two time periods with the same state the decision problem is identically the same.

$$\alpha(s_t) = \arg\max_{a_t \in A} \left[U(a_t, s_t) + \beta \int V(s_{t+1}) \; p(s_{t+1}|s_t, a_t) \right]$$

- and:

$$V(s_t) = \max_{a_t \in A} \left[U(a_t, s_t) + \beta \int V(s_{t+1}) \; p(s_{t+1}|s_t, a_t) \right]$$

Data and Unobservables

• From the point of view of the econometrician we distinguish two subsets of state variables:

$$s_t = (x_t, \varepsilon_{it})$$

• The subvector x_t groups variables that are observed by both the agent and the researcher.

• The subvector ε_t is observed only by the agent.

Data and Unobservables

• The researcher observes a sample of individuals over several periods of time:

$$Data = \{ \ a_{it} \ , \ x_{it} : i = 1, 2, ..., N \ ; \ t = 1, 2, ..., T_i \}$$

• i is the individual subindex; N is the number of individuals in the sample; and T_i is the number of periods over which we observe individual i.

• In micro-econometric applications of single-agent models, we typically have that N is relatively large and T_i is small.

Example: Market Entry-Exit

- $a_t \in \{0, 1\}$ is the indicator of the firm being in the market.
- $s_t = (a_{t-1}, z_t)$, where z_t is a vector of exogenous variables affecting demand and costs.
- The profit of being active in the market is:

$$U(1, s_t) = vp(z_t) - fc(z_t) - 1\{a_{t-1} = 0\}\ \ ec(z_t)$$

- $vp(z_t) = (p(z_t) - c(z_t))\ q(z_t)$ is variable profit; $fc(z_t)$ is fixed cost; and $ec(z_t)$ is an entry cost.
- The profit of being in-active is:

$$U(0, s_t) = a_{t-1}\ sv(z_t)$$

where $sv(z_t)$ is the scrap value of the firm.

Example: Market Entry-Exit (2)

- The Optimal decision rule is:

$$
\begin{aligned}
\{a_t &= 1\} \Leftrightarrow \\
& U(1, s_t) + \beta \int V(1, z_{t+1}) \; p_z(z_{t+1}|z_t) \\
\geq \; & U(0, s_t) + \beta \int V(0, z_{t+1}) \; p_z(z_{t+1}|z_t)
\end{aligned}
$$

- And the Bellman equation is:

$$
V(a_{t-1}, z_t) = \max_{a_t \in \{0,1\}} \left[U(a_t, a_{t-1}, z_t) + \beta \int V(a_t, z_{t+1}) \; p_z(z_{t+1}|z_t) \right]
$$

Estimation

- Given a dataset $\{a_{it}, x_{it}\}$, we are interested in estimating the unknown parameters in the primitives $\{U, p, \beta\}$.
- Let $\boldsymbol{\theta}$ be the vector of structural parameters. We distinguish three components in this vector:

$$\boldsymbol{\theta} = \{\ \boldsymbol{\theta}_u,\ \boldsymbol{\theta}_f,\ \boldsymbol{\theta}_\varepsilon, \beta\ \}$$

$\boldsymbol{\theta}_u =$ *parameters in utility function U*
$\boldsymbol{\theta}_f =$ *parameters in transition probability of observables*
$\boldsymbol{\theta}_\varepsilon =$ *parameters in distribution of observables*

Estimation

- Let $g_N(\boldsymbol{\theta})$ be an estimation criterion for this model and data, such as a likelihood or a GMM criterion.

- For instance, if the data are a random sample over individuals and the criterion is a likelihood, then $g_N(\boldsymbol{\theta}) = \sum_{i=1}^{N} l_i(\boldsymbol{\theta})$, where

$$\begin{aligned} l_i(\boldsymbol{\theta}) &= \log \Pr\left(a_{i1}, \ldots, a_{iT_i}, x_{i1}, \ldots, x_{iT_i} \mid \boldsymbol{\theta}\right) \\ &= \log \Pr\left(\begin{array}{c} \alpha(x_{i1}, \varepsilon_{i1}, \boldsymbol{\theta}) = a_{i1}, \ldots, \alpha(x_{iT_i}, \varepsilon_{iT_i}, \boldsymbol{\theta}) = a_{iT_i}, \\ x_{i1}, \ldots, x_{iT_i} \mid \boldsymbol{\theta} \end{array}\right) \end{aligned}$$

- To evaluate $g_N(\boldsymbol{\theta})$ for a particular value of $\boldsymbol{\theta}$ it is necessary to know the optimal decision rules $\alpha(s_{it}, \boldsymbol{\theta})$. Therefore, for each trial value of $\boldsymbol{\theta}$ the DP problem needs to be solved exactly, or its solution approximated in some way.

Two important issues

- Applications of DDCSM have two deal with econometric issues which are common in other models in empirical micro, e.g., measurement error, endogeneity, sample selection, etc.
- However, there are two econometric issues that are particularly important in DDCSM:

Curse of dimensionality: *Computational cost of computing exactly the solution of the DP problem increases very quickly with the dimension of the state space.*

Serially correlated unobservables *(e.g., time-invariant unobserved heterogeneity).*

Assumptions on unobservables

- The relationship between observable and unobservable state variables, and the stochastic process of the latter, are key modelling decisions in the econometrics of DDSM.
- **Additive separability** (AS) and **serial independence of the unobservables** (CI) provide the simplest framework for estimation and has been used in many applications.

Assumptions on unobservables

- *ASSUMPTION AS:*

$$U(a, x_{it}, \varepsilon_{it}) = u(a, x_{it}) + \varepsilon_{it}(a)$$

where $\varepsilon_{it}(a)$ is the a-th component the vector ε_{it}.

- *ASSUMPTION CI:*

$$p(x_{t+1}, \varepsilon_{t+1} | a_t, x_t, \varepsilon_t) = f_\varepsilon(\varepsilon_{t+1} | x_{t+1})\, f_x(x_{t+1} | a_t, x_t)$$

Conditional Choice Probabilities

• Define the *Conditional Choice Probability* (CCP) function $P(a_{it}|x_{it}, \boldsymbol{\theta})$ as the integration of the optimal decision rule over the distribution of the unobservable state variables.

$$P(a|x, \boldsymbol{\theta}) \equiv \int 1\left\{\alpha(x, \varepsilon; \boldsymbol{\theta}) = a\right\}\ f_{\varepsilon}(\varepsilon)\ d\varepsilon$$

where $1\{.\}$ is the indicator function.

• An important implication of the CI Assumption is that the CCP function is equal to the distribution of a_{it} conditional on x_{it}:

$$\Pr\left(a_{it} = a \mid x_{it} = x\right) = P(a|x, \boldsymbol{\theta})$$

• In general, this condition does not hold if ε_{it} is serially correlated because ε_{it} and x_{it} are not independent.

Assumptions on unobservables

• Under Assumption CI, the contribution of individual i to the log-likelihood function can be factored as:

$$\begin{aligned} l_i(\boldsymbol{\theta}) &= \sum_{t=1}^{T_i} \log P(a_{it}|x_{it}, \boldsymbol{\theta}) \\ &+ \sum_{t=1}^{T_i-1} \log f_x(x_{i,t+1}|x_{it}, \boldsymbol{\theta}_f) \\ &+ \log \Pr(x_{i1}|\boldsymbol{\theta}) \end{aligned}$$

• The term $\log \Pr(x_{i1}|\boldsymbol{\theta})$ is the contribution of the initial conditions to the likelihood of individual i.

Assumptions on unobservables

• Another implication of AS-CI is that the alternative-specific value functions can be decomposed as $v(a, x_{it}) + \varepsilon_{it}(a)$ as in random utility models.

• CCPs are functions of the value differences $\tilde{v}(j, x_{it}) \equiv v(j, x_{it}) - v(0, x_{it})$:

$$P(a|x_{it}, \boldsymbol{\theta}) = \Lambda(a|\{\tilde{v}(j, x_{it}, \boldsymbol{\theta}) : j \in A\})$$

where:

$$\Lambda(a|\{\tilde{v}(j, x_{it}, \boldsymbol{\theta}) : j \in A\}) \equiv$$

$$\int 1\{v(a, x_{it}, \boldsymbol{\theta}) + \varepsilon_{it}(a) > v(a', x_{it}, \boldsymbol{\theta}) + \varepsilon_{it}(a') \text{ for all } a'\} \;\; f_{\varepsilon}(\varepsilon_{it}) \; d\varepsilon_{it}$$

• When $\{\varepsilon_{it}(a)\}$ are independently distributed type 1 extreme value random variables:

$$\Lambda(a|\tilde{v}(., x_{it}, \boldsymbol{\theta})) = \frac{\exp\{\tilde{v}(a, x_{it}, \boldsymbol{\theta})\}}{1 + \sum_{j=1}^{J} \exp\{\tilde{v}(j, x_{it}, \boldsymbol{\theta})\}}$$

Time-invariant unobserved heterogeneity: Finite mixture models

• Suppose that the unobservables have a transitory and permanent: $\varepsilon_{it}(a)$ and $\omega_i(a)$, where are $\omega_i' s$ individual-specific permanent effects with a discrete distribution.

• Now, Assumption CI fails and the probability of the sequence of choices cannot be factored into a product of conditional choice probabilities.

• The observable state x_{it} is not a sufficient statistic for a_{it} because lagged choices contain information about the permanent components ω_i. However, conditional on ω_i the transitory components $\{\varepsilon_i(a)\}$ do satisfy assumption CI.

Finite mixture models

• Since ω_i has discrete support, each individual's conditional likelihood contribution can be obtained as a finite mixture of conditional likelihoods. Let $\pi_\ell \equiv \Pr(\omega_i = \omega^\ell)$, then:

$$l_i(\boldsymbol{\theta}) = \log\left(\sum_{\ell=1}^{L} \mathcal{L}_i(\boldsymbol{\theta}, \omega^\ell)\ \pi_\ell\right)$$

where

$$\begin{aligned}
\mathcal{L}_i(\boldsymbol{\theta}, \omega^\ell) &= \Pr\left(a_{i1}, \ldots, a_{iT_i}, x_{i1}, \ldots, x_{iT_i} \mid \omega^\ell, \boldsymbol{\theta}\right) \\
&= \left[\prod_{t=1}^{T_i} P(a_{it}|x_{it}, \omega^\ell, \boldsymbol{\theta})\ \prod_{t=1}^{T_i-1} f_x(x_{i,t+1}|x_{it}, \boldsymbol{\theta}_f)\right] \\
&\quad \Pr(x_{i1}|\omega^\ell, \boldsymbol{\theta})
\end{aligned}$$

• In order to evaluate the mixture of likelihoods the DP problem needs to be solved as many times as the number of components in the mixture.

Nested fixed point (NFXP) algorithm

• 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 of structural parameters, say $\hat{\boldsymbol{\theta}}_0$. A BHHH iteration is defined as:

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

where $\nabla l_i(\boldsymbol{\theta})$ is the gradient in $\boldsymbol{\theta}$ of the log-likelihood function for individual i.

Nested fixed point (NFXP) algorithm

• $\nabla l_i(\boldsymbol{\theta})$ is the sum of two terms: the gradient of the choice history and the gradient of the transitions.

$$\begin{aligned}\nabla l_i(\boldsymbol{\theta}) &= \textstyle\sum_{t=1}^{T_i} \nabla \log \Lambda(a_{it}|\tilde{v}(., x_{it}, \boldsymbol{\theta})) \\ &\quad + \textstyle\sum_{t=1}^{T_i-1} \nabla \log f_x(x_{i,t+1}|a_{it}, x_{it}, \boldsymbol{\theta}_f)\end{aligned}$$

• The second term is standard because the transition probability function is a primitive of the model. However, to obtain the first term we have to solve the DP problem for a value $\hat{\boldsymbol{\theta}}_k$ of the structural parameters.

Nested fixed point (NFXP) algorithm

• There are different ways to solve the DP problem. When the model has finite horizon the standard approach is to use backward iterations. For infinite horizon models, one can use either value function or policy function iterations, or an hybrid of both.

• To illustrate this algorithm in more detail, consider a version of Rust model which has been the most common in applications: a conditional logit model where one-period utilities are linear in the parameters $\boldsymbol{\theta}_u$:

$$u(a, x_{it}, \boldsymbol{\theta}_u) = z(a, x_{it})' \boldsymbol{\theta}_u,$$

where $z(a, x_{it})$ is a vector of known functions.

Nested fixed point (NFXP) algorithm

- For this model, the gradient $\nabla l_i(\boldsymbol{\theta})$ has the following form:

$$\nabla l_i(\boldsymbol{\theta}) \quad = \quad \sum_{t=1}^{T_i} \left[z(a_{it}, x_{it}) - \left(\sum_{j=0}^{J} P(j|x_{it}, \boldsymbol{\theta}) \; z(j, x_{it}) \right) \right]$$

$$+ \sum_{t=1}^{T_i - 1} \nabla \log f_x(x_{i,t+1} | a_{it}, x_{it}, \boldsymbol{\theta}_f)$$

- The choice probabilities $P(a|x, \boldsymbol{\theta})$ have the conditional logit form:

$$P(a|x, \boldsymbol{\theta}) = \frac{\exp\left\{ z(a, x)\boldsymbol{\theta}_u + \beta \; \mathbf{F}_x(a, x)' \mathbf{V}_\varepsilon(\boldsymbol{\theta}) \right\}}{\sum_{j=0}^{J} \exp\left\{ z(j, x)\boldsymbol{\theta}_u + \beta \; \mathbf{F}_x(j, x)' \mathbf{V}_\varepsilon(\boldsymbol{\theta}) \right\}}$$

where $\mathbf{V}_\varepsilon(\boldsymbol{\theta})$ and $\mathbf{F}_x(a, x)$ are the column vectors $\{V_\varepsilon(x, \boldsymbol{\theta}) : x \in X\}$ and $\{f_x(x'|a, x) : x' \in X\}$, respectively.

Nested fixed point (NFXP) algorithm

• The vector of values $\mathbf{V}_{\varepsilon}(\boldsymbol{\theta})$ can be obtained as the unique fixed point of the following Bellman equation in vector form:

$$\mathbf{V}_{\varepsilon}(\boldsymbol{\theta}) = \log\left(\sum_{a=0}^{J} \exp\left\{\mathbf{z}(a)\boldsymbol{\theta}_{u} + \beta\ \mathbf{F}_{x}(a)\mathbf{V}_{\varepsilon}(\boldsymbol{\theta})\right\}\right)$$

with $\mathbf{z}(a)$ and $\mathbf{F}_{x}(a)$ are the matrices $\{z(a,x) : x \in X\}$ and $\{\mathbf{F}_{x}(a,x) : x \in X\}$, respectively.

Nested fixed point (NFXP) algorithm

- The NFXP algorithm works as follows.

  (I) [Inner Algorithm] *Given* $\hat{\theta}_k$*, we obtain the vector* $\mathbf{V}_\varepsilon(\hat{\theta}_k)$ *by successive iterations in the Bellman equation: starting with* $\mathbf{V}_0 = \mathbf{0}$*, we iterate until convergence in*

  $$\mathbf{V}_{h+1} = \log(\textstyle\sum_{a=0}^{J} \exp\{\mathbf{z}(a)\hat{\theta}_{u,k} + \beta\hat{\mathbf{F}}_x(a)\mathbf{V}_h\})$$

  (II) *Then, given* $\hat{\theta}_k$ *and* $\mathbf{V}_\varepsilon(\hat{\theta}_k)$ *we construct the choice probabilities* $P(a|x, \hat{\theta}_k)$ *and the gradient* $\nabla l_i(\hat{\theta}_k)$ *using the expression above.*
  (III) [Outer iteration] *We we use the gradient* $\nabla l_i(\hat{\theta}_k)$ *to make a new BHHH iteration to obtain* $\hat{\theta}_{k+1}$.
  ** We proceed in this way until the distance between* $\hat{\theta}_{k+1}$ *and* $\hat{\theta}_k$ *is smaller than a pre-specified convergence constant.*

Nested fixed point (NFXP) algorithm

• When the model has finite horizon, we can sequentially solve for the value function using backward iterations in the inner algorithm of the NFXP. That is, the sequence of value vectors at ages T, $T-1$, etc, can be obtained starting with:

$$\mathbf{V}_T(\hat{\boldsymbol{\theta}}_k) = \log(\textstyle\sum_{a=0}^{J} \exp\{\mathbf{z}_T(a)\hat{\boldsymbol{\theta}}_{u,k}\})$$

and then using the sequential formula for $t \leq T-1$

$$\mathbf{V}_t(\hat{\boldsymbol{\theta}}_k) = \log(\textstyle\sum_{a=0}^{J} \exp\{\mathbf{z}_t(a)\hat{\boldsymbol{\theta}}_{u,k} + \beta\hat{\mathbf{F}}_{x,t}(a)\mathbf{V}_{t+1}(\hat{\boldsymbol{\theta}}_k)\})$$

Nested fixed point (NFXP) algorithm

• We have described the NFXP algorithm is in the context of full MLE. However, most applications of this algorithm have considered partial MLE. In this partial MLE approach, the parameters $\boldsymbol{\theta}_f$ in the transition probabilities are estimated by maximizing the partial likelihood $\sum_{i=1}^{N} \sum_{t=1}^{T_i-1} \log f_x(x_{i,t+1}|a_{it}, x_{it}, \boldsymbol{\theta}_f)$. This likelihood is very standard and does not require one to solve the DP problem. Given this estimator of $\boldsymbol{\theta}_f$, then the parameters in the utility function, $\boldsymbol{\theta}_u$, are estimated using the NFXP algorithm applied to the partial likelihood:

$$\sum_{i=1}^{N} \sum_{t=1}^{T_i} \triangledown \log \Lambda(a_{it}|\tilde{v}(., x_{it}, \boldsymbol{\theta}_u, \hat{\boldsymbol{\theta}}_f))$$

• This two-step approach can simplify very much the estimation problem in models with many parameters in the transition probabilities. For instance, this partial likelihood approach was used by Rust and Phelan (1997) in the model that we have described in Example 1.

Nested fixed point (NFXP) algorithm

• 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.

Sequential (Hotz-Miller) estimators

• Given the cost of solving some DP problems, this characteristic of the algorithm limits the range of applications where it can be applied.

• Hotz and Miller (1993) observed that, under the assumptions of Rust model, it is not necessary to solve the DP problem, even once, in order to estimate the structural parameters.

• A key idea in their method is that, using nonparametric estimates of choice and transition probabilities, it is possible to obtain a closed-form representation of the value function $\mathbf{V}_{\varepsilon}(\boldsymbol{\theta})$ for values of $\boldsymbol{\theta}$ around the true vector of structural parameters.

Sequential (Hotz-Miller) estimators

- This closed-form expression is particularly simple (and useful for estimation) in models where the utility function is linear-in-parameters. For this reason, here we illustrate Hotz-Miller method for this subclass of Rust models.

- For any $\boldsymbol{\theta}$, the vector of values $\mathbf{V}_{\varepsilon}(\boldsymbol{\theta})$ can be written as

$$\mathbf{V}_{\varepsilon}(\boldsymbol{\theta}) = \mathbf{W}(\mathbf{P}(\boldsymbol{\theta}), \boldsymbol{\theta}_f) \begin{pmatrix} \boldsymbol{\theta}_u \\ 1 \end{pmatrix}$$

** $\mathbf{P}(\boldsymbol{\theta}) \equiv \{P(a|x, \boldsymbol{\theta}) : (a, x) \in A \times X\}$ is the vector of conditional choice probabilities, for every state and action, associated with the optimal decision rule given $\boldsymbol{\theta}$;*
** $\mathbf{W}(\mathbf{P}, \boldsymbol{\theta}_f) \equiv \{\mathbf{W}(x, \mathbf{P}, \boldsymbol{\theta}_f) : x \in X\}$ is a valuation operator that is defined for any arbitrary value of $(\mathbf{P}, \boldsymbol{\theta}_f)$.*

Sequential (Hotz-Miller) estimators

• Each row of $\mathbf{W}(\mathbf{P}, \boldsymbol{\theta}_f)$ is associated with a value of x and it collects the expected and discounted sum of current and future $z's$ and $\varepsilon's$ which may occur along all possible future histories originating from current state x.

• The expected future $z's$ and $\varepsilon's$ are calculated under the assumption that the individual behaves today and in the future according to the choice probabilities in $\mathbf{P}$. This valuation operator is defined as the unique solution of the following contraction mapping:

$$\mathbf{W}(\mathbf{P}, \boldsymbol{\theta}_f) \;\; = \;\; \sum_{a=0}^{J} \mathbf{P}(a) * ([\mathbf{z}(a), \mathbf{e}(a, \mathbf{P})] + \beta \; \mathbf{F}_x(a) \; \mathbf{W}(\mathbf{P}, \boldsymbol{\theta}_f))$$

* $\mathbf{P}(\boldsymbol{\theta}) \equiv \{P(a|x, \boldsymbol{\theta}) : (a, x) \in A \times X\}$ is the vector of conditional choice probabilities, for every state and action, associated with the optimal decision rule given $\boldsymbol{\theta}$;

** $\mathbf{P}(a)$ is the column vector of choice probabilities $\{P(a|x) : x \in X\}$;*

** $\mathbf{e}(a, \mathbf{P}) = \{e(a|x, \mathbf{P}) : x \in X\}$, where $e(a|x, \mathbf{P})$ is the expectation of $\varepsilon_{it}(a)$ conditional on $x_{it} = x$ and ...*

Sequential (Hotz-Miller) estimators

• This conditional expectation is a function of a, $\mathbf{P}$, and the distribution G_ε only. The particular functional form of $e(a|x, \mathbf{P})$ depends on the probability distribution G_ε. A well known case where $e(a|x, \mathbf{P})$ has a closed-form expression is when $\varepsilon's$ are independently and identically distributed with extreme value distribution. In that case, $e(a|x, \mathbf{P})$ is equal to the Euler's constant minus $\log\left(P(a|x)\right)$.

• Let $\boldsymbol{\theta}^0 = (\boldsymbol{\theta}_u^0, \boldsymbol{\theta}_f^0)$ be the true value of $\boldsymbol{\theta}$ in the population of individuals under study.

• Let $\mathbf{P}^0$ be the conditional choice probabilities in the population.

Sequential (Hotz-Miller) estimators

• If we knew $(\mathbf{P}^0, \boldsymbol{\theta}_f^0)$, we could obtain the conditional choice values differences $\tilde{v}(a, x, \boldsymbol{\theta}^0)$ as a linear function of $\boldsymbol{\theta}_u^0$.

$$\tilde{v}(a, x, \boldsymbol{\theta}^0) = \tilde{z}(a, x; \mathbf{P}^0, \boldsymbol{\theta}_f^0)\ \boldsymbol{\theta}_u^0 + \tilde{e}(a, x; \mathbf{P}^0, \boldsymbol{\theta}_f^0) \tag{1}$$

where:

$$\tilde{z}(a, x; \mathbf{P}^0, \boldsymbol{\theta}_f^0) \equiv z(a, x) - z(0, x)$$

$$+\beta \textstyle\sum_{x'} [f_x(x'|a, x, \boldsymbol{\theta}_f^0) - f_x(x'|0, x, \boldsymbol{\theta}_f^0)] \mathbf{W}_z(x', \mathbf{P}^0, \boldsymbol{\theta}_f^0)$$

and

$$\tilde{e}(a, x; \mathbf{P}^0, \boldsymbol{\theta}_f^0) \equiv$$

$$\beta \textstyle\sum_{x'} [f_x(x'|a, x, \boldsymbol{\theta}_f^0) - f_x(x'|0, x, \boldsymbol{\theta}_f^0)] \mathbf{W}_e(x', \mathbf{P}^0, \boldsymbol{\theta}_f^0)$$

with $\mathbf{W}_z(\mathbf{P}, \boldsymbol{\theta}_f)$ and $\mathbf{W}_e(\mathbf{P}, \boldsymbol{\theta}_f)$ being the columns of $\mathbf{W}(\mathbf{P}, \boldsymbol{\theta}_f)$ associated with $\mathbf{z}$ and with $\mathbf{e}$, respectively.

Sequential (Hotz-Miller) estimators

• Though, we do not $(\mathbf{P}^0, \boldsymbol{\theta}_f^0)$, we can estimate consistently without having to solve the DP problem.

** Consistent estimates of transition probabilities can be obtained using a (partial) MLE of $\boldsymbol{\theta}_f^0$ that maximizes the (partial) likelihood $\sum_{i=1}^{n} \sum_{t=1}^{T_i - 1} \log f_x(x_{i,t+1} | a_{it}, x_{it}, \boldsymbol{\theta}_f)$.*
** Conditional choice probabilities can be estimated using nonparametric regression methods (i.e., $P^0(a|x) = E(I\{a_{it} = a\} | x_{it} = x))$ such as a Nadaraya-Watson kernel estimator or a simple frequency estimator.*

• Let $\hat{\mathbf{P}}_0$ and $\hat{\boldsymbol{\theta}}_{f0}$ be the estimators of $\mathbf{P}^0$ and $\boldsymbol{\theta}_f^0$, respectively. Based on these estimates, Hotz and Miller propose the GMM estimator that solves in $\boldsymbol{\theta}_u$ the sample moment conditions:

$$\sum_{i=1}^{n} \sum_{t=1}^{T_i} W_{it} \left[a_{it} - \Lambda \left(a_{it} | \{ \tilde{z}(j, x_{it}; \hat{\mathbf{P}}_0) \boldsymbol{\theta}_u + \tilde{e}(j, x_{it}; \hat{\mathbf{P}}_0) \} \right) \right] = 0$$

Sequential (Hotz-Miller) estimators

• The main advantage of this estimator is its computational simplicity. Nonparametric estimation of choice probabilities is a simple task. The main computational cost comes from the construction of $\mathbf{W}(\hat{\mathbf{P}}_0, \hat{\boldsymbol{\theta}}_{f0})$ using the valuation operator. However, $\mathbf{W}(\hat{\mathbf{P}}_0, \hat{\boldsymbol{\theta}}_{f0})$ is calculated just once and it remains fixed in the search for the Hotz-Miller estimator. In contrast, the NFXP algorithm requires one to compute this valuation operator many times, i.e., several times for each trial value $\boldsymbol{\theta}$

• Previous conventional wisdom was that Hotz-Miller estimator achieved a significant computational gain at the expense of efficiency, both in finite samples and asymptotically. Thus, researchers had the choice between two extremes: a full solution NFXP-ML estimator with the attendant computational burden, or the much faster but less efficient Hotz-Miller estimator.

• Aguirregabiria and Mira (2002) showed that a pseudo maximum likelihood (instead of GMM) version of Hotz-Miller estimator is

Sequential (Hotz-Miller) estimators

• The pseudo maximum likelihood (PML) estimator is defined as the value of $\boldsymbol{\theta}_u$ that maximizes the pseudo likelihood function:

$$l^{PL}(\boldsymbol{\theta}_u, \hat{\mathbf{P}}_0, \hat{\boldsymbol{\theta}}_{f0}) =$$

$$\sum_{i=1}^{n} \sum_{t=1}^{T_i} \log \Lambda \left(a_{it} \middle| \{ \tilde{z}(j, x_{it}; \hat{\mathbf{P}}_0) \boldsymbol{\theta}_u + \tilde{e}(j, x_{it}; \hat{\mathbf{P}}_0) \} \right)$$

• The asymptotic variance of this two-step PML estimator is just equal to the variance of the partial MLE.

• The initial nonparametric estimator of $\mathbf{P}^0$ and the PML estimator of $\boldsymbol{\theta}_u^0$ are asymptotically independent (i.e., $E(\partial^2 l_i^{PL} / \partial \boldsymbol{\theta}_u \partial \mathbf{P}') = 0$) and therefore there is not any asymptotic efficiency loss from using an inefficient initial estimator $\mathbf{P}^0$.

Sequential (Hotz-Miller) estimators

• Though the two-step PML estimator is asymptotically equivalent to partial MLE, the Monte Carlo experiments in Aguirregabiria and Mira (2002) show that the finite sample bias of the Hotz-Miller estimator can be much larger than the one of the MLE. Imprecise initial estimates of choice probabilities do not affect the asymptotic properties of the estimator, but they can generate serious small sample biases in Hotz-Miller estimator.

• This problem motivated Aguirregabiria and Mira to propose a recursive extension of the two-step method that they called the Nested Pseudo Likelihood method and that we describe below.

Sequential (Hotz-Miller) estimators

- Given the two-step PML estimator of $\boldsymbol{\theta}_u^0$ it is possible to construct estimates of choice probabilities which exploit the structure of the model.

- By using estimates of choice probabilities that exploit the structure of the model, one can get estimates of structural parameters with smaller finite sample bias an variance.

- Let $\hat{\mathbf{P}}_1 = \{\hat{\mathbf{P}}_1(a|x)\}$ such that:

$$\hat{\mathbf{P}}_1(a|x) = \Lambda(a|\{\tilde{z}(j,x;\hat{\mathbf{P}}_0,\hat{\boldsymbol{\theta}}_{f0})\hat{\boldsymbol{\theta}}_{u,1}^{PL} + \tilde{e}(j,x;\hat{\mathbf{P}}_0,\hat{\boldsymbol{\theta}}_{f0}) : j \in A\}$$

For instance, for the conditional logit model:

$$\hat{\mathbf{P}}_1(a|x) = \frac{\exp\left\{\tilde{z}(j,x;\hat{\mathbf{P}}_0,\hat{\boldsymbol{\theta}}_{f0})\hat{\boldsymbol{\theta}}_{u,1}^{PL} + \tilde{e}(j,x;\hat{\mathbf{P}}_0,\hat{\boldsymbol{\theta}}_{f0})\right\}}{1+\sum_{j=1}^{J}\exp\left\{\tilde{z}(j,x;\hat{\mathbf{P}}_0,\hat{\boldsymbol{\theta}}_{f0})\hat{\boldsymbol{\theta}}_{u,1}^{PL} + \tilde{e}(j,x;\hat{\mathbf{P}}_0,\hat{\boldsymbol{\theta}}_{f0})\right\}}$$

Sequential (Hotz-Miller) estimators

• Given the new estimates $\hat{\mathbf{P}}_1$, we can get the valuation operator $\mathbf{W}(\mathbf{P}, \boldsymbol{\theta}_f)$ now evaluated at $(\hat{\mathbf{P}}_1, \hat{\boldsymbol{\theta}}_{f0})$, and then obtain new value differences $\tilde{z}(a, x; \hat{\mathbf{P}}_1, \hat{\boldsymbol{\theta}}_{f0})$ and $\tilde{e}(a, x; \hat{\mathbf{P}}_1, \hat{\boldsymbol{\theta}}_{f0})$, a new pseudo likelihood function $l^{PL}(\boldsymbol{\theta}, \hat{\mathbf{P}}_1, \hat{\boldsymbol{\theta}}_{f0})$, and a new PML estimator that maximizes this function, say $\hat{\boldsymbol{\theta}}_{u,2}^{PL}$.

• We can proceed in this way an generate a sequence of estimators of structural parameters and conditional choice probabilities $\{\hat{\boldsymbol{\theta}}_{u,K}, \hat{\mathbf{P}}_K : K = 1, 2, ...\}$ such that for any $K \geq 1$:

$$\hat{\boldsymbol{\theta}}_{u,K} = \arg\max_{\boldsymbol{\theta} \in \hat{}} l^{PL}(\boldsymbol{\theta}, \hat{\mathbf{P}}_{K-1}, \hat{\boldsymbol{\theta}}_{f0})$$

and

$$\hat{\mathbf{P}}_K(a|x) = \Lambda(a|\{\tilde{z}(j, x; \hat{\mathbf{P}}_{K-1}, \hat{\boldsymbol{\theta}}_{f0})\hat{\boldsymbol{\theta}}_{u,K}^{PL} + \tilde{e}(j, x; \hat{\mathbf{P}}_{K-1}, \hat{\boldsymbol{\theta}}_{f0})\}$$

Sequential (Hotz-Miller) estimators

- All the estimators in this sequence are asymptotically equivalent to partial MLE and to one-step PML. Therefore, iterating in this procedure does not provide any asymptotic gain because the initial PML estimator is already asymptotically efficient.

- The NPL procedure has two interesting features:

  (I) *It is an algorithm to compute the MLE which can be computationally cheaper than NFXP.*
  (II) *It is a method to reduce the bias of the two-step Hotz-Miller estimator.*

Sequential (Hotz-Miller) estimators

- **(I)** Upon convergence this procedure provides, exactly, the MLE. This result holds regardless the initial estimator of $\mathbf{P}_0$ is consistent or not. The NPL procedure can be computationally much cheaper than NFXP. This is because the number of times that the valuation operator is solved can be much smaller under the NPL than under NFXP.

- **(II)** If the NPL is initialized with a consistent estimator of $\mathbf{P}_0$, NPL iterations reduce the finite sample bias and variance of the estimator of $\boldsymbol{\theta}_u$. This has been proved formally by Kasahara and Shimotsu (2005) using higher order expansions for the bias and variance of the sequence of PML estimators.

Patent Renewal Decisions (PAKES, 1986)

• **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. An "hedonic" approach is very limited.

• 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" is very challenging.

Pakes (1986)

• Pakes 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 **Renewal 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. It may go from a few hundred dollars to a few thousand dollars.

Model: Basic Framework

- 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.

Model: 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: $\pi_{i,t+1} = \delta \ \pi_{it}$.

Model: Stochastic process of patent profits

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

• $\tau_{i,t+1} \in \{0,1\}$ 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.

Model: Stochastic process of patent profits

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

• 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 $E(\max\left\{ \ x \ ; \ \xi_{i,t+1} \right\})$ value increases as well).

Model: Stochastic process of patent profits

• 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:

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

• 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:

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

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

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

Model: 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: $V_t(\pi_t^*) = 0$.

Model: Solution - A useful result

- 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:

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

- 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

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:

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

Estimation: 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:

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

The estimator is consistent an asymptotically normal for any R. The

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.

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).

Temporary Sales and Firm Inventories

• 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:

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

  - 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**:

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

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

- Therefore, **expected sales** are:

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

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

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

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** $(s_{it} + q_{it})/d_{it}^e$:

$$\eta_{y^e|p} \equiv \frac{-\partial y^e}{\partial p}\frac{p}{y^e} = -\left[\int I\left\{d^e \exp(\xi) \; ; \; s+q \right\} dF_\xi(\xi)\right] \frac{\partial d^e}{\partial p}\frac{p}{y^e}$$

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

- 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

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

[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

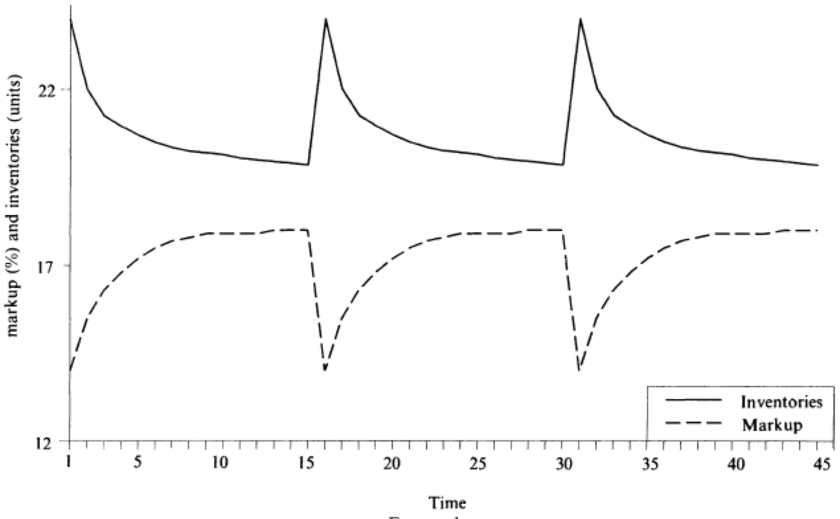

FIGURE 1

Time series of markup and inventories (without menu costs)

Evolution of inventories and price with menu cost

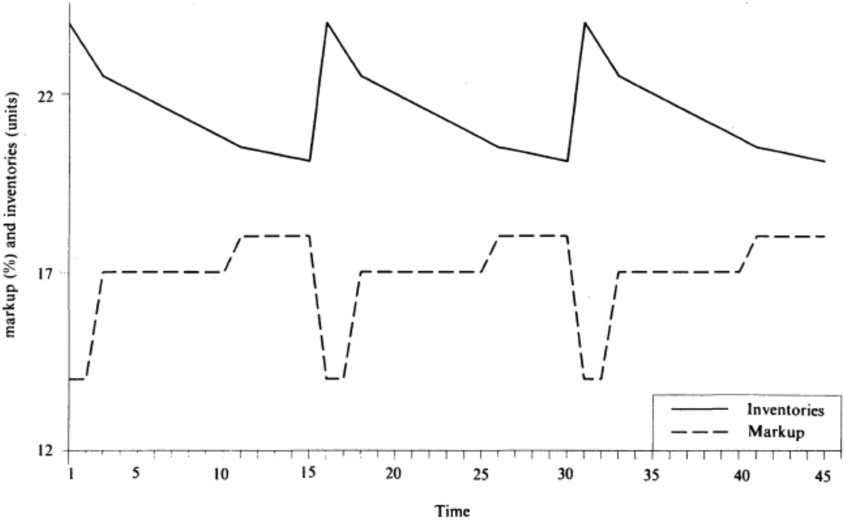

FIGURE 3

Time series of markup and inventories (with menu costs)

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

(2) Price (and markup) declines beginning of an (S,s) cycle (when an new order is made) and increases monotonically during the cycle until the next order.

• 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:

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

- 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.

$$q_{it} \in \{0 \; , \; \kappa_i\}$$

$$\Delta p_{it} \in \{0 \; , \; \delta_i^{(+)}, \; \delta_i^{(-)}\}$$

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:

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

(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.

284 REVIEW OF ECONOMIC STUDIES

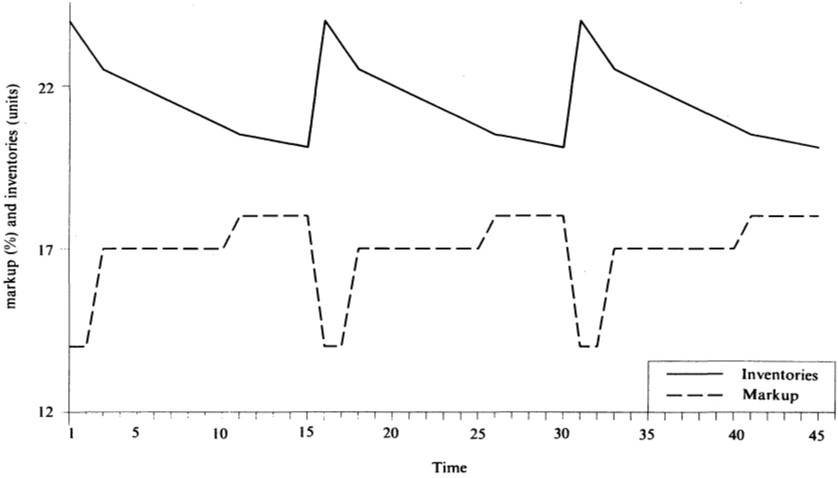

FIGURE 3

Time series of markup and inventories (with menu costs)

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

TABLE 1

Classification of brands in products and groups of products

Names of product and groups	Number of brands	Names of products and groups	Number of brands	Names of products and groups	Number of brands
1. Health, beauty	48	**5. Soft drinks**	29	**6. Food** (cont.)	
101. Eau cologne	2	501. Fruit juices	10	626. Pets food	15
102. Deodorant	2	502. Water	3	627. Salt	3
103. Soap, Shampoo	10	503. Carb. drinks 2L	5	628. Mayonnaise	9
104. Hair spray	6	504. Carb. dks small	11	629. Vinegar	2
105. Beauty creams	9	**6. Food**	277	630. Eggs	3
106. Toothpaste	3	601. Sugar	4	631. Butter, margar.	6
107. Shaving creams	3	602. Coffee	12	632. Cheese	11
108. Shaving blades	8	603. Instant coffee	10	633. Milk	6
109. Bath sponges	5	604. Malt	6	634. Powder Milk	5
2. Cleaning	69	605. Tea	2	635. Sausages	3
201. Detergent	12	606. Olive oil	7	636. Foie-gras	2
202. Bleach	12	607. Sunflower oil	4	**7. Alcohol. drinks**	89
203. Softeners	7	608. Canned fish	11	701. Cognac	10
204. Dishwashers	5	609. Dry cod	3	702. Whisky	4
205. Cleaning sprays	15	610. Canned veget.	8	703. Other spirits	12
206. Floor polish	7	611. Marmalade	9	704. Sherry	5
207. Scourers	4	612. Beans, Lentils	9	705. Champagne	6
208. Dischcloths	2	613. Rice	4	706. Vermouth	4
209. Brooms	1	614. Olives	6	707. Sidra	3
210. Insecticides	3	615. Pickles	4	708. Beer	12
211. Kitchen paper	1	616. Biscuits	23	709. wine (regular)	15
3. Pharmacy	9	617. Cakes	5	710. Wine (quality)	18
301. Bandages	4	618. Candies	11		
302. Sanit. napkins	2	619. Chocolate	30		
303. Diapers	3	620. Toasted bread	6		
Others, no food	13	621. Flour	2		
401. Alumin. paper	6	622. Pasta	14		
402. Plastic objects	2	623. Instant soups	11		
403. Batteries	4	624. Instant rice	5		
404. Stick gum	1	625. Dried fruits	6		

Descriptive Statistics

TABLE 2

Descriptive statistics: Distribution of brand-specific averages 534 brands. January 1990–May 1992 (29 months)

		Mean	Pctile. 10	Pctile. 25	Median	Pctile. 75	Pctile. 90
Orders	(1) Frequency $q > 0$ (%)	76·60	48·28	65·52	79·31	89·66	96·55
	(2) Duration (in months)	1·38	1·04	1·08	1·22	1·44	2·00
Markups	(3) Regular retail price (%)	20·50	9·46	15·69	19·95	24·77	30·31
	(4) Price under promotion (%)	5·10	−0·31	1·82	4·36	7·65	11·43
	(5) Retail price (%)	16·68	5·70	12·12	16·65	21·76	27·14
Regular Retail Price	(6) Freq. price changes (%)	46·60	17·86	28·57	46·43	60·71	78·57
	(7) Duration (in months)	2·18	1·18	1·44	1·80	2·56	3·60
Wholesale Price	(8) Freq. price changes (%)	34·63	10·71	21·43	34·14	42·86	50·00
	(9) Duration (in months)	3·46	1·82	2·10	2·86	3·80	6·00
Sales Promotions	(10) Freq. prom. >0 (%)	42·85	6·90	17·24	37·93	68·97	86·21
	(11) Ratio prom./total sales (%)	21·90	1·59	7·43	18·36	33·57	47·40
Retail Price	(12) Freq. price changes (%)	68·95	35·71	53·57	71·43	85·71	96·43
	(13) Duration (in months)	1·58	1·04	1·12	1·35	1·71	2·27

Reduced Form estimation of decision rules

TABLE 3

Fixed-Effects Probit Models for Discrete Choices (Standard errors in parentheses)

	Orders $q>0$		Regular retail price $I^{\Delta P \text{Reg}}=-1$		$I^{\Delta P \text{Reg}}=+1$		Sales promotions $I^{SP}=1$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln (s)	−0·911	−0·927	0·026	0·018	−0·044	−0·044	0·109	0·088
	(0·020)	(0·020)	(0·009)	(0·009)	(0·008)	(0·008)	(0·009)	(0·009)
b	−0·059	0·028	10·585	10·725	−11·728	−11·723	2·309	2·839
	(0·243)	(0·244)	(0·275)	(0·276)	(0·269)	(0·270)	(0·213)	(0·219)
c	−0·183	−0·059	0·668	0·840	−0·942	−0·934	0·390	0·377
	(0·228)	(0·232)	(0·206)	(0·208)	(0·199)	(0·201)	(0·206)	(0·212)
ln $y(t-1)$		0·084		0·117		0·006		0·384
		(0·024)		(0·022)		(0·021)		(0·021)
Log-lik.	−4,993	−4,985	−6,463	−6,448	−7,336	−7,336	−6,668	−6,498
Pseudo R^2 [a]	0·366	0·367	0·232	0·234	0·229	0·229	0·303	0·321
Number obs. [b]	14,056	14,056	14,868	14,868	14,952	14,952	13,888	13,888

Evolution of markup between two orders

TABLE 4

Markup behaviour between two orders

	Estimate of $\beta_{\tau,j}$ (in % points)	s.e.
$\tau = 1$		
After 1 month ($j = 1$)	−0·150	0·100
$\tau = 2$		
After 1 month ($j = 1$)	3·209	0·268
After 2 months ($j = 2$)	0·781	0·268
$\tau = 3$		
After 1 month ($j = 1$)	2·437	0·497
After 2 months ($j = 2$)	3·304	0·497
After 3 months ($j = 3$)	0·313	0·497
$\tau = 4$		
After 1 month ($j = 1$)	3·604	0·782
After 2 months ($j = 2$)	4·576	0·782
After 3 months ($j = 3$)	4·467	0·782
After 4 months ($j = 4$)	1·561	0·782

Estimation of Structural Parameters

TABLE 6

Menu costs and fixed ordering costs

Fixed ordering costs				
	Fixed OC per order	Fixed OC/ sales	Fixed OC/ items sold	Fixed OC/ variable OC
	\$278·11	3·15%	\$0·0365	3·35%

Menu costs

		MC per price change and per store	MC/sales	MC/items sold
	Total		0·70%	\$0·0081
Our study	$\Delta P > 0$	\$2·23	0·31%	\$0·0035
	$\Delta P < 0$	\$0·83	0·39%	\$0·0046
Levy *et al.*	(a)	\$0·52	0·70%	\$0·0119
	(b)	\$1·33	0·72%	\$0·0123
Slade		\$2·70		

[a] Supermarket chains where price tags are placed on the shelves but not on each individual item.
[b] Supermarket chains where a separate price tag is placed on each item (in addition to the shelf price tag).

Counterfactual Experiments

TABLE 7

Counterfactual experiments

	Scenario 1 $\eta_q = 278{\cdot}1$ $\eta^{p(-)} = 51{\cdot}5$ $\sigma_\xi = 0{\cdot}306$	Scenario 2 $\eta_q = 0{\cdot}0$ $\eta^{p(-)} = 51{\cdot}5$ $\sigma_\xi = 0{\cdot}306$	Scenario 3 $\eta_q = 278{\cdot}1$ $\eta^{p(-)} = 138{\cdot}7$ $\sigma_\xi = 0{\cdot}306$	Scenario 4 $\eta_q = 278{\cdot}1$ $\eta^{p(-)} = 51{\cdot}5$ $\sigma_\xi = 0{\cdot}0$
Frequency of positive orders (%)	70·1	100·0	64·2	69·7
Frequency of negative ΔP (%)	50·8	26·8	48·5	44·4
Frequency of positive ΔP (%)	29·6	10·9	19·4	14·2
Std. dev. of markup (%)	7·3	3·8	5·6	4·1

Demand of Storable Products and Intertemporal Price Discrimination

• Erdem, Imai, and Keane (2003): "Brand and Quantity Choice Dynamics under Price Uncertainty," Quantitative Marketing and Economics

• Hendel and Nevo (2006): "Measuring the Implications of Sales and Consumer Inventory Behavior," Econometrica.

• Hendel and Nevo (2013): "Intertemporal Price Discrimination in Storable Goods Markets," American Economic Review.

Introduction: Intertemporal Price Discrimination

• **Consumer heterogeneity** (in preferences, income, transportation, search, and storage costs) generate firm incentives to price discriminate.

• Consumer types are typically unobservable to firms. Therefore, firms need to design **screening mechanisms** to achieve separation of consumer types.

• Intertemporal Price Discrimination (IPD) is a specific screening mechanism that firms can use in markets of **durable products or storable products**.

• In the case of storable products, IPD can take the form of **temporary sales (high-low pricing)**.

Introduction: Consumers Stockpiling

• Consumers stockpiling behavior can introduce significant differences between short-run and long-run responses of demand to price changes.

• The response of demand to a price change depends on consumers' expectations/beliefs about how permanent the price change is.

• 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.

Introduction: Consumers Stockpiling (2)

• 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 temporary sales.

• 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.

Introduction: Biases of ignoring dynamics

• 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.

Introduction

• 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.

Outline

1. Introduction
2. Data and descriptive evidence
3. Model
4. Estimation
   - 4.1. Estimation of Brand Choice
   - 4.2. Estimation of Quantity Choice
5. Empirical Results

Type of Data (Consumer Scanner Data)

- 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.

Dataset in Hendel-Nevo (2006)

• Hendel and Nevo use consumer-level scanner data from Dominicks, a supermarket chain that operates in the Chicago area.

• The dataset comes from 9 supermarket stores and it covers the period June 1991 to June 1993.

• Purchases and price information is available in real (continuous) time but for the analysis in the paper it is aggregated at weekly frequency.

Dataset (structure)

- 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.

Descriptive evidence

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

Descriptive evidence

• 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 purchase has a positive effect on (1) probability of next purchase and (2) quantity purchased.

• (b) Indirect measures of low storage costs (e.g., house size) are positively correlated with households' propensity to buy on sale.

Model: Basic Assumptions

• Main challenge: allow for intertemporal substitution, but also for flexible substitution patterns between products.

• Every week a household has some level of inventories of the product and decides:

(a) how much to consume from its inventory;
(b) how much to purchase (if any) of the product,
(c) the brand to purchase.

Model: Basic Assumptions

• With J brands or products, an unrestricted dynamic deamnd model includes as state variables: the household inventories of the J products; and prices of the J products.

• This is impractical using a full-solution method.

• The authors impose important simplifying assumptions to reduce this very large state space.

• **Assumption 1:** Consumers care about brand choice when they purchase the product, but not when they consume or store it.

• Of course, the assumption imposes some restrictions on the intertemporal substitution between brands, and I will discuss this point too.

Utility function: Total utility

• 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;
• $C_h(i_{h,t+1})$ represents inventory holding costs;
• m_{ht} represents utility from product differentiation, and from the numerary good.

Utility from Consumption

- $u_h(c_{ht}, v_{ht})$ is the utility from consumption of the storable product.
- c_{ht} represents consumption (of any brand of laundry detergent).
- v_{ht} is a shock in the utility of consumption:
- Utility function from consumption is:

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

Inventory holding costs

- $C_h(i_{h,t+1})$ is the household inventory holding cost.

- $i_{h,t+1}$ is the level of inventory at the end of period t, after consumption and new purchases.

- Again, inventory does not distinguish brands, such that:

$$i_{h,t+1} = i_{ht} - c_{ht} + q_{ht}$$

- The inventory cost function is:

$$C_h(i_{h,t+1}) = \delta_{1h} \; i_{h,t+1} + \delta_{2h} \; \left[i_{h,t+1}\right]^2$$

Utility from Brand Choice

• 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 $X = 4$ is the maximum size.

• 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 h purchases x units of brand j at week t".

Utility from Brand Choice (2)

• 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.

• 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.

Utility from Brand Choice (3)

• 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 heterogeneity is simpler with consumer level data than with aggregate market share data.

• In particular, micro level datasets can include information on a rich set of household socio-economic 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.

Utility from Brand Choice (4)

- Finally, ε_{hjxt} is a consumer idiosyncratic shock that is independently and identically distributed over (h, j, x, t) with an extreme value type 1 distribution.

- This is the typical logit error that is included in most discrete models of demand of differentiated products.

- Note that while ε_{hjxt} vary over individuals, ξ_{jxt} do not.

Dynamic decision model

• 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 idiosyncratic shocks in preferences v_{ht} and ε_{ht}.

• Given this information, the household decides consumption of the storable product, c_{ht}, and how much to purchase and which product.

• 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.

Dynamic decision model (2)

• 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:

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

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.

Dynamic decision model (3)

• 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:

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

• 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.

Estimation of structural parameters

• The optimal decision rules $c_h^*(.)$ and $d_h^*(.)$ depend on the structural parameters of the model: the parameters in the utility function, and in the transition probabilities of the state variables.

• In principle, we could use the equations $c_{ht} = c_h^*(\mathbf{s}_{ht})$ and $d_{ht} = d_h^*(\mathbf{s}_{ht})$ 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.

Econometric issues (1)

- **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 infeasible 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.

Econometric issues (2)

• Though we have good data on households purchasing histories, **information on households' consumption and inventories of storable goods is not available.**

• In this application, consumption and inventories, c_{ht} and i_{ht}, are unobservable to the researchers.

• A household inventory is a 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).

Econometric issues (3)

• 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.

Reducing 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:

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

This Bellman equation is a contraction mapping.

Reducing state space

• 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 decision 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.

Quantity and Brand choice

• Two components in d_{ht}: quantity choice, x_{ht}, and brand choice j_{ht}.

• Conditional on a quantity choice, say $x_{ht} = x$, the **optimal brand choice is static**:

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

• Conditional on a quantity choice, the brand choice is a standard static demand model of differentiated product that we can estimate using standard assumptions.

Brand choice

- 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" $\{\delta_{jxt}\}$.

Quantity choice: Inclusive values approach

• If the quantity choice is made before knowledge of the $\varepsilon' s$, then the component m_{ht} of the utility function can be written as:

$$m_{ht} = \sum_{x=0}^{X} \omega_h(x, \mathbf{p}_t)$$

where $\omega_{ht}(x, \mathbf{p}_t)$ is the inclusive value:

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

Inclusive values approach (2)

- Therefore, the dynamic decision problem becomes:

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

- In words, the problem can seen as a choice between sizes, each with a utility given by the size-specific inclusive value.

- The dimension of the state space is still very 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.

Restriction on Process of Inclusive Values

• To reduce the dimension of the state space, Hendel and Nevo (2006) introduce the following assumption.

• Let $\omega_h(\mathbf{p}_t)$ be the vector with the inclusive values for every possible size $\{\omega_h(x, \mathbf{p}_t) : x = 1, 2, ..., X\}$.

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

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

Restriction on Process of Inclusive Values (2)

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

• Instead of all the prices and attributes, we only need a single index for each size.

• Two vectors of prices that yield the same (vector of) current inclusive values imply the same distribution of future inclusive values.

• This assumption is violated if individual prices have predictive power above and beyond the predictive power of $\omega_h(\mathbf{p}_t)$.

Restriction on Process of Inclusive Values (3)

• The inclusive values can be estimated outside the dynamic demand model.

• Therefore, the restriction can be tested and somewhat relaxed by including additional statistics of prices in the state space.

• Note, that $\omega_h(\mathbf{p}_t)$ is consumer specific: different consumers value a given set of products differently and therefore this assumption does not further restrict the distribution of heterogeneity.

Inclusive values approach

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

• With $X = 4$ and $J = 50$, this means a reduction of the state space from 601 to 5 continuous state variables.

Estimation of static brand choice model

- 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:

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

Estimation of static brand choice model

• 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.

Estimation of static brand choice model

• Consumer-level scanner datasets contain rich information on household socio-economic characteristics.

• Let z_h be a vector of observable socio-economic 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.

Estimation of static brand choice model

- 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:

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

where

$$\begin{aligned} \delta_{jxt} &\equiv \beta_0 a_{jxt} - \alpha_0 p_{jxt} + \xi_{jxt} \\ \sigma_{jxt} &\equiv a_{jxt}\sigma_\beta - p_{jxt}\sigma_\alpha \end{aligned}$$

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).

Estimation of static brand choice model

• 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:

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

• 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).

Estimation of static brand choice model

• 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 $\{\delta_{jxt}, \sigma_{jxt}\}$.

Estimation of static brand choice model

• Then, given these estimates of δ_{jxt} and σ_{jxt}, we can estimate demand parameters using an IV approach in the equations:

$$
\begin{aligned}
\widehat{\delta}_{jxt} &\equiv \beta_0 \; a_{jxt} - \alpha_0 \; p_{jxt} + \xi_{jxt} \\
\widehat{\sigma}_{jxt} &\equiv a_{jxt} \; \sigma_{\beta} - p_{jxt} \; \sigma_{\alpha}
\end{aligned}
$$

Construction of inclusive values

• Once we have estimated $(\beta_0, \alpha_0, \sigma_\beta, \sigma_\alpha)$, 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:

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

• Finally, we can get estimates of the inclusive values:

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

Construction of inclusive values

• And given the estimated inclusive values $\hat{\omega}_{hxt}$ we can estimate the stochastic process of these state variables.

• H&N propose an estimate a Vector Autoregressive (VAR) models for the vector of inclusive values.

Estimation of dynamic 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.

Estimation of dynamic quantity choice

• 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.

Estimation of dynamic quantity choice

• 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:

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

Estimation of dynamic quantity choice

- 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.

Estimation of dynamic quantity choice

• 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.

Estimation of dynamic quantity choice

• 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.